# Improving inventory system performance by selective purchasing of buyers' willingness to wait

Muzaffer Alım[a,*], Patrick Beullens[b]

[a] *Technology Faculty, Batman University, Batman, Turkey*
[b] *Mathematical Sciences and Southampton Business School and CORMSIS, University of Southampton, UK*
*Accepted for publication in EJOR on 12 July 2021. doi.org/10.1016/j.ejor.2021.07.027*

**Abstract**

We develop a demand postponement mechanism to improve the performance of a single item, periodic review inventory system with advance demand. The focus in the literature has been on how to stimulate customers towards advance demand. Predicting how demand will shift can be problematic, however, and backorders may still occur. We focus on how a firm can address backorders under a given advance demand pattern by a mechanism of compensation from which both the firm and the customers will benefit: the firm may offer a discount to customers for accepting later deliveries at a promised delivery date. Delivery postponement offers are made selectively, i.e. in some periods and to some customers only when there is a benefit for the firm to do so. Customers may decline the offer, but then face the probability of a backorder. In each period, the firm has to decide whether to make delivery postponement offers and for how long, and whether to order from its supplier and how much. We formulate the problem as a Markov Decision Process and solve it by backward induction. Numerical examples illustrate the properties of the state-dependent policies obtained for both uncapacitated and capacitated inventory systems. The postponement mechanism in capacitated systems leads to policies that differ from the threshold policy identified as optimal in the literature. Overall, the approach shows promise to improve system performance more efficiently compared to strategies aiming to increase advance demand in the system.

*Keywords:* Inventory, Backward Induction, Advance Demand, Discount and Replenishment Policy, Postponed Demand

[*] I am corresponding author
*Email address:* muzaffer.alim@batman.edu.tr (Muzaffer Alım)

## 1. Introduction

Inventory control is concerned with matching the supply of a product to its demand. Supplier lead-time (SLT), or the time it takes to receive an order placed at the supplier, forms an essential component of many inventory models. In general, however, also the time when customers place an order may differ from the time the order is arranged to be delivered. In the inventory literature, this time-lag is defined as the customer lead-time or 'demand lead-time' (DLT), and the techniques to control and exploit knowledge of DLTs as inventory control under Advance Demand Information (ADI).

Matching the supply to uncertain demand is a challenging task if SLT is non-zero, i.e. when supply cannot occur instantaneously. If supply exceeds demand, holding costs are incurred, and problems with warehouse capacity may arise. In the other case, backorders or lost sales occur. Improved performance may be obtained from reducing the SLT or increasing the DLT (Kremer and Van Wassenhove, 2014). Typically, the first is thought of as improving the responsiveness of the system, while the latter as improving its anticipatory power. Reducing SLT can be challenging if hard constraints are imposed, e.g. by production and transportation times. In some contexts, there might be a higher potential to increase the DLT. This usually requires renegotiating the terms offered to customers.

Changing the DLT is considered in various businesses in the context of revenue management (Talluri and Hillier, 2005). Increasing the DLT can be realized in two different ways. The first is known as advance booking or pre-ordering, whereby customers inform the company of their orders sooner, but without changing their due date. This strategy is used e.g. by Apple, Amazon and Playstation for new products, not yet released. For movies, games, or electronic products, firms may use it to induce customers to commit using these services at a later scheduled date, who in return receive a guarantee on their availability at a discount (Li and Zhang, 2013). This strategy locks in demand at earlier times, and enables the firm to better anticipate how to meet these future commitments with current decisions. A second strategy to increase DLT is known as demand postponement, and consists of offering customers incentives to postpone the date of delivery. It is a strategy applied in the airline industry as a mean to ameliorate overbooking (Tang, 2006), whereby customers are offered financial compensation if they would accept later flights. This strategy differs from pre-ordering because it is only offered selectively and when needed.

The advance demand pattern and demand postponement considered in this paper have great potential for application in real life, and for online sales channels in particular. We already see postponement policies being used by Amazon, for example, when it offers "No-Rush delivery (3-5 days delivery)" at the checkout with an Amazon gift card reward, although the customers are entitled to choose next day delivery (without the gift card award). Our paper models an extension of this approach in that it is only offered selectively and as means to better match demand with anticipated supply. Applications where a selective approach to setting the delivery date has been more commonly used includes the kitchen and home improvement/furniture sector (Alım and Beullens, 2020), but here the DLT is typically much longer, and the use of discounts in this sector, while common, is traditionally not explicitly linked to a postponement strategy. Finally, advance booking in online sales is now not limited to big brands. It is common for online sales channels in the DJ and music industry, for example, to allow customers to pre-order equipment expected to arrive in the online store by a specified future date at a lower price than the price charged by online competitors who do have the item in stock. The interesting question here would be what would happen if these competitors, who have the item in stock, offer the same terms as an option to their customers too? The latter situation is the kind of policy that is examined in this paper, although we do not consider the aspect of competition, but leave this as an idea for further research. Instead, the option to deliver later in return for a discount is offered when the firm would thereby reduce its total costs (including the price discount). The difference with the above current applications in the online sales channels is that the postponement policy considered in this paper is offered selectively at times when the firm needs it, even when the firm has still stock left. By the application of the selective postponement strategy, customers who are prepared to wait are rewarded for their patience, while companies may profit too since they use their resources more effectively. There is clearly also a competitive advantage to adopting such a policy, as argued above, although modelling these aspects would be highly complicated, and left for further research.

The literature review (Section 2) shows that the majority of studies on ADI in the context of production-inventory systems focus on getting more DLT from customers by offering them pre-announced DLT-dependent price schemes. Because these schemes are open to all customers, the level of uptake may highly affect the profitability of the firm, and it may be difficult to predict how the customer base will react to changes to the scheme. In this paper, we examine the option to use demand postponement as a means of better matching supply and demand in these systems. Provided that at least a fair proportion of customers are open to demand postponement, the financial risk involved in this approach seems less, as the firm can control the selective application of the policy. In this paper we do not model risk explicitly, but investigate the impact on the firm's expected profits.

In particular, we examine the potential of demand postponement under various DLT order patterns in a periodic-review inventory system. These patterns may range from very short DLTs (no ADI), to various heterogeneous DLT patterns (ADI). The demand postponement option is offered at some time periods only and to some particular customers only. We assume that the firm engages in such negotiations at the time when customers place their orders. From a customer service perspective, this approach is better than when such negotiations would take place afterwards and thus closer to an initially set due-date. It also signals that the firm is knowledgeable about its own supply chain constraints, and desires to make promises it can keep. As an incentive, the firm will offer a price discount on the order if a customer accepts the later delivery date. Like in other ADI models, the fact of having a policy to affect DLTs will also affect how the firm places replenishment orders at its supplier. The inventory control problem for the firm is hence to find the best joint supply ordering and ADI synchronization policy with respect to the trade-off between the loss due to awarded financial incentives and the gain in operational and backorders cost.

## 2. Literature Review

The study of ADI in inventory control is relatively recent and the core studies are presented in Table 1.

Table 1: Related Advance Demand Literature

| Paper | DLT | Review | Capacity | Early Delivery | Multi Stage | Price Discount |
|---|---|---|---|---|---|---|
| Hariharan and Zipkin (1995) | Hom. | Con. | | | ✓ | |
| Gallego and Özer (2001) | Het. | Per. | | | | |
| Gallego and Özer (2003) | Het. | Per. | | | ✓ | |
| Özer and Wei (2004) | Het. | Per. | ✓ | | | |
| Wang and Toktay (2008) | Het. | Per. | | ✓ | | |
| Xu et al. (2017) | Hom. | Per. | | ✓ | | |
| Chen (2001) | Het. | Per. | | | | A pre-announced price scheme to get ADI |
| Karaesmen et al. (2004) | Hom. | Con. | ✓ | | | Cost of getting ADI as a function of DLT |
| Alım and Beullens (2020) | | Con. | | | | Price discount to postpone some of the demand to next delivery |
| **This Paper** | Het. | Per. | ✓ | | | Price discount to obtain more ADI if needed |

Hom.: Homogeneous, Het.: Heterogeneous, Con.:Continuous, Per.:Periodic

Hariharan and Zipkin (1995) introduce the demand lead-time (DLT) concept and address the case of a continuous review system where all customers have the same DLT. They show that increasing DLT is equivalent to reducing SLT as both reduce uncertainty on future demand in the same manner. When DLT grows and becomes equal to SLT, the optimal inventory policy changes from make-to-stock to make-to-order.

The more general case of customers being heterogeneous in their DLT is addressed in Gallego and Özer (2001). They consider a periodic inventory problem and model the ADI by the vector of observed demand during period $t$, $D_t = (D_{t,t}, ..., D_{t,t+N})$ and where $D_{t,f}$ is the demand placed in period $t$ to be delivered at period $f(f \geq t)$, with a DLT of $(f - t)$. In the case of zero set-up costs, they show that a state-dependent base stock policy is optimal, and that demand information about due dates that fall behind the SLT plus a review period (which is called the *protection period*), has no operational value. For strictly positive set-up costs, the optimal policy structure becomes a state-dependent $(s, S)$ policy. For a detailed review of literature on ADI, we refer readers to Gallego and Özalp (2002) and Özer (2011).

The capacitated inventory system situations of the above two models are presented in Wijngaard and Karaesmen (2007) and Özer and Wei (2004), respectively. For positive set-up costs, the latter authors prove the optimality of a threshold policy. This policy involves ordering at full capacity when the inventory level at the end of the protection period (which is called the *modified inventory level*) falls below a threshold, and otherwise order nothing. The papers illustrate that ADI increases the effective usage of capacity. Multi-stage inventory systems with ADI are examined in Gallego and Özer (2003), Özer (2003) and Sarkar and Shewchuk (2013). Lu et al. (2003) study a multiproduct assemble-to-order system with ADI and stochastic lead times.

Other research in ADI systems investigates the benefits of allowing the firm to ship products to customers earlier than the due date. This approach is referred to as flexible delivery or early fulfillment. The exploitation of early delivery can significantly reduce inventory costs, see Karaesmen et al. (2004), Sarkar and Shewchuk (2016), Xu et al. (2017). Wang and Toktay (2008) investigate the cases with homogeneous and heterogeneous DLTs, and prove that increasing DLT is preferred over reducing SLT for the homogeneous DLT case. A variation is examined in Sarkar and Shewchuk (2013), in which there are two customer classes with equal priority, one class having customers with zero DLT, and the other class with a positive constant DLT. Xu et al. (2017) study the flexible delivery in ADI with homogeneous DLT and a penalty cost if the delivery is made later than its due date.

Most studies assume that ADI, once obtained, is fixed. In studies with imperfect ADI, such as in Tan et al. (2007), Gayon et al. (2009) and Benjaafar et al. (2011), it is however allowed that customers may cancel placed orders or make amendments to their order. The division of customers into different classes is sometimes used so that not only customers may have different DLTs but also different probabilities about cancellations or amendments. Gayon et al. (2009) exploit this in a Markov Decision Process having different transition probabilities for different demand classes. In the same customer differentiation context, Bao et al. (2018) and Van Wijk et al. (2019) determine different backlogging costs for different customers classes. This will result in backlogging some of the low priority customers even there is stock on hand for possible future high priority classes. Similarly, reservation of the on hand stock to different customer classes which each their own demand lead times has been studied by Du and Larsen (2017). Similar to this study, we consider customers as heterogeneous in terms of their DLT.

Since the results in above studies have identified benefits to the inventory system owner of having more customers adopting longer DLTs, some studies focus on methods to establish such transitions towards more ADI. Most studies have focused on offering financial incentives and evaluating the trade-off with the additional gains from increased ADI in lowering inventory or other system costs. Unlike in classic price discount models, where the overall demand may be affected, the studies in ADI typically keep total demand constant, but focus on establishing how price discounts may affect receiving future demand information earlier, or thus increasing DLT through advance booking. As customers might differ in their willingness to wait for their orders, Chen (2001) offers a price schedule where unit prices are non-increasing functions of DLT, and customers can view the price schedule in advance and self-select according to their preference. Kunnumkal and Topaloglu (2008) use price discounts to reduce the standard deviation of demand to a level which minimises the total relevant costs.

It should be noted that increasing DLT by financial incentives does not always lead to net gains. This is supported by the study of Karaesmen et al. (2004), who model the price discount as a function of DLT for the case of homogeneous customers in a production/inventory system in order to identify the optimal ADI level. Li and Zhang (2013) consider a model offering a price discount or price guarantee to customers in exchange of adopting a larger DLT through pre-ordering, but find that this policy actually may reduce the seller's profit in situations of low demand seasonality. Buhayenko and van Eikenhorst (2015) study the coordination of a supplier production schedule with deterministic demand through a price discount as an incentive for changing customers due dates and when no backorders are allowed.

The objective of this study is to take a given ADI setting as a starting point for seeking further improvements. That is, we assume that some level of ADI is present in the demand pattern, including the case when there is no ADI at all (DLT = 0). We focus on establishing when the seller would benefit from offering to 'buy' additional ADI by deciding when to make selective offers for demand postponement to some customers at the time they place their order. Note that this is different from the approach in Chen (2001) or Karaesmen et al. (2004) whereby all customers are offered an incentive scheme a priori. This makes our paper more dynamic since the discount decision is based on the current state and may not be offered if this does not result in a benefit. All previous ADI studies focus on getting more ADI with pre-announced price schemes. Unlike all these, we treat the increase of DLT as an instant decision based on system information, and the purchasing of the additional DLT is hoped to lead to a better matching of inventory and demand information on hand.

While selective demand postponement is a strategy already widely applied in the airline industry, inventory systems are very different in terms of structure. In an inventory system, for example, there are no fixed flights schedules. An inventory system may or may not have capacity constraints, and both products and orders generally differ in when they enter and leave the system. Customers may also reject the offer of demand postponement and then expect to have the products at the desired time, which the firm may or may not be able to meet. The latter case will result in a backorder.

The postponement decision in return of a price discount is similar to the policy introduced in Alım and Beullens (2020). However, the model developed in that paper is for a continuous inventory-distribution system where demand is deterministic at a constant rate. All postponed demand is dispatched with the next order arrival from the supplier. In contrast, the demand pattern we study in this paper has a more complex structure, it is stochastic and each demand can have its own lead time, and we study this in a periodic review inventory system. Another difference is that the postponement period is a decision variable in this paper, rather than all discounted demand being postponed to the same period.

We model selective demand postponement in a system with stochastic advance demand that is similar to that of Gallego and Özer (2001). We formulate the problem of finding the best combined inventory replenishment and offer synchronization strategy as a Markov Decision Process and solve it by backward induction. We examine the potential of this strategy in uncapacitated and capacitated systems, in cases of zero and positive set-up costs, and at various levels of ADI present in the system.

## 3. Problem Description

This section provides the formal description of the problem and introduces the modelling of the demand structure in the context of DLT and postponement.

A periodic review inventory problem for a single product type is considered. Table 2 summarizes the notation. The firm supplies its stock from an external supplier. A supply order placed at the beginning of period $t$ is delivered at the beginning of period $t + L$, where $L$ is the supplier lead time. With each supply order placed, the firm incurs a fixed setup cost $K_t$ and pays for the items based on a purchase cost per unit item $u_t$.

The firm has a large customer base. For each customer, it has agreed on a contract which specifies an (customer-specific) agreed DLT that applies to all orders the customer will place. The maximum possible DLT is called the *information horizon* and denoted by $N$, where we assume $N \geq L + 2$ (without loss of generality, see Section 4.1). Customers can place orders at random time periods. Most of the time, an order placed in period $t$ is accepted with the aim to deliver the products in period $t + DLT$, i.e. according to the agreed DLT. In some periods, the firm will consider making a delivery postponement offer to some customers in order for them to arrange delivery beyond the agreed DLT, who either accept or reject this offer.

The sequence of events in period $t$ can be summarized as follows: (1) Order from supplier if placed in period $t - L$ is received; (2) Decision is made to either place a supply order and the quantity, or not to place an order; (3) During period $t$, customer orders are received, if relevant then postponement offers are made, and the responses from customers to accept or reject the offer are received; (4) At the end of the period, customer orders needing delivery are fulfilled or backordered.

### *3.1. Demand Structure*

We consider the advance demand pattern structure as modelled in Gallego and Özer (2001). For the sake of clarity, we shortly review its main components. We do not yet consider the impact of the postponement policy, which will be introduced in Section 3.2.

Table 2: Notation of problem description

| | **Parameters** |
|---|---|
| $T$ | Time horizon |
| $N$ | Information horizon |
| $L$ | Supplier lead time |
| $C_t$ | Capacity limit at period $t$ |
| $\alpha$ | Opportunity cost of capital rate per period |
| $P_t$ | Fraction of demand which is accepting delivery postponement |
| $I_t$ | Inventory on hand at the beginning of period $t$ |
| $B_t$ | Backorder level at the beginning of period $t$ |
| $x_t$ | Modified Inventory position at period $t$ before the decisions |
| $y_t$ | Inventory position at period $t$ after decisions made at period $t$ |
| | **Demand** |
| $O_{t,s}$ | Cumulative observed demand for period $s$ at the beginning of period $t$ |
| $O_t$ | Observed demand beyond protection period at the beginning of period $t$ |
| $O_t^L$ | Observed protection period demand at the beginning of period $t$ |
| $D_{t,s}$ | Demand placed at period $t$ to be delivered at the end of period $s$ |
| $D_t$ | Demand placed at the end of period $t$, $D_t = (D_{t,t}, ..., D_{t,t+N})$ |
| $Q_t$ | Demand willing to accept the price discount at period $t$ |
| | **Costs** |
| $K_t$ | Setup cost per order at period $t$ (£/order) |
| $u_t$ | Purchase cost per unit at period $t$(£/product) |
| $h_t$ | Holding cost per unit at period $t$(£/product, period) |
| $p_t$ | Shortage cost per unit at period $t$(£/product, period) |
| $d_{ti}$ | Discount per unit for postponed delivery to $i$ at period $t$ (£/product) |
| | **Decision Variables** |
| $z_t$ | Order amount placed in period $t$ at supplier to be delivered at $t + L$ |
| $q_{ti}$ | Amount of demand placed in $t$ to be delivered with postponement in $i \in \{t + L + 1, ..., t + N\}$ |

During a period, the firm gathers the demand information from customers, each of whom may have a specific contract DLT. If $D_{t,s}$ denotes the total demand placed at period $t$ to be delivered according to the contract DLT at period $s \in \{t, .., t + N\}$, then at the end of period $t$, the firm has obtained a demand vector $D_t = (D_{t,t}, ..., D_{t,t+N})$.

A similar process has occurred during periods prior to $t$. Therefore, at the beginning of period $t$ and prior to obtaining $D_t$, the cumulative observed demand to be delivered at period $s \in \{t, .., t+N\}$ according to the contract DLT, will consist of demand orders placed in periods no earlier than period $s - N$ and no later than $t - 1$. If none of the demand is postponed, the cumulative observed demand at the start of period $t$ to be delivered in $s$ thus consists of:

$$O_{t,s} = \sum_{r=s-N}^{t-1} D_{r,s}. \tag{1}$$

At the beginning of any period $t$, the observed quantity $O_{t,s}$ can be expected to be only part of the demand destined to be delivered in $s$, since at any period $k \in \{t, .., s\}$ additional demand orders will arrive. Let $U_{t,s}$ denote the quantity of yet unobserved demand at the start of $t$ for delivery in $s$, then:

$$U_{t,s} = \sum_{r=t}^{s} D_{r,s}, \tag{2}$$

where these $D_{r,s}$ quantities are still undetermined random variables. The sum of $O_{t,s}$ and $U_{t,s}$, were the latter quantity known, would thus be the total quantity to be delivered in $s$ at the beginning of $t$ (prior to making any future offers to postpone).

The *protection period* refers to the period covering the next $L+1$ periods. Any customer orders to be delivered during this time can only be fulfilled from on hand stock and planned supply order arrivals. At the beginning of $t$, the total observed demand to be delivered in periods that fall within the protection period is thus:

$$O_t^L = \sum_{s=t}^{t+L} O_{t,s}, \tag{3}$$

while the total unobserved demand within this protection period equals:

$$U_t^L = \sum_{s=t}^{t+L} U_{t,s} \tag{4}$$

Figure 1 illustrates the demand pattern structure available in period $t$. Observed and unobserved demand at the beginning of this period are displayed above the x-asis, where the components of $D_t$ obtained during the period are displayed below.

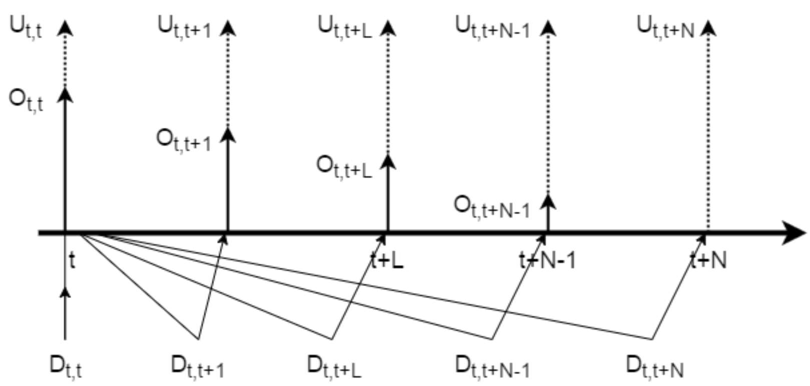


Figure 1: Demand Pattern

**Example 3.1.** *Let us assume that we have a case of* $N = 2, L = 0$*, and no postponement is offered. Then at the beginning of period* $t$*, we have* $O_{t,t}$ *and* $O_{t,t+1}$ *as observed cumulative demand from previous periods. Also we have unobserved demand* $U_{t,t} = D_{t,t}$ *and* $U_{t,t+1} = D_{t,t+1} + D_{t+1,t+1}$*.*

*During the period* $t$*, the demand vector* $D_t = (D_{t,t}, D_{t,t+1}, D_{t,t+2})$ *is observed. We deliver* $O_{t,t}+D_{t,t}$*. To move towards the next period, we update the observed demand as* $O_{t+1,t+1} = O_{t,t+1} + D_{t,t+1}$ *and* $O_{t+1,t+2} = D_{t,t+2}$*. The unobserved demand is also updated,* $U_{t+1,t+1} = D_{t+1,t+1}$ *and* $U_{t+1,t+2} = D_{t+1,t+2} + D_{t+2,t+2}$*.*

### *3.2. Demand Postponement*

Customers are offered a price discount to postpone the due date for their demand. The discount decision is only available for the protection period demand at only some times. The reason to consider offering a discount only to these demands is to protect the system against the lead time demand. The postponement period is a decision variable. Each customer is treated separately and they might have different delay periods.

During the period $t$, the demand $D_t$ is observed. Before satisfying the demand, it is decided whether to offer price discount. The discount decision is only available for protection period demand which consists of $D_{t,t}$,..., $D_{t,t+L}$. The discount decision $q_{ti}$ represents the number of discounted demand at the beginning of period $t$ delayed to period $i \in \{t + L + 1, ..., t + N\}$. Only a specific rate of the customers $P_t$ accepts the discount and the upper bound on the total demand willing to accept the postponement is therefore as shown by (5). The discount decision offered to the demand placed at period $t$ is illustrated in Figure 2.

$$Q_{t+1} = P \sum_{f=t}^{t+L+1} D_{t,f} \tag{5}$$

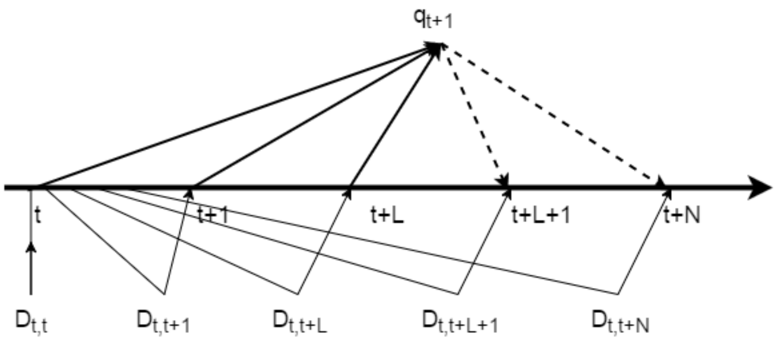


Figure 2: Discount Decision

We give the upper bound $Q$ the subscript $t + 1$ because the postponement decision on demand placed in $t$ is made at the end of period $t$, which is mathematically equivalent in information content to the start of period $t + 1$.

## 4. Model Formulation

The mathematical model and its components are presented in this section. We formulate the problem as a discrete-time Markov Decision Process (MDP). The main assumptions of the problem are as follows.

1. Supplier lead time is deterministic and constant.
2. Discounts are not announced beforehand. After all customer order placements are collected during a period, then the decision is made whether or not to offer the price discount postponement offer.
3. The amount of discount per product offered can be a function of the period to which delivery is postponed.
4. The unobserved part of the demand is not dependent on the observed part (a large customer base).
5. All unsatisfied demand is backordered.
6. Early delivery of a demand is not allowed.
7. The total warehouse capacity is limited, also known as "bounded inventory", as in Guan and Liu (2010); Gutiérrez et al. (2003).

These assumptions seem reasonable in the context of e.g. home improvement or building construction, where customers in general offer advance demand information, are reluctant to accept early deliveries, but accept backorders rather than resulting in lost sales. Also, manufacturing and building supply firms in this sector face capacity limitations on their warehouse or site.

The components of a MDP are state space, decision space, transition function/probability and reward function.

### *4.1. State and Decision Variables*

At each decision epoch, the system occupies a state from a state space. Gallego and Özer (2001) show that the state space for advance demand information has a dimension of $1 + (N - L - 1)^+$, with state variables the modified inventory position $x_t$ and the observed demand beyond protection period $O_t$. Adding the delivery postponement decision into the system obligates us to keep track of the information of maximum number of demand $Q_t$ for postponement. The state space then becomes $2 + (N - L - 1)^+$ dimensional. Notice that when $N < L + 2$, the system becomes two dimensional. However the discount decision will not take place since there is no point affecting demand beyond the protection period. Therefore the problem becomes one dimensional. We can thus focus on problems where $N \geq L + 2$.

Let the state of the MDP at the beginning of period $t$ be described by $s_t = (x_t, O_t, Q_t)$, where:

$$x_t \equiv I_t + \sum_{s=t-L}^{t-1} z_s - B_t - O_t^L \tag{6}$$

$$O_t \equiv (O_{t,t+L+1}, ..., O_{t,t+N-1}) \quad \textit{where} \quad O_{t,s} = \sum_{r=s-N}^{t-1} (D_{r,s} + q_{r,s}) \tag{7}$$

$$Q_t = P_{t-1} \sum_{f=t-1}^{t+L} D_{t-1,f} \tag{8}$$

Here, $x_t$ is the planned inventory position at the end of period $t + L$, a function of on hand inventory level, planned supply order arrivals, backorder level, and observed demand over the protection period as in (6). $O_t$ is the cumulative observed demand vector beyond protection period composed of placed demand and delayed demand at previous periods. This is the demand information needed to be considered for future periods. The last state variable is $Q_t$ which is the number of demand willing to accept delivery postponement.

Based on the state variables, order and discount decisions $a_t = (z_t, q_{ti})$ are placed, where $z_t$ is the order decision and $q_{ti}$ is the postponed demand placed at period $t$ to be delivered at period $i$. The sum of postponed demand is shown as $q_t$ by (9). The number of demand available for postponement is limited by $Q_t$, the limitation constraint is given in (10).

$$q_t = \sum_{i=t+L+1}^{t+N} q_{t,i} \tag{9}$$

$$q_t \leq Q_t \tag{10}$$

The cost placing the decision set $a_t = (z_t, q_{ti})$ is notated as $\beta(a_t)$ and calculated by (11).

$$\beta(a_t) = K_t\delta(z_t) + z_t u_t + \sum_{i=t+L+1}^{t+N} (d_{ti} q_{ti}) \tag{11}$$

After the decisions are made, the modified inventory position is updated to $y_t$ in (12). If there is no capacity constraint in the model, there are no limits imposed on both the order quantity or the warehouse storage space capacity. In the capacitated case of ADI, Özer and Wei (2004) limit the order size. In this paper, we instead model a capacity constraint as an upper bound on the modified inventory level after having made the postponement decisions. This can be thought of as a capacity reservation limit $C_t$ on $y_t$ as in (13).

$$y_t = x_t + z_t + q_t \tag{12}$$

$$y_t \leq C_t \tag{13}$$

The decision of offering discounts and the ordering decision affect the ending inventory $y_t$ as in (12). But also, the demand beyond protection period increases by price discount decisions, since the part of the current demand is postponed to those periods (see (7)) .

### *4.2. State Transitions*

The system state at the next decision stage depends on the current state, the decisions made at the current period, and the particular realisation of demand. The transition between one state to another could be either deterministic or stochastic depending on how this demand realisation is modelled. However, considering the stochastic transition increases the computational complexity of the model exponentially. Therefore, in our study, we focus on deterministic transition as Gallego and Özer (2001) and the function of moving to $s_{t+1} = (x_{t+1}, O_{t+1}, Q_{t+1})$ are shown in (14), (15), (16). (See (7) for the calculation of $O_t$ and $O_{t,s}$.)

$$x_{t+1} = y_t - D_{t,t} - \sum_{s=t+1}^{t+L+1} D_{t,s} - O_{t,t+L+1} - q_{t,t+L+1} \tag{14}$$

$$O_{t+1} = (O_{t+1,t+L+2}, ..., O_{t+1,t+N}) \tag{15}$$

$$Q_{t+1} = P_t \sum_{f=t}^{t+L+1} D_{t,f} \tag{16}$$

### *4.3. Cost Function*

After the decisions, the net inventory level at the end of period $t + L$ becomes $y_t - U_t^L$ where $U_t^L$ is the unobserved demand for protection period. The expected inventory costs at that period is calculated based on $y_t$ by (17). The expected value is due to the uncertainty on demand placed at period $t$.

$$\hat{G}_t(y_t) \equiv \frac{\alpha^{t+L}}{\alpha^t} Eg(y_t - U_t^L) \tag{17}$$

where $\alpha$ is the discount factor and $g(x)$ is total expected holding and backorder cost based on the inventory level $x$. We assume that it is a convex function and coercive since $\lim_{|x|\to\infty} g(x) = \infty$. These are the common assumptions in inventory literature (see Gallego and Özer (2001); Özer and Wei (2004)) and they are valid when the holding and backorder costs are linear.

The total expected cost including inventory and decisions cost is then shown by $c_t$ and calculated by (18).

$$c_t(s_t; a_t) = \hat{G}_t(y_t) + \beta(a_t) \tag{18}$$

*4.4. Recursive Equation*

With all the given components, the recursive equation also known as Bellman Equations are formulated as in (19) and (20).

$$V_T^*(s_T) = c_T(s_T, a_T) \quad \forall s_T \in S, \forall a_T \in A_{s_T} \tag{19}$$

$$V_t^*(s_t) = min_{a_t \in A_{s_t}} \Big\{ c_t(s_t; a_t) + \alpha E V_{t+1}(s_{t+1}) \Big\} \quad \forall t \in \{0, 1, ..., T-1\} \tag{20}$$

$V_t^*(s_t)$ is the optimal expected cost when the system in state $s_t$ at time $t$. (19) presents the termination condition when time horizon is completed.

## 5. Action elimination and algorithm

*5.1. Action Elimination*

Increasing the number of dimensions in the state space, as we have, increases computational complexity. Reducing the action space will help to narrow the searching area at each iteration, and thus alleviate computational burden. In this part, we focus on reducing the action sets in order to improve the computational complexity. The idea is to find an upper bound for the reorder level so that above that level, there will not be any order decision. To this end, we have conducted some analysis of our experimental results as to better understand the behaviour of the reorder level. This can be summarised as follows

**Proposition 1.** *The following statements seem to hold experimentally, but are not formally proven, for any $o$ and for all $t$;*

1. *If $K_2 > K_1 > 0$ then $s_t(o|K_2) \leq s_t(o|K_1)$*
2. *$s(o_2) \geq s(o_1)$ if $o_2 \leq o_1$*
3. *$s_{t-1}(o) \geq s_t(o)$*

**Remark 1.** *Statement 1 indicates that the reorder level is decreasing by increasing setup cost, all other parameters remaining the same. This is as can be expected, since higher setup cost causes fewer orders. Having less orders means that the order interval is extended, that is, the reorder level will be at lower inventory levels. This statement, if true, implies that the inventory systems with positive setup cost will have a lower reorder level than zero setup cost cases.*

**Remark 2.** *Statement 2 shows that the reorder level is increasing with observed demand beyond the protection period. Having more demand waiting in the upcoming period can cause us to place orders in the next period instead of placing an order in the current period and carrying stock to the next period. Thus, we would expect a decrease on reorder level. The numerical results in Gallego and Özer (2001) as well as in this paper support this observation.*

**Remark 3.** *For the finite-horizon (under stationary costs, as assumed in our experiments), the reorder level is reducing by time since the system is getting closer to the terminal point as referred to in Statement 3. A similar statement is found in Özer and Wei (2004).*

Gallego and Özer (2001) prove that the optimum policy is a base stock policy when the setup cost is zero. Thus, for stationary problems, the base stock parameter of the myopic policy which only minimises the cost of the current stage is optimal for a finite time horizon situation as we consider. This parameter is calculated by (21) and (22): there exists a unique value of $y^m_{min}$ which makes the gap between the minimum and maximum of the cost function zero when $G_t(x)$ is strictly convex (Gallego and Özer, 2001).

$$y^m_{min} = min\{y : G_t(y) = min_x G_t(x)\} \tag{21}$$

$$y^m_{min} = max\{y : G_t(y) = min_x G_t(x)\} \tag{22}$$

Now, when we take the statements in Proposition 1, it is concluded that the reorder level for zero setup cost is an upper bound for reorder level of positive setup cost cases. Calculating the policy parameter for $K = 0$ is as easy as shown in (21) and (22) and this could be used to reduce the action space for cases, $K > 0$. We assume that $s_{K_0}$ be the optimal base stock policy parameter for the zero setup cost case. Then, we can make an additional constraint as (23)

$$z_t = 0 \quad if \quad x_t \geq s_{K_0} \tag{23}$$

Practically, (23) will reduce the search area. Because when the modified inventory level, $x_t$ is greater than the reorder level, $s_{K_0}$ found by (21) and (22), then it is known that the optimal decision is not to order.

It should be clear that applying this search space reduction should still be considered a heuristic approach to achieving faster computational time.

*5.2. Solution Algorithm*

Solving the Bellman equations is often performed by backward induction. The backward algorithm starts to calculate the value function at the last period for each of the states. It then steps back in time by one period and calculates the values by adding it to the next periods value function. The algorithm is shown in Algorithm 1.

| **Algorithm 1:** Backward Algorithm | |
|---|---|
| **Step 1.** | Initialize the terminal values. Set $t = N$ ; |
| | $V_N(s_N) = c_N(s_N) \quad \forall\ s_N \in S$ ; |
| **Step 2.** | Substitute $t-1$ for $t$ and compute $V_t(s_t)$ $\forall\ s_t \in S$ by ; |
| | $V_t^*(s_t) = min_{a \in A_{s_t}} \left\{ c_t(s_t, a) + \alpha \sum_{s' \in S} p_t(s'\|s, a) V_{t+1}(s'_{t+1}) \right\}$ ; |
| | Set ; |
| | $a_{t,s_t} \equiv argmin_{a \in A_{s_t}} \left\{ c_t(s_t, a) + \alpha \sum_{s' \in S} p_t(s'\|s, a) V_{t+1}(s'_{t+1}) \right\}$; |
| **Step 3.** | If t=1 stop; Otherwise return to step 2 ; |

## 6. Computational Results

In this section, we carry out a number of numerical experiments to analyse the behaviour of policies under advance demand and the impact of postponement on these policies. For the numerical values of parameters, we follow the studies of Gallego and Özer (2001) and Özer and Wei (2004). The most inclusive and simple case on our problem without loss of generality is to test the case where $N = L + 2$ in which the observed demand beyond protection period has become one dimensional.

Advance demand $D_t = (D_{t,t}, D_{t,t+1}, D_{t,t+2})$ is modelled following a Poisson distribution with means $\lambda = (\lambda_0, \lambda_1, \lambda_2)$ when $N = 2$. To see the impact of more ADI available we test the model with different values of $\lambda$. While doing this, we keep the total demand rate $\lambda_0 + \lambda_1 + \lambda_2$ constant so that we are able to see the effect of the policy which shifts the urgent demands to advance demands. The value of ADI in the uncapacitated and capacitated cases are analysed and compared.

The effect of the discount is also tested under different parameters. Since we have a two dimensional discount decision, we usually offer higher discount amounts to those whose postponement period is longer, except when otherwise stated. The customers reaction to postponement is chosen from three probabilities $\{0, 0.5, 1\}$. When $P = 0$, then only at the current period, we can offer postponement if there is any demand available $(Q_0)$. For future periods, there is no demand available for postponement. This can be seen as the worst case scenario. $P = 0.5$ means that only half of the customers welcome the postponement in any of the future periods. All customers accepting the discount corresponds to the case that $P = 1$. This is the best case scenario in terms of customers acceptance of postponement.

Gallego and Özer (2001) prove that the optimal inventory policy for zero setup cost case with advance demand is the base stock policy. When the inventory level drops below the base stock level, then it is optimal to order up to the base stock level. Since the cost is calculated based on the inventory level at the end of the protection period, the ordering decision can increase the inventory level to the base stock level. As the order decision takes place free of charge when the setup cost is zero, we do not need to consider the demand postponement. Thus, our discount policy does not make any difference. So we focus on cases with positive setup costs.

For the positive setup cost without discount decision, the optimal policy is a state dependent $(s(o), S(o))$ policy which is order up to $S(o)$ if inventory level is at $s(o)$ or below. Based on the numerical results, we have noticed that adding discount decisions turns the policy to $(s(o, Q), S(o, Q), w_i(o, Q), W_i(o, Q))$ where $i \in \{L + 1, .., N\}$ refers to the delay period, and $Q$ is the demand available for price discount at the starting period. The order policy is similar to the non discount case but the parameters are now also dependent on $Q$.

Figure 3 illustrates the decisions based on the modified inventory positions. The discount policy is to offer discounts to bring the inventory position up to $W_i(o, Q)$ if the inventory level is greater than $w_i(o, Q)$ with a delay period $i$, however at any time the maximum amount by which the inventory position can be raised is limited to $Q$. The value of $w_1(o, Q)$ is the reorder level. If the inventory position is above $w_2(o, Q)$ but $w_2(o, Q) < W_1(o, Q)$, we offer the postponement with a delay period of 2. To be clear, postponement and ordering decisions never take place together in the same period. Only if the inventory position is below or equal to the reorder level $s(o, Q)$, do we place an order up to $S(o, Q)$, and then no demand postponement decision is taken.

While we have observed this policy to be found to hold in our numerical experiments, we have not been able to construct a formal proof that this is an optimal policy in general.

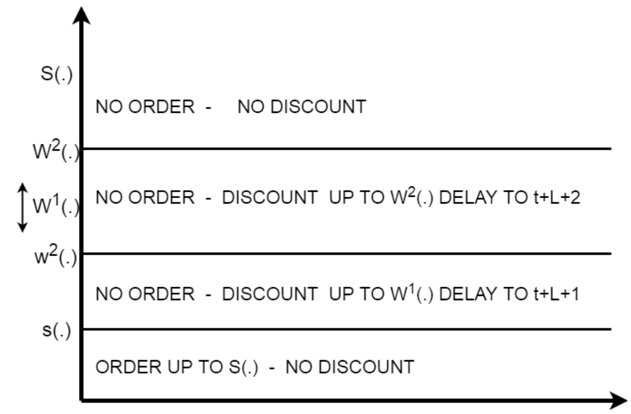


Figure 3: Order & Discount Policy

In the numerical experiments, we aim to show the value of ADI and effect of demand postponement in capacitated and uncapacitated inventory systems. Then we would like to compare these two policies to derive insight into the value of obtaining more ADI or using the postponement policy. The performance of both policies are presented under different conditions.

In setting up the experiments, we adopted the financial parameter values from Gallego and Özer (2001): the set-up cost per order by the online retailer to its supplier is £100, the unit holding cost is £1 per item per year, and the backorder cost is £9 per item. To keep the analysis tractable, we do not vary these financial parameters. However, we do test the sensitivity to varying the level of the discount price.

We can estimate the cost price of the item from the unit holding cost. Assuming that the capital rate of the firm is 0.05 per annum, then from the holding cost value we can estimate a corresponding item cost price incurred by the online firm of roughly £20[1], where the customer ordering it online is assumed to pay a higher price, say £30. Values of discounts investigated in this paper are usual in the range of £1 to £2, but we also investigate larger values up to £13. This range of discount values seems realistic to us considering the value of the item.

Many inventory models in the literature are based on considering backorder costs. It is understood that its value must include both real incurred costs experienced from satisfying the backorder, such as e.g. a more expensive delivery, but also may need to include the difficult to quantify impact of the "loss of future goodwill". The £9 remains at this point thus still a somewhat arbitrary decision, and further research may be needed to establish good methods to determine reasonable values[2].

### *6.1. Uncapacitated Inventory*

#### *6.1.1. The value of ADI*

In this section, we test the value of ADI and policy parameters for an uncapacitated inventory system. Initially we ignore the demand postponement to see the status quo of ADI. We test the no discount model with different means of demand starting from no advance demand case $\lambda = (6, 0, 0)$, through various settings for $\lambda$ as indicated in the first column of Table 3, up to the full advance demand $\lambda = (0, 0, 6)$. The percentage saving on cost is calculated by $\delta_A = 100(Cost_{noADI} - Cost_{\lambda})/Cost_{noADI}$, where $Cost_{noADI}$ measures the total cost in the case of no ADI $(6, 0, 0)$, and $Cost_{noADI}$ the cost in the case of the other ADI values reported in the first column of Table 3. The table further reports on the optimal policy structure, the minimum expected cost, and the relative benefit of ADI, $\delta_A$.

Increasing the ADI present in the system reduces cost by up to 6.79%. This value also gives an idea to the manager on how much to invest in methods to obtaining increasing levels of ADI from customers. As expected, the reorder level and order up to level decrease with the increase in advance demand since the uncertainty on demand is reducing.

#### *6.1.2. The effect of Postponement on Inventory Policy*

In this section, we evaluate the impact of demand postponement on the inventory policy and the inventory cost. We test the model for the cases that the customer accepting probabilities are $P = 0$ and $P = 1$, respectively, to show how the policy is changing between the worst and the best case scenario.

Recall that the case $P = 0$ means that only the initial state demand $Q_0$ (if $Q_0 > 0$) is ready to accept the postponement. Apart from this, no more demand in further periods is available to be

[1]Using $h = \alpha v = 1 = 0.05(20)$, see e.g. Silver et al. (1998)

[2]One such method may be based on Net Present Value techniques, see e.g. Ghiami and Beullens (2016).

Table 3: $K = 100, h = 1, p = 9$, $D_{t-1,t+1} \in \{0, ...10\}$, $T = 12$

| $\lambda$ | | $D_{t-1,t+1}$ 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Avg.Cost | $\delta_A(\%)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6,0,0 | $S(.)$ | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 289.085 | 0 |
| | $s(.)$ | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | | |
| 4,1,1 | $S(.)$ | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 286.896 | 0.76 |
| | $s(.)$ | 0 | 0 | 0 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | | |
| 3,1,2 | $S(.)$ | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 283.335 | 1.99 |
| | $s(.)$ | -1 | -1 | -1 | -1 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | | |
| 2,1,3 | $S(.)$ | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 279.632 | 3.27 |
| | $s(.)$ | -2 | -2 | -2 | -2 | -2 | -3 | -3 | -3 | -3 | -3 | -3 | | |
| 1,1,4 | $S(.)$ | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 275.789 | 4.60 |
| | $s(.)$ | -3 | -3 | -3 | -3 | -3 | -3 | -4 | -4 | -4 | -4 | -4 | | |
| 0,0,6 | $S(.)$ | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 269.454 | 6.79 |
| | $s(.)$ | -4 | -4 | -4 | -4 | -4 | -4 | -4 | -4 | -5 | -5 | -5 | | |

postponed. This means that we have a chance to use the advantage of price discount only at the current period. When $P = 1$, all the customers accept the discount when offered in any period. Therefore more improvement in costs can be expected compared to the $P = 0$ case, where the order up to level is also decreasing with a reduction of the reorder level.

In this section, we offer the same discount amount to all customers regardless of their postponement period. The costs of each case is compared with the case of no discount offerings. We measure the impact of postponement by calculating the value of $\delta_P = 100(Cost_1 - Cost_o)/Cost_o$, where $Cost_1$ is the cost when implementing the postponement policy, and $Cost_o$ is the cost of the system at the same level of ADI, but with no postponement policy introduced. Results are summarised in Table 4 for $P = 0$, and in Table 5 for $P = 1$.

As can be observed in Table 4, even with $P = 0$, there still exists a slight improvement on average inventory costs up to 1% based on the demand available to accept the postponement at the current period. The order up to level, $S$ is not changing in this case while the reorder level $s$ is decreasing when more demand is available for discount. When the system would start without having made any postponement offers in the past, then $Q_0 = 0$ and there is no benefit from the postponement policy when $P = 0$. When we consider the case $P = 1$, the reorder level is also decreasing, but now also the order up to level is decreasing. This means that additional

Table 4: $K = 100, h = 1, p = 9, d = 1, P = 0, \lambda = 4, 1, 1,\ D_{t-1,t+1} \in \{0, ...10\}$, $T = 12$

| $Q_0$ | | $D_{t-1,t+1}$ 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Avg. Cost | $\delta_P(\%)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | $S(.)$ | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 286.896 | 0 |
| | $s(.)$ | 0 | 0 | 0 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | | |
| 1 | $S(.)$ | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 286.288 | 0.21 |
| | $s(.)$ | -1 | -1 | -1 | -1 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | | |
| | $W_2(.)$ | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | | |
| | $w_2(.)$ | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | | |
| | $W_1(.)$ | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | | |
| 2 | $S(.)$ | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 285.711 | 0.41 |
| | $s(.)$ | -2 | -2 | -2 | -2 | -2 | -3 | -3 | -3 | -3 | -3 | -3 | | |
| | $W_2(.)$ | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | | |
| | $w_2(.)$ | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | | |
| | $W_1(.)$ | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | | |
| 3 | $S(.)$ | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 285.163 | 0.60 |
| | $s(.)$ | -3 | -3 | -3 | -3 | -3 | -3 | -4 | -4 | -4 | -4 | -4 | | |
| | $W_2(.)$ | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | | |
| | $w_2(.)$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | | |
| | $W_1(.)$ | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | | |
| 4 | $S(.)$ | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 284.634 | 0.79 |
| | $s(.)$ | -4 | -4 | -4 | -4 | -4 | -4 | -4 | -5 | -5 | -5 | -5 | | |
| | $W_2(.)$ | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | | |
| | $w_2(.)$ | 2 | 3 | 4 | 5 | 6 | 8 | 9 | 10 | 11 | 12 | 13 | | |
| | $W_1(.)$ | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | | |
| 5 | $S(.)$ | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 284.124 | 0.97 |
| | $s(.)$ | -5 | -5 | -5 | -5 | -5 | -5 | -5 | -5 | -6 | -6 | -6 | | |
| | $W_2(.)$ | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | | |
| | $w_2(.)$ | 1 | 2 | 3 | 4 | 5 | 7 | 8 | 9 | 10 | 11 | 12 | | |
| | $W_1(.)$ | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | | |

to ordering later, we also order less which reduces the holding cost as well. This can lead to a significant reduction on total expected cost, which is around 17% in our experiments. One may expect that the benefit will be higher when holding and setup costs are higher.

It is clear that more ADI increases the system performance as seen in Table 3. However, when we compare it with the results in Tables 4 and 5, better potential can be seen when using demand postponement. Our intuition supports this because ADI is available for all customers. On the other hand, with the discount policy, we only buy the ADI when we need it. Especially for higher setup costs, the system already needs to keep large stock most of the time. So, the uncertainty on demand does not effect the system much. But at the times when the stock is at critical levels,

Table 5: $K = 100, h = 1, p = 9, d = 1, P = 1, \lambda = 4, 1, 1$, $D_{t-1,t+1} \in \{0, ...10\}$, $T = 12$

| $Q_0$ | | $D_{t-1,t+1}$ 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Avg. Cost | $\delta_P(\%)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | $S(.)$ | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 239.296 | 16.59 |
| | $s(.)$ | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 1 | $S(.)$ | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 238.876 | 16.74 |
| | $s(.)$ | 0 | 0 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | | |
| | $W_2(.)$ | 6 | 7 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | | |
| | $w_2(.)$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 10 | 11 | 12 | | |
| | $W_1(.)$ | | 2 | 3 | 4 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | | |
| 2 | $S(.)$ | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 238.468 | 16.88 |
| | $s(.)$ | -1 | -1 | -1 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | | |
| | $W_2(.)$ | 6 | 7 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | | |
| | $w_2(.)$ | 0 | 1 | 2 | 3 | 4 | 5 | 7 | 8 | 9 | 10 | 11 | | |
| | $W_1(.)$ | | 2 | 3 | 4 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | | |
| 3 | $S(.)$ | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 238.081 | 17.02 |
| | $s(.)$ | -2 | -2 | -2 | -2 | -3 | -3 | -3 | -3 | -3 | -3 | -3 | | |
| | $W_2(.)$ | 6 | 7 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | | |
| | $w_2(.)$ | -1 | 0 | 1 | 2 | 3 | 4 | 6 | 7 | 8 | 9 | 10 | | |
| | $W_1(.)$ | | 2 | 3 | 4 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | | |
| 4 | $S(.)$ | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 237.713 | 17.14 |
| | $s(.)$ | -3 | -3 | -3 | -3 | -3 | -4 | -4 | -4 | -4 | -4 | -4 | | |
| | $W_2(.)$ | 6 | 7 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | | |
| | $w_2(.)$ | -2 | -1 | 0 | 1 | 2 | 3 | 5 | 6 | 7 | 8 | 9 | | |
| | $W_1(.)$ | | 2 | 3 | 4 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | | |
| 5 | $S(.)$ | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 236.066 | 17.72 |
| | $s(.)$ | -4 | -4 | -4 | -4 | -4 | -4 | -5 | -5 | -5 | -5 | -5 | | |
| | $W_2(.)$ | 6 | 7 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | | |
| | $w_2(.)$ | -3 | -2 | -1 | 0 | 1 | 2 | 4 | 5 | 6 | 7 | 8 | | |
| | $W_1(.)$ | | 2 | 3 | 4 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | | |

obtaining ADI at that time will be greatly beneficial. This can be achieved by our postponement policy.

### *6.2. Capacitated Inventory*

#### *6.2.1. The value of ADI and postponement*

In this section, we test ADI and the postponement policy on capacitated inventory systems. The literature on ADI mentions that ADI increases the efficient use of capacity. Similar to this, we expect that our postponement policy is going to increase the efficient usage of capacity. A particular challenge when having high levels of ADI is that postponement can only be offered to demand occurring within the protection period. Increasing levels of ADI in the system thus means less demand available for offering discount and thus exploiting the postponements policy. With the numerical experiments, we aim demonstrate this issue quantitatively and thus get a better insight into the effectiveness of the postponement policy.

In Table 6, we report the optimal replenishment and discount policy for a capacitated case. Without a capacity consideration, we should expect a reduction of reorder levels when $O_t$ is higher. However, in the capacitated system we observe an increase of reorder level when the $O_t$ approaches 10. This suggests that if higher demand is waiting for the next period, it is better to order earlier. Also, when compared to the uncapacitated case, the usage of the discount starts at higher inventory positions. This indicates that postponement is more needed in the capacitated case.

Özer and Wei (2004) show that a threshold policy which orders at full capacity if the inventory position drops to a threshold level, is optimal for the capacitated ADI case (without the postponement policy). Our findings do not verify this policy to be optimal when the discount/postponement policy is in use. Our intuition also supports the idea that if more demand is available for discount, then the order up to level will be lower than the capacity level.

Table 6: $K = 100, h = 1, p = 9, d = (1, 1.5), P = 0.5, \lambda = 4, 1, 1,\ D_{t-1,t+1} \in \{0, ...10\}, Q_t \in \{0, ..., 2\}$,$C = 20$

| $Q_0$ | | $D_{t-1,t+1}$ 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | $S(.)$ | 17 | 18 | 19 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| | $s(.)$ | 1 | 1 | 1 | 0 | 0 | 0 | -1 | -2 | -2 | -2 | -1 |
| 1 | $S(.)$ | 17 | 18 | 19 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| | $s(.)$ | 0 | 0 | 0 | 0 | -1 | -1 | -2 | -3 | -3 | -2 | -1 |
| | $W_2(.)$ | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| | $w_2(.)$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | $W_1(.)$ | 2 | 3 | 4 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| 2 | $S(.)$ | 17 | 18 | 19 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| | $s(.)$ | -1 | -1 | -1 | -1 | -2 | -2 | -3 | -4 | -3 | -2 | -1 |
| | $W_2(.)$ | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| | $w_2(.)$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| | $W_1(.)$ | 2 | 3 | 4 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |

Next, we address the benefits of ADI and demand postponement and compare them for different cases. The computational results are presented in Table 7.

The cost is increasing with the capacity constraint as expected (moving to the left in Table 7). It is also intuitive that the system can be operated at lower total costs when customers either buy with increase levels of ADI (moving down the rows in the table), or from using our demand postponement policy with increased levels of acceptance (columns with higher P values). But the surprising result exists when we compare the effectiveness of the demand postponement strategy

Table 7: Average cost of ADI and Postponement, $T = 12, K = 100$, $h = 1, p = 9, d_i = (1, 1.5)$

| | Capacity=15 | | | Capacity=20 | | | Uncapacitated |
|---|---|---|---|---|---|---|---|
| Demand | $P = 0$ | $P = 0.5$ | $P = 1$ | $P = 0$ | $P = 0.5$ | $P = 1$ | $P = 0$ |
| $6, 0, 0$ | 449.7 | 368.0 | 310.0 | 378.5 | 318.8 | 286.5 | 289.1 |
| $4, 1, 1$ | 407.0 | 360.6 | 300.6 | 353.3 | 314.8 | 281.1 | 286.9 |
| $3, 1, 2$ | 384.9 | 338.9 | 299.2 | 335.7 | 301.6 | 278.8 | 283.3 |
| $2, 1, 3$ | 366.9 | 345.7 | 299.5 | 321.0 | 305.3 | 278.3 | 279.6 |
| $1, 1, 4$ | 352.0 | 325.4 | 300.3 | 308.2 | 294.4 | 279.6 | 275.8 |
| $0, 0, 6$ | 328.2 | 328.2 | 328.2 | 292.8 | 292.8 | 292.8 | 269.5 |

relative to the uncapacitated case. When $P = 1$, our policy results in relatively small costs compared to the unconstrained case with no discount policy (the last column). This clearly shows that if a company can convince customers to accept the postponement, then they may be able to completely remove the disadvantage of the capacity constraint.

Increased levels of ADI also help to reduce the negative impact of capacity, as can be observed by comparing the difference rows in Table 7. However, when the demand pattern reaches a certain ADI, more ADI may not bring more benefit when having a postponement policy in place. In Table 7, when $P = 0.5$ and $C = 20$, the demand pattern $(3, 1, 2)$ has better cost than, for example, pattern $(2, 1, 3)$. The reason why this is possible is that the firm can purchase customers' willingness to wait when it is most needed, but not has to give these discounts at times where it is not needed. In contrast, and loosely speaking, increased levels of ADI only reduce the stress on the (capacitated) system, but offers no escape valve if the system runs hot. Note that this comparison between the postponement strategy and increased levels of ADI is in real life likely favouring the postponement policy even more since in our experiments we do account for the cost of the discount to customers accepting postponement, while the increased ADI $(2, 1, 3)$ coming from $(3, 1, 2)$ carries no extra costs in our computational experiment, while in real life this would come from offering incentives to all customers.

*6.2.2. Sensitivity Analysis on Discount Amount*

In this section, we observe the system improvements under different discount values. Since the postponement could bring significant improvements to the system, then the customers need to be encouraged to accept it. In reality, it is expected that the customers' willingness increases by getting more discount. We test the sensitivity of discount amount with different advance demand structures in the capacitated case.

Since there are two postponement periods available, we have two different discounts $d_1$ and $d_2$. We assume that $d_2$ is always equal to $d_1 + 0.5$ and we measure the impact of postponement by calculating the value of $\delta_P$. The Y axis of the figures in this section represents the $\delta$ value to show the effectiveness of the postponement decision. The X axis represents the discount value offered to customers in return for postponement. The tests are taken with a customer acceptance of $P = 0.5$, the backorder cost is £9, and the results are illustrated in Figure 4 for various advance demand patterns.

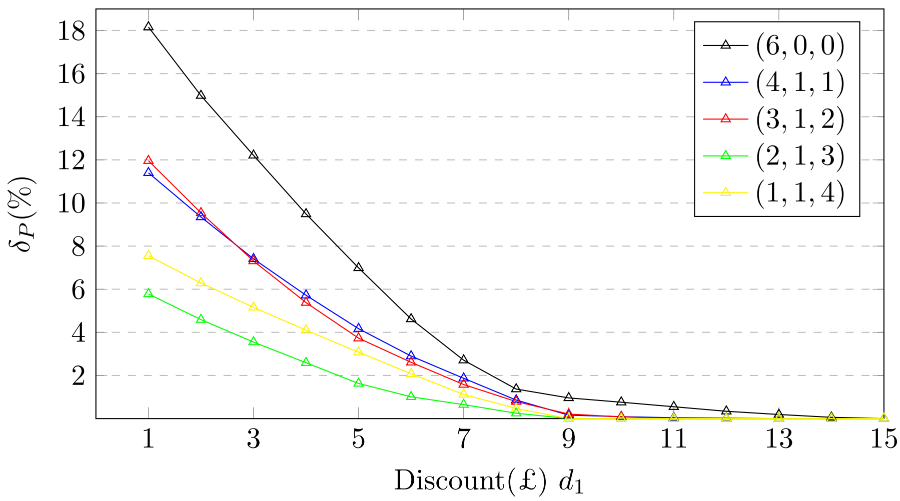


Figure 4: Sensitivity of Discount under Various ADI patterns ($d_2 = d_1 + 0.5$), $C = 15$

Although in previous experiments in Section 6.2, we have set the discount prices to $d_1 =$ £1 and $d_2 =$ £1.5, the results in Figure 4 show that the postponement strategy would remain profitable for higher values as well, but that this is quite sensitive to the level of ADI in the system. The lower the ADI level, the more effective the postponement policy. In the case $(6, 0, 0)$ (no ADI in the system), customers could receive £5 for postponement but still the firm can make more than 6% improvement. Even when the discount offered is £9, which is the same value as the backorder cost, the firm would still be able to make a small 1% additional saving. This can be understood from the fact that we can postpone either one or two periods, while a backorder is only postponed one period. Furthermore, we can postpone even if we have stock on hand.

Figure 4 also shows clearly that starting from a larger ADI in system, the usage and benefit from postponement reduces. Also, too large a discount will result in no benefit using the postponement policy, for whatever the value of ADI in the system. In this case, a discount of £13, when the backorder cost is only £9, is too costly to be used. When the company thinks of offering the discount, the balance between the level of the backorder cost and the discount amount has to be carefully considered. If a discount amount is deemed profitable, then the magnitude of benefits will depend on the level of ADI in the system. Most of the benefit can be expected for systems in which customers are used to fast deliveries.

Next, we analyse the effect of different prices for different postponement periods under a capacitated system ($C = 15$). Therefore, we set $d_1$ to £4 first, and observe the changes on improvement by different values of $d_2$ with $P = 0.5, 1$. Next, the same process is repeated to analyse the effect of $d_1$. The results are presented in Figure 5 and Figure 6, respectively.

Let us first consider the case that $d_1 = 4$. Then only the range in which $d_2 \geq 4$ becomes

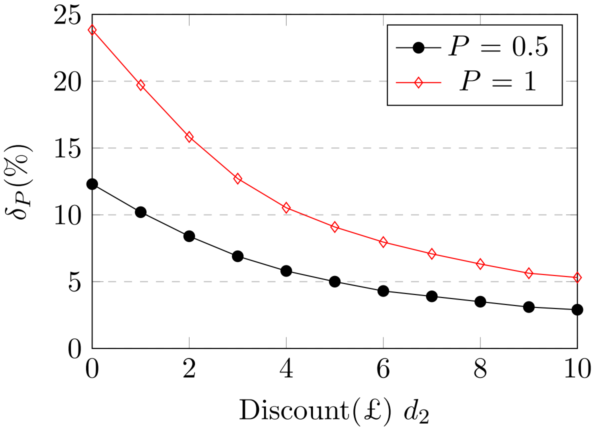


Figure 5: Analysis of $d_2$ when $d_1 = 4, \lambda = (3, 1, 2)$

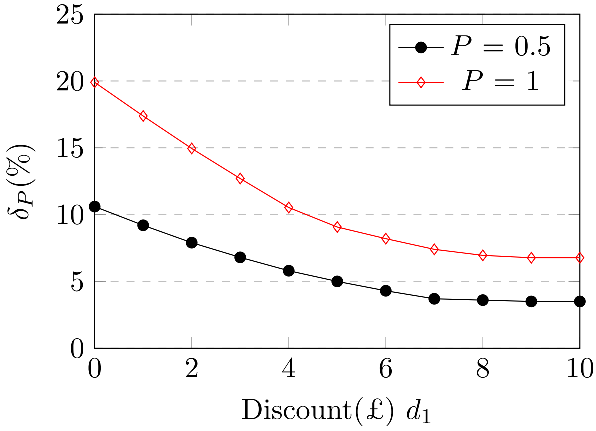


Figure 6: Analysis of $d_1$ when $d_2 = 4, \lambda = (3, 1, 2)$

realistic. Figure 5 illustrates that choosing the right value of $d_2$ would now involve estimating the correct value of $P$, and how this may be affected by the value of $d_2$. No matter the situation, we can observe from the figure that the firm can at most save 10% in costs, when $d_2 = 4$ and $P = 1$. But perhaps at $d_2 = 4$, the firm would only operate at $P = 0.5$. It may be that an acceptance close to $P = 1$ is only realised for $d_2 = 6$. This would then be a better choice, still, despite the higher discount value.

In Figure 6, we have instead fixed $d_2 = 4$. If this is a realistic value, then we should expect that $d_1 < 4$ is the range to further investigate. The maximum possible saving now would be close to 20%, but again the firm would need to estimate how the acceptance level would vary with $d_1$. Of course in practise one may not know how to set either values $d_1$ and $d_2$, and the challenge is to try to find out how P would be a function of both simultaneously.

Both figures show that in general, the sensitivity of profits to the acceptance probabilities is the largest when discount values remain small. In practise, these sensitivities are probably best discovered from real life experiments. Once these probabilities are known in function of the discount prices (illustrated in Section 6.2.3), models such as the one we developed can be used to determine when to invoke the postponement policy as to maximise the firm's overall profits.

*6.2.3. Discount-dependent customer acceptance*

So far in this paper, the rate of accepting delays was considered independent of the proposed discount. However, with the increase in the discount rate in real life, it will be easier for customers to be persuaded. Alım and Beullens (2020) presented numerical analysis based on an exponential relationship between the discount offered and the acceptance of a flexible delivery plan. Similarly, we perform some numerical tests assuming a connection with a similar formulation as $P(d) = e^{-\theta/d}$, where $\theta$ denotes the *customer resistance to discount*. With an increase in $\theta$, less customers are prone to accept a postponement, in particular noticeable at lower values of the discount. We ran numerical experiments with dependent acceptance probability $P(d)$ in which the same discount amount was offered for all delays ($d = d_1 = d_2$). The model was tested with different values $\theta$ and the percentage saving ($\delta_p$) comparing to the no discount case is measured. The results under the given parameters for a capacitated case ($C = 15$) are presented in Fig. 7.

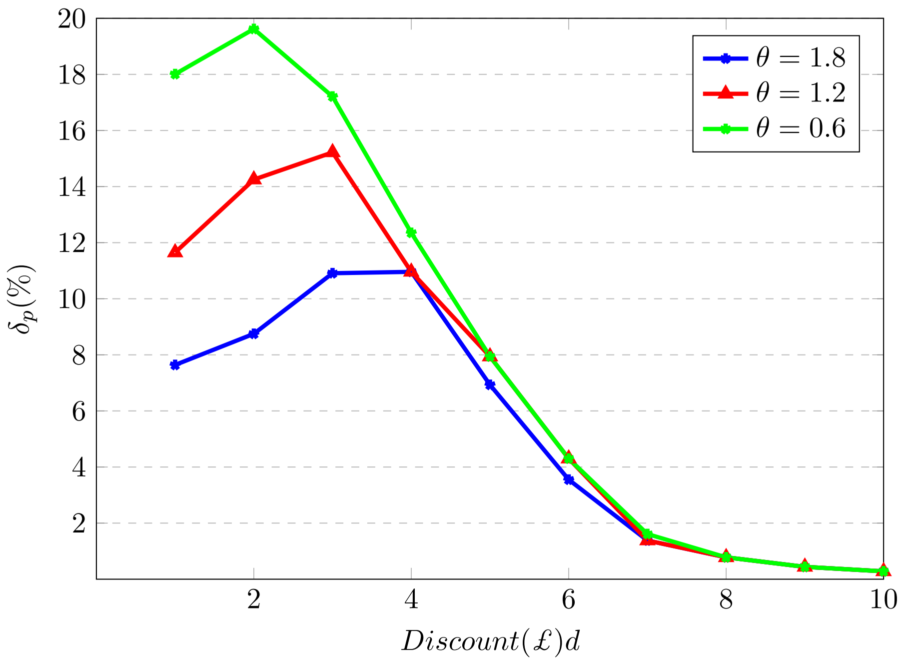


Figure 7: Results of dependent customer acceptance to discount, $d = d_1 = d_2$, $\lambda = (6, 0, 0)$

As can be seen in Fig. 7, the percentage of saving is relatively large at low amount of discounts for different $\theta$ values. Depending on the customers' resistance, the price discount that provides the best profit varies. For example, best profit can be obtained by offering $d = 2$ for $\theta = 0.6$ while it is $d = 3$ for $\theta = 1.2$ and $d = 4$ for $\theta = 1.8$. Offering a higher discount than these values does not improve the profitability. Also the difference on saving is getting smaller for higher discount amounts and identical at $d = 8$ or higher[3].

Overall, it is quite realistic that the customers' intention towards postponement can be changed by the discount amount offered. The results in Fig. 7 indicate that the companies that wish to implement such a postponement policy have to correctly define the relationship between customer acceptance rate and discount amount.

[3]Some of the data points at lower values of $d$ also coincide, as is the case for $d = 5$ and $\theta = 0.6$ and $\theta = 1.2$. This is because of the small mean demand rate and rounding. It is important to note that the impacts of small shifts between discount amount and customer acceptance will be more visible in case of having larger demand means. At higher demand rates, smoother curves can also be observed.

## 7. Conclusions

In this paper, we establish the structure of inventory and discount policies for an inventory model with advance demand information and price discounts. In return for discounts, customers are expected to wait longer for their demand to be delivered. Discount offers are made selectively and only in certain periods.

For the zero setup cost case, and without considering the discount system, a simple base stock policy is proven to be optimal by Gallego and Özer (2001) for the ADI case. Based on the numerical results, we observe that this policy does not change by the discount offer. This is in line of expectation since the ordering decision is free of charge, there is no need to use discounts for postponement. When in the ADI, without the discount offers, the setup cost is positive, then state- dependent policies are optimal. State-dependent policies are also found in our numerical experiments, adhering to a policy structure as indicated in Figure 3, but without a formal proof of its optimality. The parameters of this policy depend on the observed demand beyond the protection period and the available demand for the postponement/discount decision. Numerical examples indicate that the system cost is reducing when more advance demand and demand for discounts are present. Increasing customers acceptance of discounts results in reducing reorder and order up to levels. If firms have knowledge about how customer acceptance of postponement depends on the discount offered, an optimal discount amount can be offered within the state-dependent policy.

To our knowledge, this is the first study to develop and analyse a selective state-dependent strategy to purchasing buyer's willingness to postpone their delivery in a variety of ADI settings. This strategy seems in particular of value in capacitated inventory situations with a modest level of ADI. It is argued that this strategy can result in a more profitable system in comparison to strategies that aim to introduce more ADI in the system.

With the rise of online sales channels, companies have the ability to contact customers almost instantaneously, and to make bespoke offers based on the current state of the firm's inventory system. The discount policy proposed in this study can improve the firm's existing stock systems by offering selective discounts for postponing the delivery for some customers, should they wish to accept this offer. According to our numerical results, it was observed that the demand postponement policy can make the system more efficient especially in areas where capacity is limited. The customers who are willing to wait are rewarded, while customers not willing to do so can still opt for fast delivery. At the same time, the company can make more profit from the existing stock system more effectively. In addition, instead of the negative impact of backorders which reduce the company's reputation among the customers, the mutually agreed postponement policy at which both parties get benefits may improve the company's reputation.

Offering flexible delivery terms at various price points may offer a strong competitive advantage, as briefly discussed in the introduction, and may form a rich area for further research. Further research could be devoted to better understanding the link between discount price and customer acceptance rates, and the impact this has on system performance and policy with higher demand rates. In addition, it may also pay off in certain applications to include additional decisions with respect to early deliveries. A practical drawback from this approach is the computational complexity of the model and inventory/postponement policy. Further research may thus also be devoted to faster heuristic methods, and examining the efficiency of policies with a simpler structure.

## References


Alım, M. and Beullens, P. (2020). Joint inventory and distribution strategy for online sales with a flexible delivery option. *International Journal of Production Economics*, 222:107487.

Bao, L., Liu, Z., Yu, Y., and Zhang, W. (2018). On the decomposition property for a dynamic inventory rationing problem with multiple demand classes and backorder. *European Journal of Operational Research*, 265(1):99–106.

Benjaafar, S., Cooper, W., and Mardan, S. (2011). Production-inventory systems with imperfect advance demand information and updating. *Naval Research Logistics*, 58(2):88–106.

Buhayenko, V. and van Eikenhorst, E. (2015). Synchronisation of customers orders using discounts. *International Journal of Systems Science: Operations & Logistics*, pages 1–15.

Chen, F. (2001). Market segmentation, advance demand information, and supply chain performance. *Manufacturing and Service Operations Management*, 34(1):482–499.

Du, B. and Larsen, C. (2017). Reservation policies of advance orders in the presence of multiple demand classes. *European Journal of Operational Research*, 256(2):430–438.

Gallego, G. and Özalp, Ö.. (2002). *Supply Chain Structures; Coordination, Information and Optimization*, chapter Optimal Use of Advance Demand Information in Supply Chain Management, pages 119–160. Kluwer Academic Publisher.

Gallego, G. and Özer, Ö.. (2001). Integrating replenishment decisions with advance demand information. *Management Science*, 47(10):1344–1360.

Gallego, G. and Özer, Ö.. (2003). Optimal replenishment policies for multiechelon inventory problems under advance demand information. *Manufacturing & Service Operations Management*, pages 157–175.

Gayon, J., Benjaafar, S., and de Vericourt, F. (2009). Using imperfect advance demand information in production-inventory systems with multiple customer classes. *Manufacturing & Service Operations Management*, 11(1):128–143.

Ghiami, Y. and Beullens, P. (2016). Planning for shortages? Net present value analysis for a deteriorating item with partial backlogging. *International Journal of Production Economics*, 178:1–11.

Guan, Y. and Liu, T. (2010). Stochastic lot-sizing problem with inventory-bounds and constant order-capacities. *European Journal of Operational Research*, 207(3):1398–1409.

Gutiérrez, J., Sedeño-Noda, A., Colebrook, M., and Sicilia, J. (2003). A new characterization for the dynamic lot size problem with bounded inventory. *Computers & Operations Research*, 30(3):383–395.

Hariharan, R. and Zipkin, P. (1995). Customer-order information, leadtimes, and inventories. *Management Science*, 41(10):1599–1607.

Karaesmen, F., Liberopoulos, G., and Dallery, Y. (2004). The value of advance demand information in production/inventory systems. *Annals of Operations Research*, 126(1-4):135–157.

Kremer, M. and Van Wassenhove, L. N. (2014). Willingness to pay for shifting inventory risk: The role of contractual form. *Production and Operations Management*, 23(2):239–252.

Kunnumkal, S. and Topaloglu, H. (2008). Price discounts in exchange for reduced customer demand variability and applications to advance demand information acquisition. *International Journal of Production Economics*, 111(2):543–561.

Li, C. and Zhang, F. (2013). Advance demand information, price discrimination, and preorder strategies. *Manufacturing & Service Operations Management*, 15(1):57–71.

Lu, Y., Song, J., and Yao, D. D. (2003). Order fill rate, leadtime variability, and advance demand information in an assemble-to-order system. *Operations Research*, 51(2):292–308.

Özer, Ö. (2003). Replenishment strategies for distribution systems under advance demand information. *Management Science*, 49(3):255–272.

Özer, Ö.. (2011). *Inventory Management: Information, Coordination, and Rationality*, pages 321–365. Springer.

Özer, Ö. and Wei, W. (2004). Inventory control with limited capacity and advance demand information. *Operations Research*, 52(6):988–1000.

Sarkar, S. and Shewchuk, J. (2013). Use of advance demand information in multi-stage production-inventory systems with multiple demand classes. *International Journal of Production Research*, 51(1):57–68.

Sarkar, S. and Shewchuk, J. P. (2016). Advance demand information and early fulfillment in inventory management with two demand classes. *International Journal of Logistics Systems and Management*, 25(2):175.

Silver, E. A., Pyke, D. F., and Peterson, R. (1998). *Inventory management and production planning and scheduling*, volume 3. Wiley New York.

Talluri, K. T. and Hillier, F. S. (2005). *Theory and Practice of Revenue Management.* Springer US.

Tan, T., Gll, R., and Erkip, N. (2007). Modelling imperfect advance demand information and analysis of optimal inventory policies. *European Journal of Operational Research*, 177(2):897–923.

Tang, C. S. (2006). Perspectives in supply chain risk management. *International Journal of Production Economics*, 103(2):451–488.

Van Wijk, A., Adan, I. J., and Van Houtum, G. (2019). Optimal lateral transshipment policies for a two location inventory problem with multiple demand classes. *European Journal of Operational Research*, 272(2):481–495.

Wang, T. and Toktay, B. (2008). Inventory management with advance demand information and flexible delivery. *Management Science*, 54(4):716–732.

Wijngaard, J. and Karaesmen, F. (2007). Advance demand information and a restricted production capacity: on the optimality of order base-stock policies. *OR Spectrum*, 29(4):643–660.

Xu, H., Gong, Y. Y., Chu, C., and Zhang, J. (2017). Dynamic lot-sizing models for retailers with online channels. *International Journal of Production Economics*, 183:171–184.