IISE Transactions

ISSN: 2472-5854 (Print) 2472-5862 (Online) Journal homepage: https://www.tandfonline.com/loi/uiie21

Using imperfect advance demand information in lost-sales inventory systems with the option of returning inventory

Engin Topan, Tarkan Tan, Geert-Jan van Houtum & Rommert Dekker

IISE TRANSACTIONS
2018, VOL. 50, NO. 3, 246–264
https://doi.org/10.1080/24725854.2018.1403060

Taylor & Francis
Taylor & Francis Group

Using imperfect advance demand information in lost-sales inventory systems with the option of returning inventory

Engin Topan[a,b], Tarkan Tan[a], Geert-Jan van Houtum[a] and Rommert Dekker[c]

[a]Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands; [b]Faculty of Behavioural Management and Social Sciences, University of Twente, Enschede, The Netherlands; [c]Erasmus School of Economics, Erasmus University, Rotterdam, The Netherlands

ABSTRACT
Motivated by real-life applications, we consider an inventory system where it is possible to collect information about the quantity and timing of future demand in advance. However, this Advance Demand Information (ADI) is imperfect as (i) it may turn out to be false; (ii) a time interval is provided for the demand occurrences rather than its exact time; and (iii) there are still customer demand occurrences for which ADI cannot be provided. To make best use of imperfect information and integrate it with inventory supply decisions, we allow for returning excess stock built up due to imperfections to the upstream supplier and we propose a lost-sales inventory model with a general representation of imperfect ADI. A partial characterization of the optimal ordering and return policy is provided. Through an extensive numerical study, we investigate the value of ADI and factors that affect that value. We show that using imperfect ADI can yield substantial savings, the amount of savings being sensitive to the quality of information; the benefit of the ADI increases considerably if the excess stock can be returned. We apply our model to a spare parts case. The value of imperfect ADI turns out to be significant.

ARTICLE HISTORY
Received 28 September 2016
Accepted 14 October 2017

KEYWORDS
Value of information; imperfect advance demand information; returning excess inventory to upstream; lost-sales; inventory

1. Introduction

Developments in information technology have given rise to applications of Advance Demand Information (ADI) in inventory planning. The research in this field has also benefited from these developments and has gained significant momentum in the last two decades. Nevertheless, some important practical aspects of ADI have not yet been addressed, to the best of our knowledge. First, most papers assume that ADI is perfect. Second, the papers that consider the imperfect nature of ADI do not categorize and address the different types of imperfection. Third, clearing mechanisms, such as returning or selling excess inventory built up due to imperfect ADI to upstream suppliers, have not been addressed. Fourth, the papers are based on the assumption that unmet demand is backordered; the settings in which unmet demand is lost or satisfied by an emergency supply source have been considered to only a limited extent.

The primary motivation behind this article is our experience with ASML, a world leading original equipment manufacturer that produces lithography systems, which are critical for the production of integrated circuits for the semiconductor industry. These systems are sold with service contracts, known as Service-Level Agreements (SLAs). Through these SLAs, a certain system availability level is committed to customers, making ASML responsible for all maintenance and service activities. This imposes either a tight availability target for spare parts or a high (explicit or implicit) down time cost associated with violation of the targets. To reach the service targets or not to incur high down time costs, ASML stores parts at local warehouses close to its customers. However, spare parts of such systems are often slow moving and expensive; therefore, ASML likes to keep the stock of spare parts as low as possible, without jeopardizing its availability commitment. This can be achieved to a certain extent by employing condition monitoring. Over the past few years, ASML has been using condition monitoring for critical machine components installed at customer sites, by means of various sensors mounted on components. These sensors continuously monitor numerous condition indicators of components such as vibration, temperature, pressure, acoustic data, etc. The data are analyzed through a number of detailed data mining steps, mainly data collection, pre-processing, predictive modeling, and analysis. The main idea of the entire process is to extract useful information from the available data and then use it to predict failures in advance, often by using a prediction technique (Olson and Delen, 2008). In this way, the system can issue a warning signal in advance of an actual failure. This signal can be considered as a "demand" signal (or ADI) for the corresponding spare parts, as failures generate demand for spare parts and this can be used to optimize the spare parts supply decisions. Nevertheless, demand signals that are produced by condition monitoring can be imperfect in three ways: (i) the prediction tool may produce false signals or so-called *false positives* (warnings without resulting failures); (ii) the exact *timing* of the failure is uncertain. In addition, if the information is late, even if it is certain, it may be completely useless. For example, if an advance

CONTACT Engin Topan e.topan@utwente.nl

demand cannot be satisfied by a regular order and it has to be satisfied by an emergency shipment anyway, this information has no use; (iii) the prediction tool may fail to produce warnings for failures; hence, there are *false negatives* (failures without warnings) that need to be considered. A component may have multiple failure modes and there might be failure modes that cannot be predicted in advance. The experience of ASML is not unique. Our observations are quite common for capital goods manufacturers that are investigating the use of condition monitoring and the imperfect ADI that it provides for spare parts planning.

There are other application areas of imperfect ADI having similar characteristics. For example, companies such as Toshiba, Dell, and Océ sell industrial computers and printers to business customers. They often operate based on a configure-to-order principle and therefore reserve critical resources such as production capacity and hold inventory for intermediate products and subassemblies. In this situation, any indication of future sales or any intended order becomes important in avoiding long lead times and preventing sales losses. However, a customer who announces her intention to purchase a certain product may not buy that product (false positives) or place her order at a time different than the indicated intention (uncertain time). In addition, there are customers who place their orders without any prior warning (false negatives). The use of ADI for spare parts inventory planning at repair shops is another example. When a repairable component/subsystem fails, a service engineer can often diagnose possible causes of the failure in the field and can identify which part of the component might cause the failure and need replacement. This information can be immediately available to the repair shop, which can be very useful in supplying the spare part in advance. However, whether this component can actually be repaired and, if so, which service part is really needed to complete the repair is only known when the component is disassembled at the repair shop just before an actual repair starts.

The use of imperfect ADI raises another issue that we observe at ASML. When a spare part is ordered and kept in stock for a signaled demand and when this turns out to be a false positive, the part becomes excess stock and the system starts incurring extra holding cost by keeping that part in stock. In this situation, it may be favorable to clear the excess inventory, even at the expense of some clearing cost. In practice, if the upstream echelon of the supply chain is operated by the same company, as in the ASML case, this leads to a return to an upstream echelon where the part can be better pooled; this will cost extra forward and back shipments and a holding cost for the time that the part is on the back shipment pipeline. If not, the item may be returned or sold back to an external supplier, possibly at a lower price, leading to a high return cost.

The backordered demand assumption facilitates a relatively simple analysis, which is also true for our case. Nevertheless, the service targets that are set by SLAs explicitly or implicitly impose high penalty or down time costs for each demand that is not satisfied from stock. Unmet demand cannot be backordered; instead, it is satisfied by an emergency shipment or an emergency provisioning mode or it is simply lost to a competitor. From a modeling perspective, this can be represented by a lost sales inventory model.

We characterize three types of imperfection in ADI: (i) ADI solely results in a demand with a certain probability p, reflecting the *precision* of ADI (*proportion of true positives to sum of true and false positives*); (ii) a time interval $[\tau_l, \tau_u]$ is provided for the demand occurrences rather than its exact time, representing the time uncertainty and timeliness; i.e., the *timing* issue of ADI; (iii) only a fraction q of demand can be signaled or predicted in this way, indicating the *sensitivity* (*proportion of true positives to sum of true positives and false negatives*) of ADI. Parameters p, τ_l, and τ_u can be obtained by using a prediction tool that can typically provide a confidence interval for the remaining life time of a component. When a confidence interval is available, its lower and upper limits and confidence level can be used to estimate τ_l, τ_u, and p, respectively. In this situation, τ_l corresponds to the earliest possible time that a failure can be predicted in advance, which is more or less known by the manufacturer operating such systems. Any failure before τ_l can be considered as an unpredicted failure (i.e., false negative). τ_l might also be zero, meaning that such a lower limit does not exist and a signal may arrive and the failure corresponding to the signal may occur successively in the same period. Similarly, τ_u has the interpretation in practice that a demand signal is typically ignored when it does not become true after a sufficiently long period of time. Estimating q is rather straightforward and can be made by looking at the ratio of predicted demand to total demand based on historical observations. Our imperfect ADI setting is quite general and it applies to the other two motivating examples that we have discussed above.

By considering a general representation of imperfect ADI, we build a single-item, single-location, periodic-review lost sales inventory model with a positive lead time where excess stock built up due to imperfections can be returned to an upstream supplier. The objective is to find the optimal ordering and return policy under imperfect ADI. Using our model, we study the following questions about the use of imperfect ADI and the benefit of return under imperfect ADI: How can the imperfect ADI be best used? What is the value of using this information? How is the value of ADI influenced by imperfections? How useful is returning excess stock in coping with the consequences of imperfectins? How can the optimal ordering and returning policy be characterized?

This article contributes to five main fields of research: value of (imperfect) ADI in inventory planning, inventory models with negative inventory flow (using our terminology, inventory systems with returns to upstream), lost sales inventory systems, use of condition monitoring in spare parts inventory planning, and inventory management with forecast updating. In most of the papers on the value of ADI, the information is perfect (e.g., Buzacott and Shanthikumar (1994), Hariharan and Zipkin (1995), Gallego and Özer (2001), Özer (2003), and Karaesmen (2013)). There are a few papers that do take the imperfect nature of ADI into consideration. Among these papers, Donselaar *et al.* (2001), Thonemann (2002), Tan *et al.* (2007, 2009), and Song and Zipkin (2012) study the use of imperfect ADI for inventory systems; Liberopoulos and Koukoumialos (2008) study the use of imperfect ADI for a capacitated production/inventory system; Gao *et al.* (2012) study the use of imperfect ADI for an assembly system; and Bernstein and DeCroix (2015) study the use of imperfect ADI for a multiproduct system in a single-period setting. These papers assume that unmet demand is backordered or there is a single period. Gayon *et al.* (2009) and Benjafaar *et al.* (2011) study the value of ADI under a multi-period lost sales setting (the latter as an extension). Both study the time and quantity uncertainty of imperfect ADI and consider a continuous-review model and assume at most one outstanding order and an exponentially distributed demand lead time, which are assumptions that do not hold for our setting. In this article, we consider a more general representation of imperfect ADI by also addressing the timing of ADI, whether the information is available before the time to make the ordering decision.

This article also contributes to inventory models with disposal (Fukuda, 1961) and to the stochastic cash balance problem (Eppen and Fama, 1969); i.e., inventory models that allow negative inventory flow. Although these models gained a fair amount of attention in the past, they have surprisingly received little attention in recent years. As we experience in practice, an inventory system operating under an imperfect ADI setting may benefit from both negative and positive flows. However, this subject has not been thoroughly addressed. To our knowledge, Song and Zipkin (2012) is the only paper that considers both ADI and returning excess inventory to an upstream supplier. They consider a newsboy setting with return possibility where a procurement decision is made only at a single procurement epoch while canceling excess inventory is possible when some partial ADI is revealed. In contrast, we consider a multi-period problem where both procurement and return decisions can be made at each period. In this sense, this article is the first to consider returning excess inventories for a general (in)finite-horizon inventory model with ADI.

The analysis of lost sales inventory systems is more difficult than that of backorder systems, as the optimal inventory policy depends on the number of outstanding replenishment orders and on-hand inventory, and the state space grows very rapidly, which is also true for inventory systems with ADI. Papers on structural analysis of the optimal policy for lost-sales systems are rare (Karlin and Scarf, 1956; Morton, 1969; Zipkin, 2008b). Most of the papers in this field propose useful heuristics (Morton, 1971; Johansen, 2001; Zipkin, 2008a; Bijvank and Vis, 2011). These heuristics are often based on myopic policies, base stock policies, and their variations. More recently, Zipkin (2008b) provided a new approach for the structural analysis of lost-sales models by applying a state transformation and using the notion of $L^\natural$-convexity, a property implying both convexity and submodularity. This considerably simplifies the analysis. Zipkin does not consider imperfect ADI and returning excess inventory to an upstream supplier; however, he provides an extension for a Markov-modulated demand process. Unlike Zipkin (2008b), the number of demand signals in this article changes due to demand realizations; therefore, we do not have a Markov-modulated demand process. Consequently, the structural analysis in Zipkin (2008b) is not directly applicable to our model. Hence, we propose a different state transformation making it possible to use $L^\natural$-convexity. To the best of our knowledge, this article is the first to characterize the optimal ordering and return policy for a periodic-review lost sales inventory system with imperfect ADI.

The use of condition monitoring in maintenance optimization has been extensively studied in the literature (Elwany and Gebraeel, 2008). However, studies on the consequences of using condition monitoring in spare parts inventory planning are rare (Deshpande *et al.*, 2006; Li and Ryan, 2011; Louit *et al.*, 2011; Lin *et al.*, 2017) and all assume perfect information. To our knowledge, this article is the first to investigate the imperfect nature of the information provided by condition monitoring and the consequences of using it in the optimal control of spare parts inventories. Therefore, we also contribute to the vast literature on spare parts inventory systems (Muckstadt, 2005).

Inventory management with ADI has similarities with inventory management with forecast updating, as the demand is formulated as a function of information that changes with time, using probabilistic models in both (Hausman, 1969; Heath and Jackson, 1994; Güllü, 1996; Toktay and Wein, 2001; Zhu and Thonemann, 2004; Wang and Tomlin, 2009). However, this article differs from this stream in four ways:

1. The inventory management with forecast updating is based on point estimation of demand realization and using it as an input in decision making. In contrast, we incorporate information directly in decision making by using ADI, which is the reason why there is a different stream of literature on ADI.
2. Consequently, each predicted demand realization is coupled with an ADI; therefore, the demand realizations affect the number of active demand signals in the system, which is not necessarily the case in forecast models.
3. Our ADI model has a unique characteristic that captures the uncertainties concerning the materialization of demand signals; e.g., timing and likelihood. Therefore, how long an individual piece of information remains in the system and how this affects the demand realizations in future periods is uncertain, which is not captured by forecast models.
4. To our best knowledge, returning excess stock to an upstream supplier (would then be due to changes in forecast update) has not been considered in the papers on inventory management with forecast updating.

The main contributions of this article are thus as follows:

1. By categorizing the types of imperfection of ADI and addressing all at the same time, we consider a general representation of imperfect ADI that can be used to model a wide range of ADI applications in practice. We assume a general probability distribution for the interarrival time between signals (ADI) and the demand lead time; we do not have any restriction on the size of outstanding orders; in addition, we make return decisions (in addition to ordering), all of which are in line with our observations in practice. With this model, we provide a methodological recipe for companies on how they can use imperfect ADI to plan their inventory supplies.
2. We propose a state transformation under which the cost-to-go function is proven to be $L^\natural$-convex for given numbers of demand signals from multiple periods. We derive a number of structural monotonicity properties of the optimal ordering and return policy with respect to inventory levels by using $L^\natural$-convexity. Our findings indicate that the optimal policy has a quite complex, state-dependent structure: The optimal policy is dependent not only on on-hand stock but also on pipeline stock. We further show that optimal order (return) size and inventory levels are economic substitutes

(complements). Finally, base stock policies and myopic policies, which are commonly used in practice, are not necessarily optimal and they may yield poor performance.

3. We generate useful managerial insights that can be used as input in design and improvement of inventory systems with imperfect ADI. The most important observations among all are that the timing of ADI is highly influential on the value of ADI and returning excess inventory is quite effective in coping with consequences of false ADI.

The rest of this article is organized as follows. In Section 2, we present our model. In Section 3, we characterize the structural properties of the optimal policy. In Section 4, we provide our numerical results and the spare parts case. Finally, in Section 5, we draw conclusions.

2. The model

We consider a single-item, single-location, periodic-review inventory system. An information collection mechanism makes it possible to issue a demand signal (or ADI) indicating that a demand is likely. Time is divided into periods, which are indexed by $t = 1, 2, \ldots, T$. Time horizon T can be finite or infinite as $T \to \infty$. For simplicity, we use generic variables that are defined for each $t = 1, 2, \ldots, T$ as long as it makes sense. The number of demand signals that are (collected during period $t-1$ but) first available in the system at the beginning of period t is denoted by the generic random variable W, which can follow any probability distribution. A demand signal that is first available at the beginning of period t (i) either turns out to be true and materializes as an actual demand in period $t+\tau$ with probability $p_\tau > 0$ for $\tau \in \{\tau_l, \ldots, \tau_u\}$ and $p_\tau = 0$ otherwise, where $\tau_l, \tau_u \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and $\tau_u \geq \tau_l$, or (ii) it eventually leaves the system as a false positive at the beginning of period $t + \tau_u + 1$. In this setting, τ, which is the delay between when a demand signal arrives and when it becomes a demand realization or leaves the system without becoming a demand realization, corresponds to the demand lead time (Hariharan and Zipkin, 1995); with one exception: we limit the definition to demand lead times of true demand signals since we also have false positives); $[\tau_l, \tau_u]$ is the prediction interval for the demand lead time; and $p = \sum_{\tau=\tau_l}^{\tau_u} p_\tau \leq 1$ is the probability that a demand signal will ever become a demand realization, which we refer to as the *precision* of the signal. The demand type whose occurrence is prognosed and, hence, whose probability distribution depends on the accumulated demand signals is called the predicted demand. We assume that every signal corresponds to at most one demand and when a demand is realized it is known which demand signal it belongs to unless it is a false negative (demand without any prior warning). To formulate the dynamics for the flow of signals and the predicted demand for each period t, we define generic random variable A_τ as the number of demand signals that are in the system for exactly $\tau \in \{0, \ldots, \tau_u\}$ periods; this refers to signals that became available at the beginning of period $t-\tau$ and have not yet materialized. Then, R_τ denotes the number of demand signals of A_τ that materialized into an actual demand in period t. Letting a_τ be the realization of A_τ in period t, R_τ has a binomial distribution with parameters a_τ and $p_\tau/(1 - \sum_{k=\tau_l}^{\tau-1} p_k)$ for $\tau \in \{\tau_l, \ldots, \tau_u\}$ and it is zero for $\tau \in \{0, \ldots, \tau_l - 1\}$. Then, the total number of predicted demands in period t is given by $\sum_{\tau=\tau_l}^{\tau_u} R_\tau$.

Apart from predicted demand, there are unpredicted demand occurrences that cannot be signaled in advance. The unpredicted demand in period t is denoted by the generic random variable D^u, which can follow any probability distribution. We assume that the two demand types are independent. Since the consequences and the costs of these two demand types do not differ, they are treated equally and served based on the first-come first-served rule. As a result of the ADI setting explained above, the expected predicted demand per period for $\delta \leq \tau_u$ periods ahead from the present period depends on the number of demand signals that arrived at most $\tau_u - \delta$ periods earlier and have not yet materialized, i.e., $(a_0, \ldots, a_{\tau_u - \delta})$—and of course on the realizations of those signals. The expected predicted demand per period for $\delta > \tau_u$ periods ahead from the present period equals $pE[W]$ and, therefore, the expected total demand per period is expressed by $\lambda = pE[W] + E[D^u] > 0$. The ratio of expected predicted demand to expected total demand per period is denoted by a constant q where

$$q = \frac{pE[W]}{pE[W] + E[D^u]} \geq 0,$$

which we refer as to the *sensitivity* of the demand signal.

The demand for an item is immediately satisfied from stock if there is an available item in stock. The stock is replenished from an appropriate supplier within a constant (regular) replenishment lead time $L \in \mathbb{N}_0$ at a unit procurement cost $c\ (>0)$. When an item is requested but there is no available stock on hand, the demand is satisfied by an emergency supply source or it is lost. In this situation, a penalty cost $c_e\ (>0)$ is incurred per unit of unmet demand. In the context of spare parts demand from technical systems, this cost involves a cost for the emergency supply source and a downtime cost incurred during the emergency lead time (which is short compared with the length of the review period). In the general lost sales case for complex products, this cost involves loss of profit margin and goodwill. In each period t, the size of the regular replenishment order placed in period $t - L + l$ and due in period $t + l$ is denoted by generic variable z_l for $l = 0, \ldots, L$. A holding cost $h\ (>0)$ is incurred for each unit of inventory carried from one period to the next. An excess stock can be returned to the central warehouse or to the supplier at a per unit return cost c_r. In the case where a part is purchased back by the supplier, there might be a revenue associated with the return. Therefore, we allow a negative unit return cost c_r. Furthermore, we assume $c + c_r \geq h \times L$. This implies that it is cheaper to keep an item in stock (at the expense of holding one extra item during a lead time, $h \times L$) than to return that item and at the same time to place a new order (at the expense of return and procurement costs, $c + c_r$). Note that the assumption facilitates our analysis (see Lemma 1) and it also eliminates making speculative profit by returns. The on-hand inventory (before the arrival of a due order and return of excess stock) at the beginning of period t is denoted by generic variable $x\ (\geq 0)$. The size of the return made in period t is denoted by generic variable $y\ (\geq 0)$. We assume that the size of the return y cannot exceed the available stock x. For notational convenience, we assume that $T \geq \max(L, \tau_u)$. Realizations of random variables are denoted by lowercase letters. Our notation is summarized in Table 1.

Table 1. General notation.

T	End of planning horizon
t	Time index, $t = 0, \ldots, T$
λ	Expected total demand per period
p	Probability that a signal will ever become a demand realization (precision)
p_τ	Probability that a signal will become a demand realization in τ periods after its first arrival
q	Ratio of expected predicted demand to expected total demand (sensitivity)
L	(Regular) replenishment lead time
τ	Demand lead time
τ_l	Lower limit for the demand lead time
τ_u	Upper limit for the demand lead time
c	Unit procurement (ordering) cost
c_e	Unit penalty cost of an emergency supply
c_r	Unit return cost
h	Holding cost for each unit carried from one period to the next
D^u	Unpredicted demand per period
W	Number of demand signals collected per period
A_τ	Number of demand signals that are in the system for τ periods
R_τ	Number of demand signals of A_τ that materialize into an actual demand in period t
f_t	Optimal cost-to-go function from period t to the end of the planning horizon T
x	On-hand inventory (before the arrival of order due and the returning decision) in period t
z_l	Size of the replenishment order placed in period $t - L + l$ and due at $t + l$
y	Size of the return made in period t

The sequence of events in period t is as follows:

1. The signals (collected during $t-1$), W, are announced to the system and registered as a_0.
2. The replenishment order that will arrive in period $t+L$, z_L, and the size of the return y are determined. These orders are placed accordingly.
3. The replenishment order that has been placed at $t-L$ and due at t, z_0, arrives,
4. Both the predicted and unpredicted demands, $\sum_{\tau=\tau_l}^{\tau_u} r_\tau$ and d^u, respectively, are realized.
5. The procurement, inventory holding, and penalty costs are incurred accordingly.

For notational simplicity, we let $\mathbf{a} = (a_{\tau_u}, \ldots, a_0)$ and $\mathbf{z} = (x, z_0, \ldots, z_{L-1})$. Then, the system state is described by $(\mathbf{a}, \mathbf{z})$, and the state space by the Cartesian product of $\mathcal{U} = \{\mathbf{a} : \mathbf{a} \in \mathbb{N}_0^{\tau_u+1}\}$ and $\mathcal{Z} = \{\mathbf{z} : \mathbf{z} \in \mathbb{N}_0^{L+1}\}$. Our objective is to determine the order size z_L and the size of the return y that will minimize the total inventory holding, penalty, and return costs. Therefore, the action space is given by $\mathcal{A}_x = \{(z_L, y) : z_L, y \in \mathbb{N}_0, y \le x\}$. For a given state $(\mathbf{a}, \mathbf{z})$, let $f_t(\mathbf{a}, \mathbf{z})$ be the optimal cost-to-go (value) function from period t to the end of the planning horizon T. Then, for all $t = 1, \ldots, T$ the optimal cost-to-go function is given by the dynamic programming recursion

$$
\begin{aligned}
f_t(\mathbf{a}, \mathbf{z}) &= \min_{(z_L, y) \in \mathcal{A}_x} \{J_t(\mathbf{a}, \mathbf{z}, z_L, y)\}, \\
J_t(\mathbf{a}, \mathbf{z}, z_L, y) &= c z_L + c_r y + L(a_{\tau_u}, \ldots, a_{\tau_l}, x + z_0 - y) \\
&\quad + E\Bigg[f_{t+1}(\bar{\mathbf{a}} - \bar{\mathbf{R}}, W, (x + z_0 - y \\
&\quad - \sum_{\tau=\tau_l}^{\tau_u} R_\tau - D^u)^+, z_1, \ldots, z_L) \Bigg], \qquad (1)
\end{aligned}
$$

where

$$
\begin{aligned}
L&(a_{\tau_u}, \ldots, a_{\tau_l}, x + z_0 - y) \\
&= hE\left[(x + z_0 - y - \sum_{\tau=\tau_l}^{\tau_u} R_\tau - D^u)^+ \right] \\
&\quad + c_e E\left[\left(\sum_{\tau=\tau_l}^{\tau_u} R_\tau + D^u - x - z_0 + y \right)^+ \right]
\end{aligned}
$$

is the one-period holding penalty cost, and $f_{T+1}(\mathbf{a}, \mathbf{z}) = 0$, $\bar{\mathbf{a}} = (a_{\tau_u-1}, \ldots, a_0)$, $\bar{\mathbf{R}} = (R_{\tau_u-1}, \ldots, R_0)$, noting that $R_\tau = 0$ for $\tau \in \{0, \ldots, \tau_l - 1\}$. Let $z_L^*(\mathbf{a}, \mathbf{z})$ and $y^*(\mathbf{a}, \mathbf{z})$ be an optimal combination for the order size and the return size for any state $(\mathbf{a}, \mathbf{z}) \in \mathcal{U} \times \mathcal{Z}$, respectively. In the case of multiple optima, we take a smallest vector solution. (In Lemma 5, we show that there is always a unique smallest vector solution.)

The following lemma indicates that $z_L^*(\mathbf{a}, \mathbf{z})$ and $y^*(\mathbf{a}, \mathbf{z})$ cannot be both strictly positive.

Lemma 1. *For each $(\mathbf{a}, \mathbf{z}) \in \mathcal{U} \times \mathcal{Z}$ and $t = 1, \ldots, T$, the optimal decisions are characterized by $z_L^*(\mathbf{a}, \mathbf{z}) \times y^*(\mathbf{a}, \mathbf{z}) = 0$.*

3. Characterization of the optimal policy

We contribute to the literature in the following way: We characterize the optimal policy with respect to the on-hand and pipeline inventory levels by using $L^\natural$-convexity (Murota, 2003), a notion that implies both discrete convexity and submodularity (Topkis, 1998). Note that $L^\natural$-convexity has been used for the analysis of several inventory models (Zipkin, 2008b; Li and Yu, 2014). Zipkin (2008b) uses the notion for structural analysis of the standard single-item lost sales inventory system by applying a state transformation. He also provides an extension for a more general Markov-Modulated Demand Process (MMDP). Our problem is more complicated than the one in Zipkin (2008b). First, the demand process in this article cannot be modeled via an MMDP. One could see the states $\mathbf{a}$ as the states of a Markov chain that models the world. In that case, the demand in each period is fully determined by the world state $\mathbf{a}$ at the beginning of the period. However, if many (or few) demand signals result in actual demands in that period, then that gives a high (low) total demand in that period but also leads to a transition to a world state $\mathbf{a}'$ with few (many) demand signals. In other words, given a world state $\mathbf{a}$ at the beginning of a period, there is a coupling between the probabilities for the total demand in that period and the probabilities for the transitions to the next world state. This is not allowed in an MMDP. Second, we allow the return of excess stock and thus we have an additional decision variable. Hence, the structural analysis and the state transformation in Zipkin (2008b) are not directly applicable to our setting. Therefore, we propose a different state transformation. Pang *et al.* (2012) use the same transformation for a different setting in which they consider a pricing and ordering decision for an inventory system with backorders. In contrast with their paper, we have a lost sales inventory system with returning and ordering decisions and, consequently, our analysis is different from that of Pang *et al.* (2012). We extend Zipkin's analysis in the following ways: We show that for given values of numbers of demand signals from multiple periods the transformed cost function is $L^\natural$-convex (Theorem 1), the optimal order (return) size is monotone decreasing (increasing) in the on-hand and pipeline inventory levels with a slope of no more than one and more sensitive to recent (early) orders (Corollary 1).

In Section 3.1, the notation, $L^\natural$-convexity, and submodularity are introduced. Then, the structural properties of the optimal policy are characterized in Section 3.2. All proofs are provided in the Appendix.

3.1. General properties

Before proceeding, we remind the reader that $L^\natural$-convexity can be defined on integer lattices (Murota, 2003) as well as on real numbers (Zipkin, 2008b). In this article, we stick to the original definition in Murota (2003) that is based on integer lattices. However, different from Murota (2003), we work with non-negative integer variables (see Remark 1). First, we start with some definitions and notation. Let $\mathbf{X} \subseteq \mathbb{N}_0^l$ be a partially ordered set of vectors with a component-wise ordering of vectors; i.e., this means $\mathbf{x} \geq \mathbf{w}$ if and only if $x_i \geq w_i$ for all $i = 1, \ldots, l$ for each $\mathbf{x}$ and $\mathbf{w} \in \mathbf{X}$. This partially ordered set $\mathbf{X}$ forms a lattice if it contains the component-wise maximum $\mathbf{x} \vee \mathbf{w}$ and minimum $\mathbf{x} \wedge \mathbf{w}$ of each pair $\mathbf{x}$ and $\mathbf{w} \in \mathbf{X}$. If a subset of $\mathbf{X}$ contains the component-wise maximum and minimum of each pair of its elements, then this subset is a sublattice (of $\mathbf{X}$) and itself forms a lattice. A partially ordered set is a chain (ordered set) if either $\mathbf{x} \geq \mathbf{w}$ or $\mathbf{x} \leq \mathbf{w}$ holds for each pair $\mathbf{x}$ and $\mathbf{w} \in \mathbf{X}$. Let $\mathbf{e}_i$ be a vector having all entries zero except for 1 in its ith entry, and let $\mathbf{e}$ denote a vector of ones. A function $f : \mathbb{N}_0^l \to \mathbb{R}$ is said to be *increasing* (*decreasing*) in $x_i \in N_0$ with $i = 1, \ldots, l$ if $f(\mathbf{x} + \mathbf{e}_i) - f(\mathbf{x}) \geq 0$ (≤ 0) for all $\mathbf{x} \in \mathbf{X}$. Let m and n be positive integers and $\mathbf{M}$ and $\mathbf{N}$ be sublattices of $\mathbb{N}_0^m$ and $\mathbb{N}_0^n$, respectively. Then, their Cartesian product $\mathbf{M} \times \mathbf{N}$ also forms a lattice. Also, let $\overline{\mathbf{N}} = \{(\mathbf{y}, \varepsilon) : \mathbf{y} \in \mathbf{N}, \varepsilon \in \mathbb{N}_0; \varepsilon \leq y_j \forall j\}$. Then, $\overline{\mathbf{N}}$ is a sublattice (of $\mathbf{N} \times \mathbb{N}_0$), as it involves constraints of type $\varepsilon - y_j \leq 0$ having at most two variables with opposite signs (this also holds for $\varepsilon - y_j \leq b$ for constant $b \in \mathbb{N}_0$); see Topkis (1998), Example 2.2.7(b). Also, for any set $\mathbf{S} \subseteq \mathbb{N}_0^l$, let $\mathbf{S}_i$ denote the set of values of the ith argument of all vectors in $\mathbf{S}$ for all $i = 1, \ldots, l$. For a function $g : \mathbf{S} \to \mathbb{R}$, let $\Delta_{x_i} g(\mathbf{x}) = g(\mathbf{x} + \mathbf{e}_i) - g(\mathbf{x})$ and $\Delta_{x_i}\Delta_{x_j} g(\mathbf{x}) = g(\mathbf{x} + \mathbf{e}_i + \mathbf{e}_j) - g(\mathbf{x} + \mathbf{e}_j) - g(\mathbf{x} + \mathbf{e}_i) + g(\mathbf{x})$ denote first and second-order differences, respectively, for each $i = 1, \ldots, l$ and $j = 1, \ldots, l$. We say that $g(\mathbf{x})$ has *increasing* (*decreasing*) *differences* in x_i and x_j for any $i \neq j$ if $\Delta_{x_i}\Delta_{x_j} g(\mathbf{x}) \geq 0$ (≤ 0).

Next, we define submodularity, $L^\natural$-convexity, and some properties regarding these notions:

Definition 1. A function $g : \mathbf{M} \times \mathbf{N} \to \mathbb{R}$ is submodular in $\mathbf{y} \in \mathbf{N}$ for each $\mathbf{x} \in \mathbf{M}$ if $g(\mathbf{x}, \mathbf{y}) + g(\mathbf{x}, \mathbf{v}) \geq g(\mathbf{x}, \mathbf{y} \wedge \mathbf{v}) + g(\mathbf{x}, \mathbf{y} \vee \mathbf{v})$ for all $\mathbf{y}$ and $\mathbf{v} \in \mathbf{N}$ for each $\mathbf{x} \in \mathbf{M}$.

The following property follows from Corollary 2.6.1 of Topkis (1998).

Property 1. A function $g : \mathbf{M} \times \mathbf{N} \to \mathbb{R}$ is submodular in $\mathbf{y} \in \mathbf{N}$ for each $\mathbf{x} \in \mathbf{M}$ if each set $\mathbf{N}_i$ (of $\mathbf{N}$) forms a chain and $g(\mathbf{x}, \mathbf{y})$ has decreasing differences for all y_i and y_j; i.e., $\Delta_{y_i}\Delta_{y_j} g(\mathbf{x}, \mathbf{y}) \leq 0$, for all $i \neq j \in \{1, \ldots, n\}$ for each $\mathbf{x} \in \mathbf{M}$.

In what follows, we give the definition of $L^\natural$-convexity. The original definition is based on L-convexity (Murota, 2003). Here, we skip this step and also linearity in direction $\mathbf{e}$ and make the definition by directly relating the notion with submodularity (see also Remark 1).

Definition 2. A function $g : \mathbf{M} \times \mathbf{N} \to \mathbb{R}$ is $L^\natural$-*convex* in $\mathbf{y} \in \mathbf{N}$ for each $\mathbf{x} \in \mathbf{M}$ if $\psi(\mathbf{x}, \mathbf{y}, \varepsilon) = g(\mathbf{x}, \mathbf{y} - \varepsilon\mathbf{e})$ is submodular in $(\mathbf{y}, \varepsilon) \in \overline{\mathbf{N}}$ for each $\mathbf{x} \in \mathbf{M}$.

Remark 1. Our definition is slightly different from the original definition (Murota, 2003): We limit ourselves to non-negative integer variables. We define the dummy variable $\varepsilon \in \mathbb{N}_0$ such that $\varepsilon \leq y_j$ for all $j = 1, \ldots, n$. In this way, we guarantee that $\mathbf{y} - \varepsilon\mathbf{e}$, the second argument of $g(\mathbf{x}, \mathbf{y} - \varepsilon\mathbf{e})$, is a non-negative vector and $g(\mathbf{x}, \mathbf{y} - \varepsilon\mathbf{e})$ is defined on lattice $\mathbf{M} \times \mathbf{N}$. In the proof of Theorem 1, the dummy variable corresponds to a physical value; i.e., amount deducted from stock. But even there, constraints $\varepsilon \leq y_j$ for all $j = 1, \ldots, n$ are automatically satisfied. The condition that requires linearity in direction $\mathbf{e}$ is not considered since this is automatically satisfied as in Zipkin (2008b).

The following property indicates that $L^\natural$-convexity implies submodularity and the proof can be given along the same line as that of Theorem 7.1 in Murota (2003).

Property 2. If $g : \mathbf{M} \times \mathbf{N} \to \mathbb{R}$ is $L^\natural$-convex in $\mathbf{y} \in \mathbf{N}$ for each $\mathbf{x} \in \mathbf{M}$, then $g(\mathbf{x}, \mathbf{y})$ is also submodular in $\mathbf{y}$ for each $\mathbf{x} \in \mathbf{M}$.

Next, we proceed with some lemmas to develop our results in Section 3.2. Lemma 2 is a crucial stepping stone in the proof of Theorem 1 (and also that of Lemma 3). Lemma 3 shows that $L^\natural$-convexity is preserved under minimization. Lemma 4 indicates that the minimizer of an $L^\natural$-convex function with respect to a set of its arguments is monotone increasing in other arguments, with limited sensitivity. Lemmas 2, 3, and 4 extend Zipkin's results (Lemmas 1, 2, 3 in his paper) to our setting in which we have an additional state vector–i.e., demand signals–and an additional decision variable; i.e., return size.

Lemma 2. *If $g : \mathbf{M} \times \mathbf{N} \to \mathbb{R}$ is $L^\natural$-convex in $\mathbf{y} \in \mathbf{N}$ for each $\mathbf{x} \in \mathbf{M}$, then $\psi(\mathbf{x}, \mathbf{y}, \varepsilon) = g(\mathbf{x}, \mathbf{y} - \varepsilon\mathbf{e})$ is $L^\natural$-convex in $(\mathbf{y}, \varepsilon) \in \overline{\mathbf{N}}_0$ for each $\mathbf{x} \in \mathbf{M}$.*

Let $\mathbf{U} = \mathbb{N}_0^u$, with u being a positive integer, be a lattice. Let $\widehat{\mathbf{N}}$ be a sublattice of $\mathbf{N} \times \mathbf{U}$.

Lemma 3. *If $h : \mathbf{M} \times \widehat{\mathbf{N}} \to \mathbb{R}$ is $L^\natural$-convex in $(\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}$ for each $\mathbf{x} \in \mathbf{M}$, then $g : \mathbf{M} \times \mathbf{N} \to \mathbb{R}$ with $g(\mathbf{x}, \mathbf{y}) := \min_{\boldsymbol{\xi}:(\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}} \{h(\mathbf{x}, \mathbf{y}, \boldsymbol{\xi})\}$ is $L^\natural$-convex in $\mathbf{y} \in \mathbf{N}$ for each $\mathbf{x} \in \mathbf{M}$.*

Lemma 4. *Suppose that $h : \mathbf{M} \times \widehat{\mathbf{N}} \to \mathbb{R}$ is $L^\natural$-convex in $(\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}$ and also suppose that $\min_{\boldsymbol{\xi}:(\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}} \{h(\mathbf{x}, \mathbf{y}, \boldsymbol{\xi})\}$ has a unique smallest vector solution denoted by $\boldsymbol{\xi}^*(\mathbf{x}, \mathbf{y})$. Then, for each $\mathbf{x} \in \mathbf{M}$:*

- *(a) $\boldsymbol{\xi}^*(\mathbf{x}, \mathbf{y})$ is increasing in $\mathbf{y} \in \mathbf{N}$.*
- *(b) $0 \leq \xi_i^*(\mathbf{x}, \mathbf{y} + k\mathbf{e}) - \xi_i^*(\mathbf{x}, \mathbf{y}) \leq k$ for $k \in \mathbb{N}_0^+$ and for all $i = 1, \ldots, u$ and $j = 1, \ldots, n$.*
- *(c) $0 \leq \boldsymbol{\xi}^*(\mathbf{x}, \mathbf{y} + k\mathbf{e}) - \boldsymbol{\xi}^*(\mathbf{x}, \mathbf{y}) \leq k\mathbf{e}$ for $k \in \mathbb{N}_0^+$ and for all $j = 1, \ldots, n$.*

The following property implies that $L^\natural$-convexity is preserved under expectation and it follows from Corollary 2.6.2 of Topkis (1998).

Property 3. Suppose that $\mathbf{R}$ is a random vector with domain $\mathbf{U}$ having an arbitrary distribution function $f(\mathbf{r})$. If a function $h : \mathbf{M} \times \mathbf{N} \times \mathbf{U} \to \mathbb{R}$ is $L^\natural$-convex in $\mathbf{y} \in \mathbf{N}$ for each $\mathbf{x} \in \mathbf{M}$ and for each realization $\mathbf{r}$ of $\mathbf{R}$, then $E[h(\mathbf{x}, \mathbf{y}, \mathbf{R})]$ is $L^\natural$-convex in $\mathbf{y} \in \mathbf{N}$ for each $\mathbf{x} \in \mathbf{M}$.

3.2. The state transformation and the structural properties with respect to inventory levels

Next, we apply our state transformation. Here, we use a different transformation than the one in Zipkin (2008b), see Remark 2. We let $v_l = x + \sum_{t=0}^{l} z_t$ for $l = -1, \ldots, L$ and $\mathbf{v} = (v_{-1}, \ldots, v_{L-1})$. Then, the state space is defined by the Cartesian product of $\mathcal{U} = \{\mathbf{a} : \mathbf{a} \in \mathbb{N}_0^{\tau_u+1}\}$ and $\mathcal{V} = \{\mathbf{v} : \mathbf{v} \in \mathbb{N}_0^{L+1}, v_{-1} \le v_0 \le \cdots \le v_{L-1}\}$; the action space is given by $\bar{\mathcal{A}}_{v_{-1},v_{L-1}} = \{(v_L, y) : v_L, y \in \mathbb{N}_0, y \le v_{-1}, v_L \ge v_{L-1}\}$; the Cartesian product of $\mathcal{V}$ and $\bar{\mathcal{A}}_{v_{-1},v_{L-1}}$ is given by $\mathcal{Q} = \{(\mathbf{v}, v_L, y) : \mathbf{v} \in \mathcal{V}, (v_L, y) \in \mathbb{N}_0^2, y \le v_{-1}, v_L \ge v_{L-1}\}$; and an optimal solution (a smallest vector solution in case of multiple optima) for any state $(\mathbf{a}, \mathbf{v})$ is denoted by $(v_L^*(\mathbf{a}, \mathbf{v}), y^*(\mathbf{a}, \mathbf{v}))$.

The optimal total cost function from time t onwards is defined by

$$\bar{f}_t(\mathbf{a}, \mathbf{v}) = \min_{(v_L, y) \in \bar{\mathcal{A}}_{v_{-1},v_{L-1}}} \{\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y)\}, \quad (2)$$

$$\begin{aligned}\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y) = {}& c(v_L - v_{L-1}) + c_r y + L(a_{\tau_u}, \ldots, a_{\tau_l}, v_0 - y) \\ & + E\Bigg[\bar{f}_{t+1}\Bigg(\bar{\mathbf{a}} - \bar{\mathbf{R}}, W, \\ & \left(v_0 - y - \sum_{\tau=\tau_l}^{\tau_u} R_\tau - D^u\right)^+, v_1 - v_0 \\ & + \left(v_0 - y - \sum_{\tau=\tau_l}^{\tau_u} R_\tau - D^u\right)^+, \\ & \ldots, v_L - v_0 + \left(v_0 - y - \sum_{\tau=\tau_l}^{\tau_u} R_\tau - D^u\right)^+\Bigg)\Bigg], \end{aligned} \quad (3)$$

where

$$\begin{aligned} L(a_{\tau_u}, \ldots, a_{\tau_l}, v_0 - y) = {}& hE\left[(v_0 - y - \sum_{\tau=\tau_l}^{\tau_u} R_\tau - D^u)^+\right] \\ & + c_e E\left[\left(\sum_{\tau=\tau_l}^{\tau_u} R_\tau + D^u + y - v_0\right)^+\right]. \end{aligned}$$

We note that $\mathcal{V}$ is a lattice on N_0^{L+1} and $\mathcal{Q}$ is a sublattice of $\mathcal{V} \times \mathbb{N}_0^2$ since each involves constraints having at most two variables with opposite signs; see Topkis (1998), Example 2.2.7(b). Also, note that $\bar{f}_t(\mathbf{a}, \mathbf{v}) = f_t(\mathbf{a}, \mathbf{z})$.

Now we can establish one of our key results by using the transformed model.

Theorem 1.

(a) *$\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y)$ is $L^\natural$-convex in $(\mathbf{v}, v_L, y) \in \mathcal{Q}$ for each $\mathbf{a} \in \mathcal{U}$ and $t = 1, \ldots, T$.*

(b) *$\bar{f}_t(\mathbf{a}, \mathbf{v})$ is $L^\natural$-convex in $\mathbf{v} \in \mathcal{V}$ for each $\mathbf{a} \in \mathcal{U}$ and $t = 1, \ldots, T+1$.*

(c) *$J_t(\mathbf{a}, \mathbf{z}, z_L, y)$ is component-wise convex in z_L and y, i.e., $\Delta_{z_L}\Delta_{z_L}J_t(\mathbf{a}, \mathbf{z}, z_L, y) \ge 0$ and $\Delta_y\Delta_y J_t(\mathbf{a}, \mathbf{z}, z_L, y) \ge 0$, for each $\mathbf{a} \in \mathcal{U}$ and $t = 1, \ldots, T$.*

(d) *$f_t(\mathbf{a}, \mathbf{z})$ is multimodular; hence, it has increasing differences and component-wise convexity, for each $\mathbf{a} \in \mathcal{U}$ and $t = 1, \ldots, T+1$.*

Remark 2. Stating $\bar{f}_t(\mathbf{a}, \mathbf{v})$ as a function of $\bar{f}_{t+1}(\bar{\mathbf{a}} - \bar{\mathbf{r}}, w, \mathbf{v} - \varepsilon\mathbf{e})$ (along with additional functions that are $L^\natural$-convex) is a key step in the proof of $L^\natural$-convexity (see the proof of Theorem 1 for the details and also the corresponding variable for ε). The proof is based on induction: First, we start with the assumption that $\bar{f}_{t+1}(\mathbf{a}, \mathbf{v})$ is $L^\natural$-convex in $\mathbf{v}$ for each $\mathbf{a}$ for all $t = 1, \ldots, T+1$ eventually to show that this also holds for $\bar{f}_t(\mathbf{a}, \mathbf{v})$. By Lemma 2 and $L^\natural$-convexity of $\bar{f}_{t+1}(\mathbf{a}, \mathbf{v})$ in $\mathbf{v}$ for each $\mathbf{a}$, we establish that $\bar{f}_{t+1}(\bar{\mathbf{a}} - \bar{\mathbf{r}}, w, \mathbf{v} - \varepsilon\mathbf{e})$ is $L^\natural$-convex in $(\mathbf{v}, \varepsilon)$ for each $(\bar{\mathbf{a}} - \bar{\mathbf{r}}, w)$. Finally, using the expression that defines $\bar{f}_t(\mathbf{a}, \mathbf{v})$ as a function of $\bar{f}_{t+1}(\bar{\mathbf{a}} - \bar{\mathbf{r}}, w, \mathbf{v} - \varepsilon\mathbf{e})$, we show that $\bar{f}_t(\mathbf{a}, \mathbf{v})$ is $L^\natural$-convex in $\mathbf{v}$ for each $\mathbf{a}$. This simple idea works well because we define $v_l = x + \sum_{t=0}^{l} z_t$ for $l = -1, \ldots, L$, and this enables y to appear as $-y$ (embedded inside $-\varepsilon$) in all arguments of $\mathbf{v} - \varepsilon\mathbf{e}$ in the expression $\bar{f}_{t+1}(\mathbf{a}, \mathbf{v} - \varepsilon\mathbf{e})$. On the contrary, had we worked with Zipkin's transformation (the state variable would have then been $v_l = \sum_{t=l}^{L-1} z_t$ for $l = 0, \ldots, L$ and $v_{-1} = x$), our additional decision variable y, which does not exists in Zipkin's model, would have appeared as $-y$ only in the first argument of $\mathbf{v}$ and hence we would not have had this nice form. This explains why we need a different state transformation than Zipkin (2008b) and the importance of the state transformation step in the analysis. Our state transformation seems appropriate for inventory models in which return size is also a decision variable in addition to the order size.

Lemma 5. *For each $(\mathbf{a}, \mathbf{z}) \in \mathcal{U} \times \mathcal{Z}$ and $t = 1, \ldots, T$, there is a unique smallest vector solution of $\min_{(z_L, y) \in \mathcal{A}_x}\{J_t(\mathbf{a}, \mathbf{z}, z_L, y)\}$, which is denoted by $(z_L^*(\mathbf{a}, \mathbf{z}), y^*(\mathbf{a}, \mathbf{z}))$.*

Lemma 5 holds also for the optimal solution of our transformed model (see the proof of Lemma 5). Thus, $(v_L^*(\mathbf{a}, \mathbf{v}), y^*(\mathbf{a}, \mathbf{v}))$ denotes the smallest vector solution for each $(\mathbf{a}, \mathbf{v})$.

Next, by using the results of Theorem 1, we define the monotonicity properties of the optimal policy with respect to on-hand and pipeline inventory levels.

Corollary 1. *For each $\mathbf{a} \in \mathcal{U}$ and $t = 1, \ldots, T$:*

(a) $0 \le \Delta_{v_i} v_L^*(\mathbf{a}, \mathbf{v}) \le 1$ *for* $i = -1, \ldots, L-1$.

(b) $0 \le \Delta_{v_i} y^*(\mathbf{a}, \mathbf{v}) \le 1$ *for* $i = -1, \ldots, L-1$.

(c) $-1 \le \Delta_{z_{L-1}} z_L^*(\mathbf{a}, \mathbf{z}) \le \cdots \le \Delta_{z_0} z_L^*(\mathbf{a}, \mathbf{z}) \le \Delta_x z_L^*(\mathbf{a}, \mathbf{z}) \le 0$,

(d) $0 \le \Delta_{z_{L-1}} y^*(\mathbf{a}, \mathbf{z}) \le \cdots \le \Delta_{z_0} y^*(\mathbf{a}, \mathbf{z}) \le \Delta_x y^*(\mathbf{a}, \mathbf{z}) \le 1$.

Corollary 1(c) shows that the optimal order quantity is decreasing in on-hand and pipeline inventories (with a rate less than one) and it is more sensitive to earlier orders. In that sense, Theorem 1 and Corollary 1(c) generalize Zipkin's results (Theorem 4 and Corollary 5 in Zipkin (2008b)) to a lost sales system with imperfect ADI and inventory returns to the upstream supplier. Corollary 1(d) is completely new to the literature. It contributes to studies on (imperfect) ADI, lost sales inventory systems, and inventory systems with inventory returns to the upstream supplier by showing that, in contrast with the optimal order quantity, the optimal return quantity is increasing in on-hand and pipeline inventories (with an increase less than one) and it is more sensitive to earlier orders.

Our results in this section are as follows:

1. We show that optimal order (return) size and inventory levels are economic substitutes (complements).
2. Since $v_L^*(\mathbf{a}, \mathbf{v})$ corresponds to the inventory position, Corollary 1(a) indicates that the inventory position is increasing (or changing) with $\mathbf{v} \in \mathcal{V}$. This indicates that the optimal ordering (also return) decision is dependent not only on on-hand inventory but also on where the previous replenishment order(s) are in the pipeline. Hence, the optimal inventory position is not necessarily a constant and, therefore, a simple (state-independent) base stock policy, which is widely used in practice, is not necessarily optimal.
3. By Theorem 1(c), $J_t(\mathbf{a}, \mathbf{z}, z_L, y)$ is component-wise convex in $z_L \in \mathbb{N}_0$ and $y \in \{y \in \mathbb{N}_0 : y \leq x\}$ and this can be exploited to speed up the search for the optimal $z_L^*(\mathbf{a}, \mathbf{z})$ and $y^*(\mathbf{a}, \mathbf{z})$ at each iteration of the value iteration algorithm.
4. Bounds can be obtained for $z_L^*(\mathbf{a}, \mathbf{z})$. Since a lexicographic order is followed for $(\mathbf{a}, \mathbf{z})$ to find $z_L^*(\mathbf{a}, \mathbf{z})$, $z_L^*(\mathbf{a}, \mathbf{z} - \mathbf{e}_i)$ with $i \in \{1, \ldots, L+1\}$ is always obtained in earlier steps. By the monotonicity, $z_L^*(\mathbf{a}, \mathbf{z} - \mathbf{e}_i)$ can be used as an upper bound on $z_L^*(\mathbf{a}, \mathbf{z})$ and $y^*(\mathbf{a}, \mathbf{z} - \mathbf{e}_i)$ can be used as a lower bound on $y^*(\mathbf{a}, \mathbf{z})$ for $i \in \{1, \ldots, L+1\}$.

Each iteration of the value iteration algorithm takes polynomial time with an order of $\mathcal{O}(|\mathcal{A}_x| \times |\mathcal{U} \times \mathcal{Z}^2|)$ where $|\mathcal{A}_x|$ is the size of the action space (for a given x value) and $|\mathcal{U} \times \mathcal{Z}^2|$ is the size of the state space. However, the number of iterations grows exponentially with the discount factor (Bertsekas and Tsitsiklis, 1996), which is implicitly one in our case. Therefore, the value iteration algorithm is not guaranteed to run in polynomial time. In contrast, the myopic policy, in which the number of iterations is $\max(L, \tau_u) + 1$ and therefore the running time is $\mathcal{O}(\max(L, \tau_u) \times |\mathcal{A}_x| \times |\mathcal{U} \times \mathcal{Z}^2|)$, guarantees a polynomial-time solution. The lower and the upper bounds proposed for the optimal order and return sizes decrease the running time by reducing the size of the action space $|\mathcal{A}_x|$; however, they do not have any effect on the order of complexity.

4. Computational study

We conduct an experimental study to investigate the value of imperfect ADI and the benefit of returning excess stock under our imperfect ADI setting. While using our model in Section 2 to conduct our analysis, we consider two alternatives: First, we use a value iteration algorithm to obtain the optimal long-run average cost. The algorithm is run until it converges with a specified accuracy, as described in Puterman (1994). Second, for large-scale problem instances where the optimal solution becomes intractable, we consider the myopic solution of the problem (1) as a heuristic, which takes into account only the maximum of the lead time ahead and the prediction horizon. Hence, we solve the recursion (1) for $T = \max(L, \tau_u) + 1$. In our computational study, we explore the performance of the myopic policy.

The long-run average per period cost is considered as a performance measure. Therefore, we define

$$g_{ADI} = \lim_{T \to \infty} \frac{f_0(\mathbf{a}, \mathbf{z})}{T}$$

as the optimal long-run average cost per period under imperfect ADI, which is obtained by using a value iteration algorithm. Similarly, we define g_{NoADI} as the optimal long-run average cost per period for the system without imperfect ADI. To obtain g_{NoADI}, we take $q = 0$ (leading to $E[W] = 0$ and $E[D^u] = \lambda$) and consider all demand to be unpredicted. The value of imperfect ADI is evaluated in terms of the percentage cost reduction:

$$PCR_{ADI} = \frac{g_{NoADI} - g_{ADI}}{g_{NoADI}}$$

or simply *PCR*. The long-run average cost per period for the myopic policy under imperfect ADI, g_{MADI}, is obtained by running this policy in the infinite-horizon problem. Its performance is tested in terms of the percentage cost reduction relative to the optimal policy under no ADI:

$$PCR_{MADI} = \frac{g_{NoADI} - g_{MADI}}{g_{NoADI}}.$$

Apart from relative cost differences, we consider the absolute differences. In all experiments, our observations are similar in both measures.

Our computational study includes an extensive experiment to fully investigate the effects of parameters (Sections 4.2 and 4.3) and a spare-part case study based on the data of ASML (Section 4.4) to test with a case from practice. In Section 4.1, we explain our experimental design used in Sections 4.2 and 4.3.

4.1. Experimental design

We consider eight parameters for our experiment: lead time L, prediction interval $[\tau_l, \tau_u]$, return cost c_r, total demand rate λ, unit holding cost h, penalty cost c_e, precision p, and sensitivity q. While generating values of L and τ_l, we consider two cases: $L \leq \tau_l$ (case 1) and $L > \tau_l$ (case 2). Case 1 corresponds to the ideal situation where the demand signal is received sufficiently in advance so that it can be responded to using a regular replenishment order. Case 2 corresponds to the situation where the demand lead time can be shorter than the regular supply lead time; hence, only some (or none if $L > \tau_u$) of the demand signals can be responded to using a regular replenishment order (of course unless we keep safety stock). We consider two probability distributions for $\{p_\tau\}$: a truncated geometric distribution with $p_{\tau+1} = p \times p_\tau$ (here we take the success probability to be the same as p) and a uniform distribution with $p_{\tau+1} = p_\tau$ for all $\tau = \tau_l, \ldots, \tau_u - 1$, in both cases by setting p_τ such that $\sum_{\tau=\tau_l}^{\tau_u} p_\tau = p$. When we consider the spare parts case, these distributions correspond to having a constant or increasing failure rate, respectively. Once p, q, and λ are known, the predicted and unpredicted demand parameters are obtained by $E[D^u] = \lambda(1-q)$ and $E[W] = \lambda(q/p)$, respectively, as

$$\lambda = pE[W] + E[D^u] \text{ and } q = \frac{pE[W]}{pE[W] + E[D^u]}.$$

In this manner, average demand rate is set to be λ. In all experiments, we assume that W and D^u have Poisson distributions. For each case and probability distribution of $\{p_\tau\}$, we consider two levels of L and $[\tau_l, \tau_u]$; three levels of λ, h, c_e, p, and q; and four levels of c_r, resulting in a total of $2^2 \times 4 \times 3^5 = 3888$ problem instances for both case 1 and case 2. For two reasons, the values for c_r are defined as a multiple of h: When a return is made to a central warehouse, the return cost mainly consists of the pipeline inventory holding cost. In the case of a return to an external supplier, this cost is often higher, and it is expressed as a ratio of unit purchasing cost, still as a multiple of holding cost. A carrying charge of 0.4% per week (20% per year) is assumed. We set $c_r = 2.5h$, $25h$, and $125h$ (all satisfying $c_r \geq h \times L$), representing return costs of 1% ($2.5 \times 0.4\%$), 10%, and 50% of the unit purchasing cost, respectively. Higher return costs are representative of cases where a return is made to an external supplier, whereas $c_r \to \infty$ represents the situation in which returning excess inventory is not allowed. For simplicity, we exclude the unit procurement cost—i.e., purchasing priceand the regular transportation cost and set $c = 0$ in the experiments. Similarly, we exclude the purchasing price and the amount equivalent to regular cost from c_e. Table 2 summarizes the values of the parameters used in our experiment.

For computational purposes, the state space is truncated by taking $z_l \leq 5$ for all $l \in \{1, \ldots, L\}$ and $a_\tau \leq 5$ for all $\tau \in \{0, \ldots, \tau_u\}$, which are not restrictive considering the values of the demand parameters $E[D^u] = \lambda(1-q)$ and $E[W] = \lambda(q/p)$ taken in the experiments. We take the computational precision as $\epsilon = 10^{-6}$. As in all numerical experiments, we do not claim our observations to be valid outside the problem setting and the range of problem instances we consider.

4.2. The value of imperfect ADI and benefit of returning excess stock

4.2.1. $L \leq \tau_l$ (case 1)

A summary of the results for this case is presented in Figure 1, which illustrates the average *PCR* for each level of parameters L, $[\tau_l, \tau_u]$, λ, h, c_e, p, and q for different levels of c_r. The results of the experiments are detailed in Table A.1 in the Appendix. The main observations drawn from the factorial experiment are given as follows:

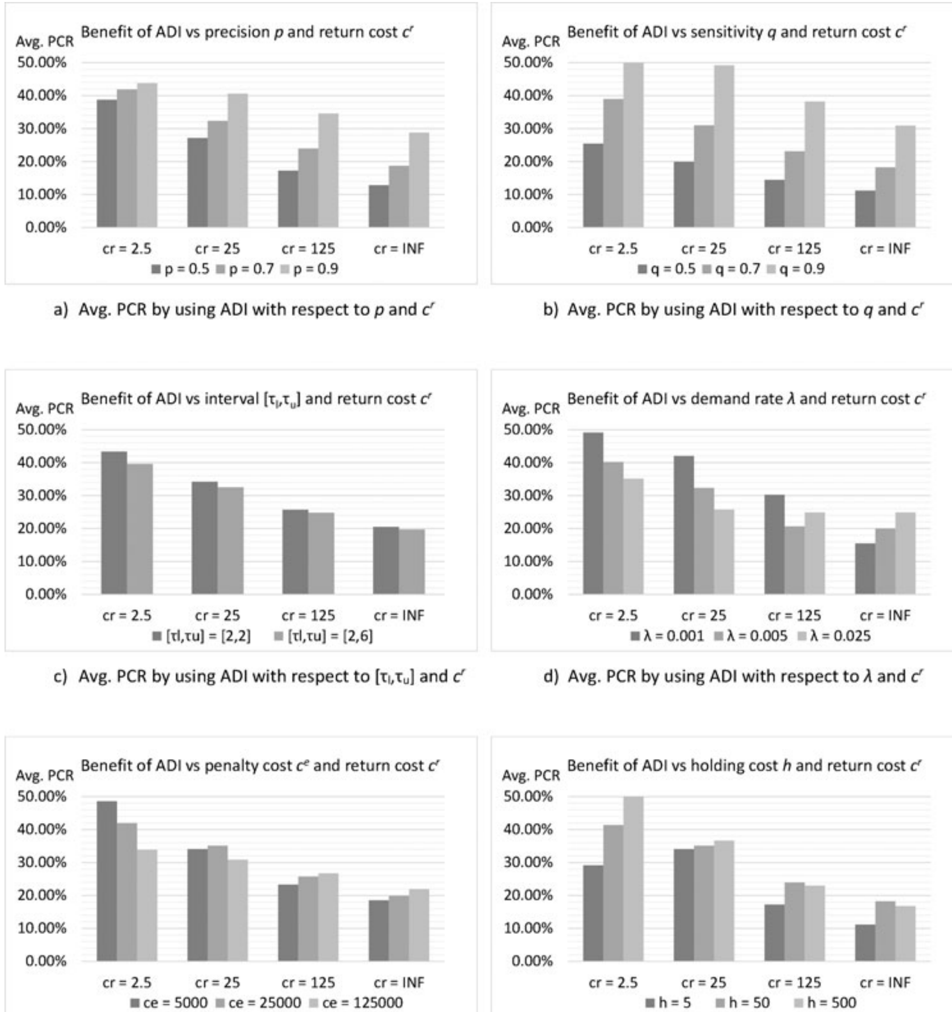
1. The average benefit of ADI is very high. Despite the imperfection, the average *PCR* is found to be 30.06%, and the maximum *PCR* is 89.96%.
2. The value of ADI declines sharply with increased levels of imperfection. Among the two measures used to measure the extent of the imperfection, the ratio of predicted demand over total demand is found to be very important, even more than the precision of the ADI. This result suggests that the parameters of the prediction tools' warning limits should be set such that the model detects as many failures as possible, and this might be achieved at the expense of some level of precision. Hence, it might be favorable to place more cheaper, less accurate sensors than fewer more accurate, expensive ones.
3. For high values of q and p, the value of imperfect demand signals increases with q and p with an increasing rate. Considering that q has some correspondence with the fraction of customers that provide ADI, our observation for q extends the results in Gayon *et al.* (2009), who report that the benefit of imperfect ADI increases linearly with, the fraction of customers providing ADI for a system when the demand lead time is exponentially distributed.
4. Provided that L is shorter than τ_l, knowing the exact time of a demand occurrence does not have a significant impact on the benefit of the information. The reason for this behavior is that when $L \leq \tau_l$, it is possible to react to ADI anyway. In Section 4.2.2, we illustrate that this is not true for $L > \tau_l$.
5. Returning excess stock is influential on the value of imperfect ADI. The benefit is quite substantial. As return cost decreases, the value of information significantly increases and it becomes less sensitive to the precision of the information. Therefore, particularly for very slow-moving items, the value of information is extremely high for low return costs, making returning excess stock even more attractive for slow-moving items. Furthermore, for lower return costs, we observe that the optimal policy has a less simple structure that has a lower dependence on the state.
6. The parameters λ, h, and c_e are highly influential on the value of imperfect ADI. However, their effects are non-monotonic and highly dependent on the value of c_r. For example, when returning excess stock is a viable option, the system benefits from using imperfect demand signals more for expensive parts; however, if returning excess inventory is not possible, the system responds in exactly the opposite way. Our observations are similar for demand rate and emergency cost. Therefore, the possibility to return is described as a game changer.

Table 2. Parameter values for the testbed.

Parameters	*Values*
L (week)	1, 2
$[\tau_l, \tau_u]$ (week)	[2, 2], [2, 6] for $L \leq \tau_l$, case 1 [0, 0], [0, 4] for $L > \tau_l$, case 2
c_r (€/unit)	$2.5h$, $25h$, $125h$, ∞
λ (units/week)	0.001, 0.005, 0.025
h (€/unit/week)	5, 50, 500
c_e(€/unit)	5000, 25 000, 125 000
p	0.5, 0.7, 0.9
q	0.5, 0.7, 0.9

4.2.2. $L > \tau_l$ (case 2)

The results of the experiments are summarized in Figure 2 (the detailed results are presented in Table A.2 in the Appendix). Since the average *PCR* differs significantly for each level of $[\tau_l, \tau_u]$, we present these values separately. The main observations are thus as follows:



a) Avg. PCR by using ADI with respect to p and c^r

b) Avg. PCR by using ADI with respect to q and c^r

c) Avg. PCR by using ADI with respect to $[\tau_l,\tau_u]$ and c^r

d) Avg. PCR by using ADI with respect to λ and c^r

e) Avg. PCR by using ADI with respect to c^e and c^r

f) Avg. PCR by using ADI with respect to h and c^r

Figure 1. Effect of parameters on the value of imperfect ADI (case 1, $L \leq \tau_l$).

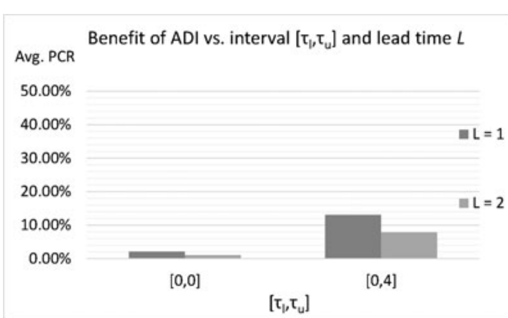


Figure 2. Effect of parameters on the value of imperfect ADI (case 2, $L > \tau_l$).

1. The benefit of using imperfect ADI is lower when $L > \tau_l$. The average PCR is found to be 6.03%. This shows that the delivery time of the ADI has a high impact on the value of imperfect ADI, even more than the other two measures of imperfect behavior p and q. Based on this observation, we make the following important suggestion for the design of prediction tools: A warning limit should be set such that it can issue a signal far enough, if possible, a regular replenishment lead time, in advance and this might be achieved at the expense of some precision p and sensitivity q of the ADI.
2. Late demand signals still have some value. This is because demand signals can still be useful in predicting the lead time demand and, hence, also the inventory level at the end of the lead time. The benefit of ADI is influenced not only by τ_l being less than L but also by how much it is less than L. Provided that $L > \tau_l$, the value of ADI increases monotonically with τ_l. Note that when $\tau_l = \tau_u$, our setting corresponds to the case where the demand lead time is constant. In this sense, our observation is in line with Hariharan and Zipkin (1995), who report monotonic increase of value of ADI with demand lead time.

Furthermore, we make the following observations:

1. The value of ADI is found to be slightly higher for a uniform distribution when $L > \tau_l$. This can be explained as follows: under a uniform distribution, which has an increasing failure rate, a customer's demand is likely to occur later. This increases the chance that the demand is satisfied by a regular replenishment order, hence also increasing the benefit of ADI.

Table 3. Summary of the results (case 1, $L = 2$, $\tau_l = 2$, $\tau_u = 2$).

		With ADI			
		Without return	With return		
Policy	Without ADI Avg. PCR (%)	$c_r = \infty$ Avg. PCR (%)	High ($c_r = 125h$) Avg. PCR (%)	Med. ($c_r = 25h$) Avg. PCR (%)	Low ($c_r = 2.5h$) Avg. PCR (%)
Optimal	0.00	21.03	26.24	34.81	44.07
Myopic	0.00	1.40	9.18	27.77	43.02

2. It is noteworthy that not only substantial cost savings are achieved by using imperfect ADI but customer responsiveness (measured by the average rate of demand satisfied from stocks or simply $E[\max(D, y)]/E[D]$) is slightly improved.

4.3. Performance of the myopic policy

As a part of the numerical analysis, we also test the performance of using the myopic solution of the problem. Table 3 summarizes our results regarding how the optimal policy, the myopic policy, and returning excess inventory can be used as a tool to make best use of imperfect ADI when $L = 2$ and $[\tau_l = 2, \tau_u = 2]$: The myopic policy does not perform well when c_r is high. Therefore, when c_r is high, using the optimal policy, which has a complex structure and requires computational effort, inevitably benefits from imperfect ADI. However, if c_r is low, it is possible to achieve high benefits from imperfect ADI by also using the myopic policy. We note that the figures that we report here for the performance of the myopic policy are significantly lower than those for lost sales inventory systems without ADI (Zipkin, 2008a). This is attributed to the fact that our system involves high demand variability, where the performance of the myopic policies is known to be relatively poor (Levi *et al.*, 2007).

4.4. Case study on ASML

In this section, we perform a case study by using the data provided by ASML. The data set involves data for four parts that are representative and reflect different characteristics of spare parts that ASML supplies to its customers all over the world. We abbreviate these parts as P, T, X, and W. Our aim is to analyze potential cost savings for a single stock point based on these four parts that are important for the company. The values in the data set are for a relatively small local warehouse. Precision p, sensitivity q, and lower and upper limits for failure time, τ_l and τ_u, are obtained from the prediction tool in use at ASML and c_e is calculated as the average of emergency cost from a nearby local warehouse and that from the central warehouse, each consisting of a transportation cost and a downtime cost incurred while waiting for part shipment. Per unit return cost is defined as the sum of transportation cost and the pipeline holding cost for parts returned to the central warehouse. We exclude the purchasing price of the parts and include only the transportation cost in the experiments. Time unit is week and costs are in euros as before. Based on these considerations, we take $c_e = 75\,000$, $L = 2$, $c = 100$ for all parts. We define g_{ADINR} as the optimal long-run average cost per period under imperfect ADI with no return. To evaluate the performance of the value of ADI under no return case, we define

$$PCR_{ADINR} = \frac{g_{NoADI} - g_{ADINR}}{g_{NoADI}}.$$

For problem instances with $\tau_l > L$, we run our model by substracting $\tau_l - L$ from τ since the ADI available more than lead time in advance is useless; e.g., while running our model for part T, we take $(\tau_l = 2, \tau_u = 6)$. Values of the part-specific parameters and results of the experiment are summarized in Table 4.

Our observations are similar to previous observations: (i) timing of the ADI is highly important; e.g., the value of ADI is very low for parts X and W, which have $L > \tau_l$ (whereas very high for parts P, which have $L \leq \tau_l$) and (ii) returning excess inventory is quite powerful in coping with unprecise ADI; e.g., for part P, for which p is low and q is high, the value of ADI is high only when returning excess inventory is allowed. Furthermore, when we make the comparison against the (optimal) base stock policy, which is the policy in use at ASML, our observations are similar: the benefit of using the optimal policy is slightly higher than the one against the optimal solution under no ADI.

We also illustrate the characteristics of the optimal policy for four ASML parts. Tables 5(a) to 5(g) demonstrate the optimal values of the decision variables for part X for different values of $(\mathbf{a}, \mathbf{z})$. In each cell, a positive value indicates the order size, a negative value indicates the return size, and zero stands for no action. As seen in Table 5(d), when there are three signals that have arrived at the beginning of the period ($\mathbf{a} = (0, 0, 0, 0, 3)$) and available inventory is zero ($x + z_0 = 0$) and pipeline stock is zero ($z_1 = 0$), then the optimal action is to order two units ($z_2^*(\mathbf{a}, \mathbf{z}) = 2$). This shows that the optimal action may involve ignoring a signal. Note that this is due to imperfection in p and/or timing. Also, as seen in Table 5(g), when there are

Table 4. Results of the experiment with ASML case data.

Part	h (€/unit/week)	$[\tau_l, \tau_u]$ (week)	λ (unit/week)	p	q	c_r	g_{NoADI} (€/week)	g_{ADINR} (€/week)	g_{ADI} (€/week)	PCR_{ADINR} (%)	PCR_{ADI} (%)
P	2720	[2, 6]	0.0188	0.42	0.44	5500	1406.01	1399.69	963.05	0.45	31.50
T	112	[8, 12]	0.0600	0.90	0.90	325	248.13	144.37	137.03	41.82	44.78
X	152	[0, 4]	0.0019	0.45	0.43	400	145.57	142.05	134.47	2.42	7.63
W	646	[0, 1]	0.0036	0.90	0.50	1400	274.28	274.28	274.28	0.00	0.00

Table 5. Values of (z_2^*, y^*) for different $(\mathbf{a}, \mathbf{z})$ values for part X.

Part X		$y+z_0$			
		0	1	2	3
z_1	**0**	0	−1	−2	−3
	1	0	−1	−2	−3
	2	0	−1	−2	−3
	3	0	−1	−2	−3

a) $\mathbf{a} = (0, 0, 0, 0, 0)$

Part X		$y+z_0$			
		0	1	2	3
z_1	**0**	1	0	−1	−2
	1	0	0	−1	−2
	2	0	0	−1	−2
	3	0	0	−1	−2

b) $\mathbf{a} = (0, 0, 0, 0, 1)$

Part X		$y+z_0$			
		0	1	2	3
z_1	**0**	1	1	0	−1
	1	1	0	0	−1
	2	0	0	0	−1
	3	0	0	0	−1

c) $\mathbf{a} = (0, 0, 0, 0, 2)$

Part X		$y+z_0$			
		0	1	2	3
z_1	**0**	2	1	1	0
	1	1	1	0	0
	2	0	0	0	0
	3	0	0	0	0

d) $\mathbf{a} = (0, 0, 0, 0, 3)$

Part X		$y+z_0$			
		0	**1**	**2**	**3**
z_1	**0**	0	0	−1	−2
	1	0	0	−1	−2
	2	0	0	−1	−2
	3	0	0	−1	−2

e) $\mathbf{a} = (1, 0, 0, 0, 0)$

Part X		$y+z_0$			
		0	**1**	**2**	**3**
z_1	**0**	0	0	−1	−2
	1	0	0	−1	−2
	2	0	0	−1	−2
	3	0	0	−1	−2

f) $\mathbf{a} = (3, 0, 0, 0, 0)$

Part X		$y+z_0$			
		0	**1**	**2**	**3**
z_1	**0**	0	0	−1	−2
	1	0	0	−1	−2
	2	0	0	−1	−2
	3	0	0	−1	−2

g) $\mathbf{a} = (3, 0, 0, 0, 0)$

three signals that are at end of the demand signal pipeline ($\mathbf{a} = (3, 0, 0, 0, 0)$) and available inventory is three ($x + z_0 = 3$) and pipeline stock is zero ($z_1 = 0$), the optimal decision is to return one to stock ($y^*(\mathbf{a}, \mathbf{z}) = 1$). That is, a part that was ordered for a signal can be returned before this signal is withdrawn. This is because when a signal is close to the end of its demand signal pipeline, the likelihood that it will materialize as a demand gets smaller and keeping an extra item due to the signal becomes more expensive than returning it and taking the risk of demand materialization. Finally, as we move through Tables 5(a) to 5(d) or 5(a) and 5(e) to 5(g), we can see that as the number of demand signals increases, the optimal order quantity increases (or return quantity decreases) with a slope less than one.

The optimal policy for W is simply to ignore the demand signals. This is because demand signals are too late to be met by regular replenishment and therefore an emergency shipment is required if and when the demand signal materializes. Since W is expensive, the optimal policy is to keep zero stock and, therefore, to meet demand by emergency shipments, which is the current situation in ASML for very expensive parts. Regarding part P, as long as returning excess stock is economically feasible, the optimal policy resembles that of part X. Otherwise, the optimal policy resembles that of part W: keep zero stock despite the demand signals. For part T, we have similar observations as parts X and P. Different from X and P, a signal almost always triggers an order since p is high. Also, different from P, an excess stock can be cleared naturally by demand occurrences since demand rate is high (an excess stock is cleared on average in 1/0.06 weeks). Therefore, the optimal policy does not regard whether an excess stock can be returned or not.

Apart from these observations, our findings for parts X, P, and T show that most of the time the local warehouse does not carry stock. A spare part is shipped to the local warehouse only if a demand signal is issued in the system (when returning is allowed, our observation holds even for low p). Note that this is exactly what a typical capital goods manufacturer, such as ASML, wants to have for the supply of its expensive spare parts that requires high availability. Using ADI makes it possible to control the inventories centrally by shipping a spare part to the warehouse when it is necessary. This implies a transition from a decentralized static inventory planning to a more centralized dynamic one where a spare parts supply network benefits also from lower operating costs and more risk pooling. Our findings show the potential value of imperfect ADI, which is very interesting for ASML in their process to discover the value of condition monitoring information. However, a structural redesign of the spare parts planning at the operational level is possible after ASML extends these predictions to more parts. Therefore, developing this redesign of the spare parts planning is another step that needs to be taken and would require additional research by ASML.

5. Conclusion

In this article, we investigate the benefit of using imperfect demand information (ADI). We consider three aspects of imperfection: false ADI (false positives), demand occurrences without ADI (false negatives), and timing of ADI. Using this setting, we propose a lost sales inventory model with a general representation of imperfect ADI that can apply to wide range of ADI applications in practice. We allow excess stock built up due to imperfections to be returned to an upstream supplier. First, we provide a partial characterization of the structure of the optimal ordering and return policy. We show that the optimal policy is dependent not only on on-hand stock but also on pipeline stock. The optimal order (return) size increases (decreases) with inventory levels with a slope less than one. Base stock policies and myopic policies, which are commonly used in practice, do not always perform well. Second, through an extensive computational study, we obtain several insights that can be used as input in design and improvement of inventory systems with imperfect ADI. We reveal that using imperfect ADI yields substantial savings while the amount of savings is sensitive to the levels of imperfectness aspects; having fewer false negatives is more desirable than having fewer false positives; returning excess inventory is quite effective in coping with the consequences of false ADI, particularly for slow-moving items. Third, we apply our model to a spare parts case. Our analysis reveals that provided that ADI is timely, a typical manufacturer subject to low demand can keep minimum, most of the time zero, stock while still maintaining the responsiveness to customers; using imperfect ADI leads to a transition from a decentralized static inventory system to a more centralized dynamic one, where spare parts are mainly stored at the central warehouse and are shipped to the customer only when there is a demand signal.

This article has two main limitations. First, although computing the optimal policy was not a burden in our setting, it might easily become intractable in applications with higher demand volumes. Second, we did not analyze the effect of demand signals on the optimal policy, which might be important in many applications. However, the analysis of the effect of demand signals on the optimal policy would be complicated since demand signals affect demand expectations in multiple periods, each with a different rate. These two points deserve further attention.

Acknowledgments

The authors would like to thank Gert Streutker and Ruud van Sommeren from ASML for the discussions that formed the basis of this study.

Funding

This research is part of the project on Proactive Service Logistics for Advanced Capital Goods and has been sponsored by the TKI-Dutch Institute for Advanced Logistics.

Notes on contributors

Engin Topan is an assistant professor of reliability, maintenance, and service logistics at the University of Twente. He received his Ph.D. from the Middle East Technical University, Turkey, in 2010. He worked at different positions at Cankaya University, Turkish Military Academy, and Eindhoven University of Technology and also took part as a researcher in projects on proactive service logistics for capital goods (ProSeLo, ProSeLoNext) and proactive maintenance planning (MANTIS). His research interests include spare parts management, maintenance, and service logistics.

Tarkan Tan is an associate professor of sustainable operations management at the Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology. He received his Ph.D. from the Middle East Technical University, Ankara, Turkey, in 2002. He spent a year at Columbia University, New York, as a Fulbright Scholar. He joined Eindhoven University of Technology in 2003 and he is currently the director of the operations management and logistics M.Sc. program at this university. He has visited the University of California, Los Angeles, and the University of Sydney as a visiting scholar. He is an executive board member of the European Supply Chain Forum. His research interests include inventory theory, capacity management, spare parts management, and supply chain management with a particular focus on sustainability.

Geert-Jan van Houtum is a professor of maintenance and reliability at the Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology. His research is focused on spare parts management, production–inventory systems, maintenance and availability management of capital goods, and the effect of design decisions on the total cost of ownership of capital goods. Much of his research is based on a close collaboration with, for example, ASML, Dutch Railways, Philips, and the Royal Airforce. He is associate editor of *Manufacturing and Service Operations Management*, *OR Spectrum*, and *Flexible Services and Manufacturing*.

Rommert Dekker is a professor of operations research, quantitative logistics, and IT at the Erasmus School of Economics. He currently leads an industry-sponsored research program on service logistics. Professor Dekker began his career working at Shell Research. During his 7 years at the company he published over 100 papers on topics including reverse logistics, service logistics, inventory control, maintenance optimization, container logistics, and transport optimization. He was also co-founder of the well-known network for reverse logistics REVLOG.

References

Altman, E., Gaujal, B. and Hordijk, A. (2000) Multimodularity, convexity, and optimization properties. *Mathematics of Operations Research*, **25**(2), 324–347.

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

Bernstein, F. and DeCroix, G.A. (2015) Advance demand information in a multiproduct system. *Manufacturing and Service Operations Management*, **17**(1), 52–65.

Bertsekas, D.P. and Tsitsiklis, J.N. (1996) *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA.

Bijvank, M. and Vis, IFA (2011) Lost-sales inventory theory: A review. *European Journal of Operational Research*, **215**(3), 1–13.

Buzacott, J.A. and Shanthikumar, J.G. (1994) Safety stock versus safety time in MRP controlled production systems. *Management Science*, **40**(12), 1678–1689.

Deshpande, V., Iyer, A.V. and Cho, R. (2006) Efficient supply chain management at the U.S. Coast Guard using part-age dependent supply replenishment policies. *Operations Research*, **54**(6), 1028–1040.

Donselaar, K.H. van Kopczak, L.R. and Wouters, M.J.F. (2001) The use of advance demand information in a project-based supply chain. *European Journal of Operational Research*, **130**(3), 519–538.

Elwany, A.H. and Gebraeel, N.Z. (2008) Sensor-driven prognostic models for equipment replacement and spare parts inventory. *IIE Transactions*, **40**, 629–639.

Eppen, G.D. and Fama, E.F. (1969) Cash balance and simple dynamic portfolio problems with proportional costs. *International Economic Review*, **10**(2), 119–133.

Fukuda, Y. (1961) Optimal disposal policies. *Naval Research Logistics Quarterly*, **8**(3), 221–227.

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

Gao, L., Xu, S.H. and Ball, M.O. (2012) Managing an available-to-promise assembly system with dynamic short-term pseudo-order forecast. *Management Science*, **58**(4), 770–790.

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

Güllü, R. (1996) On the value of information in dynamic production inventory problems under forecast evolution. *Naval Research Logistics*, **43**, 289–303.

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

Hausman, W.H. (1969) Sequential decision problems: A model to exploit existing forecasters. *Management Science*, **16**, B93–B111.

Heath, D.C. and Jackson, P.L. (1994) Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions*, **26**, 17–30.

Johansen, S.G. (2001) Pure and modified base-stock policies for the lost sales inventory system with negligible set-up costs and constant lead times. *International Journal of Production Economics*, **71**, 391–399.

Karaesmen, F. (2013) Value of advance demand information in production and inventory systems with shared resources, in J. M. Smith, and B. Tan, eds., *Handbook of Stochastic Models in Manufacturing System Operations*, Springer, New York, NY, pp. 139–165.

Karlin, S. and Scarf, H. (1958) Inventory models of the Arrow-Harris-Marschak type with time lag, in Arrow, S. Karlin, H. Scarf, eds., *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford, CA, pp. 155–178.

Levi, R., Pál, M., Roundy, R. O., and Shmoys, D.B. (2007) Approximation algorithms for stochastic inventory control models. *Mathematics of Operations Research*, **32**, 284–302.

Li, Q. and Yu, P. (2014) Multimodularity and its applications in three stochastic dynamic inventory problems. *Manufacturing and Service Operations Management*, **16**(3), 455–463.

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

Liberopoulos, G. and Koukoumialos, S. (2008) On the effect of variability and uncertainty in advance demand information on the performance of a make-to-stock supplier. *MIBES Transactions International Journal*, **2**(1), 95–114.

Lin, X., Basten, R.J.I., Kranenburg, A.A. and van Houtum, G.J. (2017) Condition based spare parts supply. *Reliability Engineering and System Safety* **168**, 240–248.

Louit, D., Pascual, R., Banjevic, D. and Jardine, A.K.S. (2011) Condition based spares ordering for critical components. *Mechanical Systems and Signal Processing*, **25**, 1837–1848.

Morton, T.E. (1969) Bounds on the solution of the lagged optimal inventory equation with no demand backlogging and proportional costs. *SIAM Review*, **11**(4), 572–596.

Morton, T.E. (1971) The near-myopic nature of the lagged-proportional cost inventory problem with lost sales. *Operations Research*, **19**(7), 1708–1716.

Muckstadt, J.A. (2005) *Analysis and Algorithms for Service Parts Supply Chains*, Springer, New York, NY.

Murota, K. (2003) *Discrete Convex Analysis*, SIAM, Philadelphia, PA.

Olson, D. L., Delen, D. (2008) *Advanced Data Mining Techniques*, Springer, Heidelberg.

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

Pang, Z., Chen, F.Y. and Feng, Y. (2012) Technical Note: A note on the structure of joint inventory-pricing control with leadtimes. *Operations Research*, **60**(3), 581–587.

Puterman, M. (1994) *Markov Decision Processes*, John Wiley & Sons, New York, NY.

Song, J.S. and Zipkin, P.H. (2012) Newsvendor problems with sequentially revealed demand information. *Naval Research Logistics*, **59**, 601–612.

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

Tan, T., Güllü, R. and Erkip, N.K. (2009) Using imperfect advance demand information in ordering and rationing decisions. *International Journal of Production Economics*, **121**, 665–677.

Thonemann, U.W. (2002) Improving supply-chain performance by sharing advance demand information. *European Journal of Operational Research*, **142**(1), 81–107.

Toktay, L.B. and Wein, L.M. (2001) Analysis of a forecasting-production-inventory system with stationary demand. *Management Science*, **47**, 1268–1281.

Topkis, D.M. (1998) *Supermodularity and Complementarity*, Princeton University Press, Princeton, NJ.

Wang, Y. and Tomlin, B. (2009) To wait or not to wait: Optimal ordering under lead time uncertainty and forecast updating. *Naval Research Logistics*, **56**, 766–779.

Zhu, K. and Thonemann, U.W. (2004) An adaptive forecasting algorithm and inventory policy for products with short life cycles. *Naval Research Logistics*, **51**, 633–653.

Zipkin, P.H. (2008a) Old and new methods for lost-sales inventory systems. *Operations Research*, **56**(5), 1256–1263.

Zipkin, P.H. (2008b) On the structure of lost-sales inventory models. *Operations Research*, **56**(4), 937–944.

Appendix

A.1. Proofs

Proof of Lemma 1. Let $(\mathbf{a}, \mathbf{z}) \in \mathcal{U} \times \mathcal{Z}$ and z_L and y both be strictly positive. Then, by reducing each z_L and y by one unit, keeping that one unit in stock as a reserved stock for L periods and releasing it after L periods, we have one return less, we order one unit less, and we have an extra part on stock for L periods and therefore less likelihood of shortage and for the rest everything remains the same. This reduces the costs at least by $c + c_r - h \times L \geq 0$. This shows that $(z_L - 1, y - 1)$ is at least equally as good as (z_L, y). Note that (z_L, y) cannot be a smallest minimizer of Equation (1). □

Proof of Lemma 2. Let $\mathbf{x} \in \mathbf{M}$. Assume $g : \mathbf{M} \times \mathbf{N} \to \mathbb{R}$ is $L^\natural$-convex in $\mathbf{y} \in \mathbf{N}$ for $\mathbf{x}$. Also, let $l \leq y_j$ for all $j = 1, \ldots, n, l \leq \varepsilon$, and $\widetilde{\mathbf{N}} = \{(\mathbf{y}, \varepsilon, l) \in \overline{\mathbf{N}} \times \mathbb{N}_0 : l \leq y_j \forall j; l \leq \varepsilon\}$. By Definition 2, we need to show that $\phi(\mathbf{x}, \mathbf{y}, \varepsilon, l) = \psi(\mathbf{x}, \mathbf{y} - l\mathbf{e}, \varepsilon - l) : \mathbf{M} \times \widetilde{\mathbf{N}} \to \mathbb{R}$ is submodular in $(\mathbf{y}, \varepsilon, l) \in \widetilde{\mathbf{N}}$ for $\mathbf{x}$. First, we note that $\widetilde{\mathbf{N}}$ is a sublattice (of $\overline{\mathbf{N}} \times \mathbb{N}_0$) since constraints $l - y_j \leq 0$ and $l - \varepsilon \leq 0$ have at most two variables with opposite signs (see Topkis (1998), Example 2.2.7(b)). Also, note that

$$\begin{aligned} \phi(\mathbf{x}, \mathbf{y}, \varepsilon, l) &= \psi(\mathbf{x}, \mathbf{y} - l\mathbf{e}, \varepsilon - l) = g(\mathbf{x}, \mathbf{y} - l\mathbf{e} - (\varepsilon - l)\mathbf{e}) \\ &= g(\mathbf{x}, \mathbf{y} - \varepsilon\mathbf{e}) = \psi(\mathbf{x}, \mathbf{y}, \varepsilon). \end{aligned} \tag{A1}$$

Since $g(\mathbf{x}, \mathbf{y})$ is $L^\natural$-convex in $\mathbf{y}$ for $\mathbf{x}$, by Definition 2, $\psi(\mathbf{x}, \mathbf{y}, \varepsilon) = g(\mathbf{x}, \mathbf{y} - \varepsilon\mathbf{e})$ is submodular in $(\mathbf{y}, \varepsilon) \in \overline{\mathbf{N}}$ for $\mathbf{x}$. Then, from Equation (A1), $\phi(\mathbf{x}, \mathbf{y}, \varepsilon, l)$ is submodular in $(\mathbf{y}, \varepsilon, l) \in \widetilde{\mathbf{N}}$ for $\mathbf{x}$. □

Proof of Lemma 3. Let $\mathbf{x} \in \mathbf{M}$. Assume $h : \mathbf{M} \times \widehat{\mathbf{N}} \to \mathbb{R}$ is $L^\natural$-convex in $(\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}$ for $\mathbf{x}$. We want to show that $g(\mathbf{x}, \mathbf{y}) = \min_{\boldsymbol{\xi}:(\mathbf{y},\boldsymbol{\xi}) \in \widehat{\mathbf{N}}} \{h(\mathbf{x}, \mathbf{y}, \boldsymbol{\xi})\}$ is $L^\natural$-convex in $\mathbf{y} \in \mathbf{N}$ for $\mathbf{x}$. By Definition 2, it suffices to show that

$$\omega(\mathbf{x}, \mathbf{y}, l) = g(\mathbf{x}, \mathbf{y} - l\mathbf{e}) = \min_{\boldsymbol{\xi}:(\mathbf{y}-l\mathbf{e},\boldsymbol{\xi}) \in \widehat{\mathbf{N}}} \left\{h(\mathbf{x}, \mathbf{y} - l\mathbf{e}, \boldsymbol{\xi})\right\},$$

is submodular in $(\mathbf{y}, l) \in \overline{\mathbf{N}}' = \{(\mathbf{y}, l) \in \mathbf{N} \times \mathbb{N}_0 : l \leq y_j \forall j\}$ for $\mathbf{x}$. Note that

$$\begin{aligned} \omega(\mathbf{x}, \mathbf{y}, l) = g(\mathbf{x}, \mathbf{y} - l\mathbf{e}) &= \min_{\boldsymbol{\xi}:(\mathbf{y}-l\mathbf{e},\boldsymbol{\xi}) \in \widehat{\mathbf{N}}} \left\{h(\mathbf{x}, \mathbf{y} - l\mathbf{e}, \boldsymbol{\xi})\right\} \\ &= \min_{\boldsymbol{\xi}:(\mathbf{y}-l\mathbf{e},\boldsymbol{\xi}) \in \widehat{\mathbf{N}}} \left\{h(\mathbf{x}, \mathbf{y} - l\mathbf{e}, (l\mathbf{e} + \boldsymbol{\xi}) - l\mathbf{e})\right\}. \end{aligned}$$

Let $\bar{\boldsymbol{\xi}} = l\mathbf{e} + \boldsymbol{\xi}$. Then, $\bar{\boldsymbol{\xi}} \geq l\mathbf{e}$, i.e., $\bar{\xi}_k \geq l \, \forall k$—therefore, $\bar{\boldsymbol{\xi}} - l\mathbf{e} \in \mathbf{U}$ and $(\mathbf{y} - l\mathbf{e}, \bar{\boldsymbol{\xi}} - l\mathbf{e}) \in \widehat{\mathbf{N}}$. Hence, we can write

$$\omega(\mathbf{x}, \mathbf{y}, l) = g(\mathbf{x}, \mathbf{y} - l\mathbf{e}) = \min_{\bar{\boldsymbol{\xi}}:(\mathbf{y}-l\mathbf{e},\bar{\boldsymbol{\xi}}-l\mathbf{e}) \in \widehat{\mathbf{N}}} \{h(\mathbf{x}, \mathbf{y} - l\mathbf{e}, \bar{\boldsymbol{\xi}} - l\mathbf{e})\}.$$

Now, let $\widetilde{\mathbf{N}} = \{(\mathbf{y}, \bar{\boldsymbol{\xi}}, l) \in \widehat{\mathbf{N}} \times \mathbb{N}_0 : l \leq y_j \ \forall j; l \leq \bar{\xi}_k \ \forall k\}$. Note that $\widetilde{\mathbf{N}}$ involves constraints with two variables having opposite signs; therefore, it is a sublattice (of $\widehat{\mathbf{N}} \times \mathbb{N}_0$) for $\mathbf{x}$. Since $h : \mathbf{M} \times \widehat{\mathbf{N}} \to \mathbb{R}$ is $L^\natural$-convex in $(\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}$ for $\mathbf{x}$, $\varsigma(\mathbf{x}, \mathbf{y}, \bar{\boldsymbol{\xi}}, l) = h(\mathbf{x}, \mathbf{y} - l\mathbf{e}, \bar{\boldsymbol{\xi}} - l\mathbf{e})$ is $L^\natural$-convex in $(\mathbf{y}, \bar{\boldsymbol{\xi}}, l) \in \widetilde{\mathbf{N}}$ for $\mathbf{x}$ (by Lemma 2), and hence also submodular (by Property 2) in $(\mathbf{y}, \bar{\boldsymbol{\xi}}, l) \in \widetilde{\mathbf{N}}$ for $\mathbf{x}$. By preservation of submodularity under minimization (Topkis, 1998, Theorem 2.7.6), we conclude that

$$\begin{aligned} \omega(\mathbf{x}, \mathbf{y}, l) = g(\mathbf{x}, \mathbf{y} - l\mathbf{e}) &= \min_{\bar{\boldsymbol{\xi}}:(\mathbf{y}-l\mathbf{e},\bar{\boldsymbol{\xi}}-l\mathbf{e}) \in \widehat{\mathbf{N}}} \{h(\mathbf{x}, \mathbf{y} - l\mathbf{e}, \bar{\boldsymbol{\xi}} - l\mathbf{e})\} \\ &= \min_{\bar{\boldsymbol{\xi}}:(\mathbf{y},\bar{\boldsymbol{\xi}},l) \in \widetilde{\mathbf{N}}} \{\varsigma(\mathbf{x}, \mathbf{y}, \bar{\boldsymbol{\xi}}, l)\} \end{aligned}$$

is submodular in $(\mathbf{y}, l) \in \overline{\mathbf{N}}'$ for $\mathbf{x}$. □

Proof of Lemma 4. Let $\mathbf{x} \in \mathbf{M}$. Assume $h : \mathbf{M} \times \widehat{\mathbf{N}} \to \mathbb{R}$ is $L^\natural$-convex in $(\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}$ and $\boldsymbol{\xi}^*(\mathbf{x}, \mathbf{y})$ is the smallest vector solution of $\min_{\boldsymbol{\xi}:(\mathbf{y},\boldsymbol{\xi}) \in \widehat{\mathbf{N}}} \{h(\mathbf{x}, \mathbf{y}, \boldsymbol{\xi})\}$ for $\mathbf{x}$.

Part (a). From Property 2, $h(\mathbf{x}, \mathbf{y}, \boldsymbol{\xi})$ is submodular in $(\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}$ for $\mathbf{x}$. This implies that $\xi_i^*(\mathbf{x}, \mathbf{y})$ is increasing in $\mathbf{y} \in \mathbf{N}$ for $\mathbf{x}$ (Topkis, 1998, Theorem 2.8.2).

Part (b). The inequality on the left is due to part (a). The proof of the inequality on the right is by contradiction. Let $(\mathbf{x}, \mathbf{y}) \in \mathbf{M} \times \mathbf{N}$ and $k \in \mathbb{N}_0^+$. Let ξ_i and $\xi_i^*(\mathbf{x}, \mathbf{y})$ be the ith argument of $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^*(\mathbf{x}, \mathbf{y})$ (the latter is the smallest minimizer of $h(\mathbf{x}, \mathbf{y}, \boldsymbol{\xi})$), respectively, for $i = 1, \ldots, u$. Also, let $\zeta_i(\mathbf{x}, \mathbf{y}, \xi_i) = \min_{\xi_k, \forall k \neq i : (\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}} \{h(\mathbf{x}, \mathbf{y}, \boldsymbol{\xi})\}$. Then, $\xi_i^*(\mathbf{x}, \mathbf{y})$ is a minimizer of $\zeta_i(\mathbf{x}, \mathbf{y}, \xi_i)$. The rest holds for each $i = 1, \ldots, u$. Assume an arbitrary $\bar{\xi}_i > \xi_i^*(\mathbf{x}, \mathbf{y}) + k$. Then, the proof will be complete if we show that such an arbitrary $\bar{\xi}_i$ cannot be the smallest minimizer of $\zeta_i(\mathbf{x}, \mathbf{y} + k\mathbf{e}, \xi_i)$ for $\mathbf{x}$, $\mathbf{y}$, and k. Let $\psi_i(\mathbf{x}, \mathbf{y}, \xi_i, l) = \zeta_i(\mathbf{x}, \mathbf{y} - l\mathbf{e}, \xi_i - l)$. Note that $h(\mathbf{x}, \mathbf{y}, \boldsymbol{\xi})$ is $L^\natural$-convex in $(\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}$ for $\mathbf{x}$. By Lemma 3, $\zeta_i(\mathbf{x}, \mathbf{y}, \xi_i) = \min_{\xi_k, \forall k \neq i : (\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}} \{h(\mathbf{x}, \mathbf{y}, \boldsymbol{\xi})\}$ is $L^\natural$-convex in $(\mathbf{y}, \xi_i) \in \{(\mathbf{y}, \xi_i) : (\mathbf{y}, \boldsymbol{\xi}) \in \widehat{\mathbf{N}}\}$ for $\mathbf{x}$. Therefore, $\psi_i(\mathbf{x}, \mathbf{y}, \xi_i, l) = \zeta_i(\mathbf{x}, \mathbf{y} - l\mathbf{e}, \xi_i - l)$ is $L^\natural$-convex (Lemma 2) and hence also submodular (Property 2) in $(\mathbf{y}, \xi_i, l) \in \{(\mathbf{y}, \xi_i, l) : (\mathbf{y}, \boldsymbol{\xi}, l) \in \widehat{\mathbf{N}} \times \mathbb{N}_0; l \le y_j \; \forall j; l \le \xi_i\}$ for $\mathbf{x}$. Then, we can write

$$\psi_i(\mathbf{x}, \mathbf{y} + k\mathbf{e}, \bar{\xi}_i, 0) + \psi_i(\mathbf{x}, \mathbf{y} + k\mathbf{e}, \xi_i^*(\mathbf{x}, \mathbf{y}) + k, k) \\ \ge \psi_i(\mathbf{x}, \mathbf{y} + k\mathbf{e}, \xi_i^*(\mathbf{x}, \mathbf{y}) + k, 0) + \psi_i(\mathbf{x}, \mathbf{y} + k\mathbf{e}, \bar{\xi}_i, k),$$

which is equivalent to

$$\zeta_i(\mathbf{x}, \mathbf{y} + k\mathbf{e}, \bar{\xi}_i) - \zeta_i(\mathbf{x}, \mathbf{y} + k\mathbf{e}, \xi_i^*(\mathbf{x}, \mathbf{y}) + k) \\ \ge \zeta_i(\mathbf{x}, \mathbf{y}, \bar{\xi}_i - k) - \zeta_i(\mathbf{x}, \mathbf{y}, \xi_i^*(\mathbf{x}, \mathbf{y})). \quad \text{(A2)}$$

Since $\xi_i^*(\mathbf{x}, \mathbf{y})$ is the smallest minimizer of $\zeta_i(\mathbf{x}, \mathbf{y}, \xi_i)$, the term on the right-hand side of Equation (A2) is non-negative. Hence, $\zeta_i(\mathbf{x}, \mathbf{y} + k\mathbf{e}, \bar{\xi}_i) - \zeta_i(\mathbf{x}, \mathbf{y} + k\mathbf{e}, \xi_i^*(\mathbf{x}, \mathbf{y}) + k) \ge 0$. This inequality shows that $\bar{\xi}_i$ cannot be the smallest minimizer of $\zeta_i(\mathbf{x}, \mathbf{y} + k\mathbf{e}, \xi_i)$ for $\mathbf{x}$, $\mathbf{y}$, and k. Note that this holds for all $i = 1, \ldots, u$.

Part (c). This follows directly from part (b).

□

Proof of Theorem 1.

Parts (a) and (b). The proof is by induction on t. Since $\bar{f}_{T+1}(\mathbf{a}, \mathbf{v}) = 0$, the result certainly holds for $t = T+1$ and each $\mathbf{a} \in \mathcal{U}$. Then, given the induction hypothesis that $\bar{f}_{t+1}(\mathbf{a}, \mathbf{v})$, $1 \le t \le T$, is $L^\natural$-convex in $\mathbf{v} \in \mathcal{V}$ for each $\mathbf{a} \in \mathcal{U}$, we will prove that $\bar{f}_t(\mathbf{a}, \mathbf{v}, v_L, y)$ is $L^\natural$-convex in $(\mathbf{v}, v_L, y) \in \mathcal{Q}$ for each $\mathbf{a} \in \mathcal{U}$ (Theorem 1(a)) and $\bar{f}_t(\mathbf{a}, \mathbf{v})$ is $L^\natural$-convex in $\mathbf{v} \in \mathcal{V}$ for $\mathbf{a}$ for each $\mathbf{a} \in \mathcal{U}$ (Theorem 1(b)).

For convenience, we consider a modified problem that is equivalent to the original problem denoted by Equations (2) and (3). The problem is reformulated as a two-step nested optimization problem where the inner problem is trivial and its optimal solution is obvious. For the inner problem, we suppose that the ordering and returning decisions v_L and y have been made (but we keep them as variables) and the values of the predicted demand $\mathbf{r} = (r_{\tau_u}, \ldots, r_0) \in \mathcal{R}$ with $\mathcal{R} = \{\mathbf{r} \in \mathcal{U} : \mathbf{r} \le \mathbf{a}\}$, unpredicted demand $d^u \in \mathbb{N}_0$, and new signals $w \in \mathbb{N}_0$ are all observed. The problem is to decide how much demand has to be fulfilled at the end of period t. Let $u \in \mathbb{N}_0$ denote this amount; i.e., the demand to be satisfied. Also, let

$$d = d^u + \sum_{\tau=\tau_l}^{\tau_u} r_\tau \in \mathbb{N}_0, \text{ and } \bar{\mathbf{r}} = (r_{\tau_u - 1}, \ldots, r_0) \in \bar{\mathcal{R}}$$

with $\bar{\mathcal{R}} = \{\bar{\mathbf{r}} \in \mathbb{N}_0^{\tau_u} : \bar{\mathbf{r}} \le \bar{\mathbf{a}}\}$ where $\bar{\mathbf{a}} \in \mathbb{N}_0^{\tau_u}$ and it is defined as in Problem (1). The inner problem is formulated by

$$K_t\left(\mathbf{a}, \mathbf{v}, v_L, y, d^u, \mathbf{r}, w\right) \\ = \min_u \{\psi(\mathbf{a}, \mathbf{v}, v_L, y, u, d^u, \mathbf{r}, w) = c(v_L - v_{L-1}) \\ + h(v_0 - y - u) + c_e(d - u) \\ + c_r y + \bar{f}_{t+1}(\bar{\mathbf{a}} - \bar{\mathbf{r}}, w, v_0 - y - u, \ldots, v_L - y - u) : \\ u \in \mathbb{N}_0, \; u \le d, \quad u \le v_0 - y\}, \quad \text{(A3)}$$

for each $\mathbf{a} \in \mathcal{U}$, $(\mathbf{v}, v_L, y) \in \mathcal{Q}$ and $\mathbf{r} \in \mathcal{R}$, $d^u \in \mathbb{N}_0$, $w \in \mathbb{N}_0$. The optimal decision for this problem will be to satisfy demand to the maximum extent possible; i.e., $u^* = \min\{d, v_0 - y\}$. Otherwise, one would reserve stock for the next period's demand while applying an emergency supply for (some of) the current period demand. Note that such a solution is never optimal for our optimization problem as a whole and also not for the inner problem. To simplify Equation (A3) further, we define $v_0^+ = u + y \in \mathbb{N}_0$, which denotes the amount deducted from stock by either fulfilling demand or returning extra stock. We use v_0^+ to eliminate variable u by simply replacing u with $v_0^+ - y \in \mathbb{N}_0$. Then, the inner problem is restated as

$$K_t\left(\mathbf{a}, \mathbf{v}, v_L, y, d^u, \mathbf{r}, w\right) \\ = \min_{v_0^+} \{\psi(\mathbf{a}, \mathbf{v}, v_L, y, v_0^+, d^u, \mathbf{r}, w) = c(v_L - v_{L-1}) \\ + h(v_0 - v_0^+) + c_e(d - (v_0^+ - y)) \\ + c_r y + \bar{f}_{t+1}(\bar{\mathbf{a}} - \bar{\mathbf{r}}, w, v_0 - v_0^+, \ldots, v_L - v_0^+) : \\ v_0^+ \in \mathbb{N}_0, v_0^+ \le y + d, v_0^+ \le v_0\}$$

for each $\mathbf{a} \in \mathcal{U}$, $(\mathbf{v}, v_L, y) \in \mathcal{Q}$ and $\mathbf{r} \in \mathcal{R}$, $d^u \in \mathbb{N}_0$, $w \in \mathbb{N}_0$. Let

$$\widehat{\mathcal{V}} = \{(\mathbf{v}, v_L, y, v_0^+) \in \mathcal{Q} \times \mathbb{N}_0 : v_0^+ \le y + d, v_0^+ \le v_0\},$$

and also let $\bar{\mathbf{v}} = (v_0, \ldots, v_L) \in \mathcal{V}$ and

$$\bar{\mathcal{V}} = \{(\bar{\mathbf{v}}, v_0^+) \in \mathcal{V} \times \mathbb{N}_0 : v_0^+ \le v_0\}.$$

Note that $\widehat{\mathcal{V}}$ and $\bar{\mathcal{V}}$ are sublattices (of $\mathcal{Q} \times \mathbb{N}_0$ and $\mathcal{V} \times \mathbb{N}_0$, respectively) since all constraints have at most two variables with opposite signs; see Topkis (1998), Example 2.2.7(b). By the induction hypothesis, $\bar{f}_{t+1}(\mathbf{a}, \mathbf{v})$ is $L^\natural$-convex in $\mathbf{v} \in \mathcal{V}$ for each $\mathbf{a} \in \mathcal{U}$. Note also that $\bar{\mathbf{v}} - v_0^+\mathbf{e} \in \mathcal{V}$ (Remark 1 and Remark 2). Then, by Lemma 2, $\phi(\mathbf{a}, \bar{\mathbf{v}}, v_0^+) = \bar{f}_{t+1}(\mathbf{a}, \bar{\mathbf{v}} - v_0^+\mathbf{e})$ is $L^\natural$-convex in $(\bar{\mathbf{v}}, v_0^+) \in \bar{\mathcal{V}}$ for each $\mathbf{a} \in \mathcal{U}$. Consequently, $\phi(\bar{\mathbf{a}} - \bar{\mathbf{r}}, w, \bar{\mathbf{v}}, v_0^+) = \bar{f}_{t+1}(\bar{\mathbf{a}} - \bar{\mathbf{r}}, w, \bar{\mathbf{v}} - v_0^+\mathbf{e})$ is $L^\natural$-convex in $(\bar{\mathbf{v}}, v_0^+) \in \bar{\mathcal{V}}$ for each $\bar{\mathbf{a}} \in \mathbb{N}_0^{\tau_u}$ and $\bar{\mathbf{r}} \in \bar{\mathcal{R}}$ (hence for $\bar{\mathbf{a}} - \bar{\mathbf{r}} \in \mathcal{U}$), $w \in \mathbb{N}_0$. The remaining terms that define $\psi(\mathbf{a}, \mathbf{v}, v_L, y, v_0^+, d^u, \mathbf{r}, w)$ are separable and linear; therefore, they are $L^\natural$-convex in $\{(v_0, y, v_0^+) \in \mathbb{N}_0^3 : 0 \le v_0^+ \le y + d, v_0^+ \le v_0\}$ for $d \in \mathbb{N}_0$. Therefore, $\psi(\mathbf{a}, \mathbf{v}, v_L, y, v_0^+, d^u, \mathbf{r}, w)$ is $L^\natural$-convex in $(\mathbf{v}, v_L, y, v_0^+) \in \widehat{\mathcal{V}}$ for each $\mathbf{a} \in \mathcal{U}$ and $\mathbf{r} \in \mathcal{R}$, $d^u \in \mathbb{N}_0$,

$w \in \mathbb{N}_0$. By Lemma 3, $K_t(\mathbf{a}, \mathbf{v}, v_L, y, d^u, \mathbf{r}, w)$ is $L^\natural$-convex in $(\mathbf{v}, v_L, y) \in \mathcal{Q}$ for each $\mathbf{a} \in \mathcal{U}$ and $\mathbf{r} \in \mathcal{R}$, $d^u \in \mathbb{N}_0$, $w \in \mathbb{N}_0$. In the first step, we assumed that values of the random variables $\mathbf{R}$, D^u, and W in period t are known. In the second step, we remove this assumption and we rewrite Equation (3) as

$$\bar{J}_t\left(\mathbf{a}, \mathbf{v}, v_L, y\right) = E[K_t\left(\mathbf{a}, \mathbf{v}, v_L, y, D^u, \mathbf{R}, W\right)].$$

Note that Equation (2) remains the same. It follows from Property 3 that $\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y)$ is also $L^\natural$-convex in $(\mathbf{v}, v_L, y) \in \mathcal{Q}$ for each $\mathbf{a} \in \mathcal{U}$. By Lemma 3, we find that $\bar{f}_t(\mathbf{a}, \mathbf{v})$ is $L^\natural$-convex in $\mathbf{v} \in \mathcal{V}$ for each $\mathbf{a} \in \mathcal{U}$.

Part (c). By Theorem 1(a), $\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y)$ is component-wise convex in $v_L \in \mathbb{N}_0$ and $y \in \mathbb{N}_0$; i.e., $\Delta_{v_L}\Delta_{v_L}\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y) \geq 0$ and $\Delta_y\Delta_y\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y) \geq 0$, respectively, for all t, $0 \leq t \leq T$. By using $\bar{J}_t(\mathbf{a}, v_{-1}, \dots, v_L, y) = J_t(\mathbf{a}, v_{-1}, v_0 - v_{-1}, \dots, v_L - v_{L-1}, y)$, we also have

$$\Delta_{z_L}\Delta_{z_L}J_t\left(\mathbf{a}, \mathbf{z}, z_L, y\right) = \Delta_{v_L}\Delta_{v_L}\bar{J}_t\left(\mathbf{a}, v_{-1}, \dots, v_L, y\right) \geq 0$$

and

$$\Delta_y\Delta_yJ_t\left(\mathbf{a}, \mathbf{z}, z_L, y\right) = \Delta_y\Delta_y\bar{J}_t\left(\mathbf{a}, v_{-1}, \dots, v_L, y\right) \geq 0.$$

Part (d). Note that $f : \mathcal{U} \times \mathcal{Z} \to \mathbb{R}$ is represented as

$$\begin{aligned} &f_t\left(\mathbf{a}, x, z_0, \dots, z_{L-1}\right) \\ &\quad = \bar{f}_t\left(\mathbf{a}, x, x + z_0, x + z_0 + z_1, \dots, x + z_0 + \dots + z_{L-1}\right), \end{aligned}$$

and $\bar{f}_t : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ is $L^\natural$-convex for $\mathbf{a} \in \mathcal{U}$. Then, it follows from page 183 of Murota (2003) that $f : \mathcal{U} \times \mathcal{Z} \to \mathbb{R}$ is multimodular for $\mathbf{a} \in \mathcal{U}$. Having increasing differences and component-wise convexity is a direct consequence of multimodularity (Lemma 2.2.b.ii of Altman *et al.*, 2000). □

Proof of Lemma 5. The proof is by contradiction. Let $(\mathbf{a}, \mathbf{z}) \in \mathcal{U} \times \mathcal{Z}$ and $t = 1, \dots, T$. Let $(z_L^*(\mathbf{a}, \mathbf{z}), 0)$ be a smallest vector minimizer of $J_t(\mathbf{a}, \mathbf{z}, z_L, y)$ with $z_L^*(\mathbf{a}, \mathbf{z}) > 0$ (and also by definition $z_L^*(\mathbf{a}, \mathbf{z}) \in \mathbb{N}_0$). Let $(0, \bar{y})$ be an alternative smallest vector minimizer of $J_t(\mathbf{a}, \mathbf{z}, z_L, y)$ with $\bar{y} > 0$ (and also by definition $\bar{y} \in \mathbb{N}_0, \bar{y} \leq x$). The proof will be complete if we show that $(0, \bar{y})$ cannot be a smallest vector minimizer. (Note that, by Lemma 1 and because we are interested only in smallest vector minimizers, this is the only form that can express two alternative smallest minimizers.) Here, first we apply our transformation $v_l = x + \sum_{t=0}^{l} z_t$ for $l = -1, \dots, L$, $\mathbf{v} = (v_{-1}, \dots, v_{L-1})$ and $v_L^*(\mathbf{a}, \mathbf{v}) = v_{L-1} + z_L^*(\mathbf{a}, \mathbf{z})$. From Theorem 1(a) and Property 2, $\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y)$ is submodular in $(\mathbf{v}, v_L, y)$. Then by Definition 1, we write

$$\begin{aligned} &\bar{J}_t\left(\mathbf{a}, \mathbf{v}, v_{L-1}, \bar{y})\right) - \bar{J}_t\left(\mathbf{a}, \mathbf{v}, v_{L-1}, 0\right) \geq \bar{J}_t\left(\mathbf{a}, \mathbf{v}, v_{L-1} + z_L^*(\mathbf{a}, \mathbf{z}), \bar{y}\right) \\ &\quad - \bar{J}_t\left(\mathbf{a}, \mathbf{v}, v_{L-1} + z_L^*(\mathbf{a}, \mathbf{z}), 0\right). \end{aligned} \tag{A4}$$

Since $(v_{L-1} + z_L^*(\mathbf{a}, \mathbf{z}), 0)$ is a minimizer of $\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y)$, we have

$$\bar{J}_t\left(\mathbf{a}, \mathbf{v}, v_{L-1} + z_L^*(\mathbf{a}, \mathbf{z}), \bar{y}\right) - \bar{J}_t\left(\mathbf{a}, \mathbf{v}, v_{L-1} + z_L^*(\mathbf{a}, \mathbf{z}), 0\right) \geq 0.$$

Then from Equation (A4), we also have

$$\bar{J}_t\left(\mathbf{a}, \mathbf{v}, v_{L-1}, \bar{y})\right) - \bar{J}_t\left(\mathbf{a}, \mathbf{v}, v_{L-1}, 0\right) \geq 0.$$

This shows that $(v_{L-1}, \bar{y})$ cannot be a smallest vector minimizer of $\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y)$. Note that this also indicates that $(0, \bar{y})$ cannot be a smallest vector minimizer of $J_t(\mathbf{a}, \mathbf{z}, z_L, y)$. □

Proof of Corollary 1. Let $\mathbf{a} \in \mathcal{U}$.

Part (a). By Theorem 1(a), $\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y)$ is $L^\natural$-convex in $(\mathbf{v}, v_L, y) \in \mathcal{Q}$ for $\mathbf{a}$. By Lemma 4(a), $(v_L^*(\mathbf{a}, \mathbf{v}), y^*(\mathbf{a}, \mathbf{v})) = \min_{v_L, y}\{\bar{J}_t(\mathbf{a}, \mathbf{v}, v_L, y)\}$ is increasing in $\mathbf{v} \in \mathcal{V}$ for $\mathbf{a}$; hence:

$$\Delta_{v_i}v_L^*\left(\mathbf{a}, \mathbf{v}\right) = v_L^*(\mathbf{a}, \mathbf{v} + \mathbf{e}_i) - v_L^*(\mathbf{a}, \mathbf{v}) \geq 0$$

for all $i = -1, \dots, L-1$ with $(\mathbf{v} + \mathbf{e}_i) \in \mathcal{V}$. Using this, we can write

$$v_L^*(\mathbf{a}, \mathbf{v} + \mathbf{e}_i) \leq v_L^*(\mathbf{a}, \mathbf{v} + \mathbf{e}),$$

for all $i = -1, \dots, L-1$. By subtracting $v_L^*(\mathbf{a}, \mathbf{v})$ from both sides of the inequality, we get

$$v_L^*(\mathbf{a}, \mathbf{v} + \mathbf{e}_i) - v_L^*(\mathbf{a}, \mathbf{v}) \leq v_L^*(\mathbf{a}, \mathbf{v} + \mathbf{e}) - v_L^*(\mathbf{a}, \mathbf{v}).$$

By Lemma 4(b), the expression on the right is bounded by one. Therefore, we establish $\Delta_{v_i}v_L^*(\mathbf{a}, \mathbf{v}) = v_L^*(\mathbf{a}, \mathbf{v} + \mathbf{e}_i) - v_L^*(\mathbf{a}, \mathbf{v}) \leq 1$ for all $i = -1, \dots, L-1$.

Part (b). The proof follows the same steps as above.

Part (c). First, we prove the leftmost inequality. Recall that $v_i = x + \sum_{t=0}^{i} z_t$ for $i = -1, \dots, L$. Everything else remaining the same, increasing v_{L-1} by one means increasing z_{L-1} by 1. Therefore, we establish

$$\begin{aligned} \Delta_{v_{L-1}}v_L^*\left(\mathbf{a}, \mathbf{v}\right) &= v_L^*\left(\mathbf{a}, \mathbf{v} + \mathbf{e}_{L+1}\right) - v_L^*\left(\mathbf{a}, \mathbf{v}\right) \\ &= x + \sum_{t=0}^{L-1} z_t + 1 + z_L^*\left(\mathbf{a}, \mathbf{z} + \mathbf{e}_{L+1}\right) \\ &\quad - x - \sum_{t=0}^{L-1} z_t - z_L^*\left(\mathbf{a}, \mathbf{z}\right) \\ &= z_L^*\left(\mathbf{a}, \mathbf{z} + \mathbf{e}_{L+1}\right) + 1 - z_L^*\left(\mathbf{a}, \mathbf{z}\right) \\ &= \Delta_{z_{L-1}}z_L^*(\mathbf{a}, \mathbf{z}) + 1. \end{aligned}$$

From part 1(a), $\Delta_{v_{L-1}}v_L^*(\mathbf{a}, \mathbf{v}) \geq 0$. Therefore, we have $\Delta_{z_{L-1}}z_L^*(\mathbf{a}, \mathbf{z}) \geq -1$.

Second, we prove the rightmost inequality. Everything else remaining the same, increasing each argument of $\mathbf{v}$ by one means increasing x by one. Therefore, we write

$$\begin{aligned} &v_L^*\left(\mathbf{a}, \mathbf{v} + \mathbf{e}\right) - v_L^*\left(\mathbf{a}, \mathbf{v}\right) \\ &\quad = x + \sum_{t=0}^{L-1} z_t + z_L^*\left(\mathbf{a}, \mathbf{z} + \mathbf{e}_1\right) + 1 - x - \sum_{t=0}^{L-1} z_t - z_L^*\left(\mathbf{a}, \mathbf{z}\right) \\ &\quad = z_L^*\left(\mathbf{a}, \mathbf{z} + \mathbf{e}_1\right) + 1 - z_L^*\left(\mathbf{a}, \mathbf{z}\right) \\ &\quad = \Delta_x z_L^*(\mathbf{a}, \mathbf{z}) + 1. \end{aligned}$$

By Theorem 1(a) and Lemma 4(b), $v_L^*(\mathbf{a}, \mathbf{v} + \mathbf{e}) - v_L^*(\mathbf{a}, \mathbf{v}) \leq 1$. Therefore, we have $\Delta_x z_L^*(\mathbf{a}, \mathbf{z}) \leq 0$.

Next, we prove the inequalities that define the monotonic relationship between $\Delta_{z_i}z_L^*(\mathbf{a}, \mathbf{z})$ and $\Delta_{z_{i+1}}z_L^*(\mathbf{a}, \mathbf{z})$ for all $i = 0, \dots, L-2$. For ease of exposition, we introduce an additional notation: Let $\bar{\mathbf{e}}_i$ be a vector having zero for the first $i-1$ entries and one for the rest. From part 1(a), $v_L^*(\mathbf{a}, \mathbf{v})$ is increasing in $\mathbf{v} \in \mathcal{V}$. Therefore, we can write $v_L^*(\mathbf{a}, \mathbf{v} + \bar{\mathbf{e}}_{i+2}) - v_L^*(\mathbf{a}, \mathbf{v} + \bar{\mathbf{e}}_{i+3}) \geq 0$ for all $i = -1, \ldots, L-2$. Thus,

$$\begin{aligned} v_L^* (\mathbf{a}, \mathbf{v} + \bar{\mathbf{e}}_{i+2}) &- v_L^* (\mathbf{a}, \mathbf{v} + \bar{\mathbf{e}}_{i+3}) \\ &= x + \sum_{t=0}^{L-1} z_t + 1 + z_L^* (\mathbf{a}, \mathbf{z} + \mathbf{e}_{i+2}) \\ &\quad - x - \sum_{t=0}^{L-1} z_t - 1 - z_L^* (\mathbf{a}, \mathbf{z} + \mathbf{e}_{i+3}) \\ &= z_L^* (\mathbf{a}, \mathbf{z} + \mathbf{e}_{i+2}) - z_L^* (\mathbf{a}, \mathbf{z} + \mathbf{e}_{i+3}) \\ &= z_L^* (\mathbf{a}, \mathbf{z} + \mathbf{e}_{i+2}) - z_L^* (\mathbf{a}, \mathbf{z}) + z_L^* (\mathbf{a}, \mathbf{z}) - z_L^* (\mathbf{a}, \mathbf{z} + \mathbf{e}_{i+3}) \\ &= \Delta_{z_i} z_L^*(\mathbf{a}, \mathbf{z}) - \Delta_{z_{i+1}} z_L^*(\mathbf{a}, \mathbf{z}) \geq 0, \end{aligned} \tag{A5}$$

for all $i = 0, \ldots, L-2$. Similarly, for $i = -1$:

$$v_L^* (\mathbf{a}, \mathbf{v} + \bar{\mathbf{e}}_{i+2}) - v_L^* (\mathbf{a}, \mathbf{v} + \bar{\mathbf{e}}_{i+3}) = \Delta_x z_L^*(\mathbf{a}, \mathbf{z}) - \Delta_{z_0} z_L^*(\mathbf{a}, \mathbf{z}) \geq 0.$$

Part (d). The proof is similar to above. Again, we start with the proofs of the leftmost and the rightmost inequalities. Since increasing v_{L-1} by one means increasing z_{L-1} by one:

$$\begin{aligned} y^* (\mathbf{a}, \mathbf{v} + \mathbf{e}_{L+1}) &- y^* (\mathbf{a}, \mathbf{v}) \\ &= y^* (\mathbf{a}, \mathbf{z} + \mathbf{e}_{L+1}) - y^* (\mathbf{a}, \mathbf{z}) = \Delta_{z_{L-1}} y^*(\mathbf{a}, \mathbf{z}). \end{aligned}$$

From $\Delta_{v_{L-1}} y^*(\mathbf{a}, \mathbf{v}) \geq 0$, we also have $\Delta_{z_{L-1}} y^*(\mathbf{a}, \mathbf{z}) \geq 0$. Similarly, since increasing each argument of $\mathbf{v}$ by one means increasing x by one, $y^*(\mathbf{a}, \mathbf{v} + \mathbf{e}) - y^*(\mathbf{a}, \mathbf{v}) = y^*(\mathbf{a}, \mathbf{z} + \mathbf{e}_1) - y^*(\mathbf{a}, \mathbf{z}) = \Delta_x y^*(\mathbf{a}, \mathbf{z})$. By Theorem 1(a) and Lemma 4(b), $y^*(\mathbf{a}, \mathbf{v} + \mathbf{e}) - y^*(\mathbf{a}, \mathbf{v}) \leq 1$; hence, we also have $\Delta_x y^*(\mathbf{a}, \mathbf{z}) \leq 1$.

Next, we prove the inequalities that define the monotonic relationship between $\Delta_x y^*(\mathbf{a}, \mathbf{z})$ and $\Delta_{z_{i+1}} y^*(\mathbf{a}, \mathbf{z})$ for all $i = 0, \ldots, L-2$. From part 1(b), $y^*(\mathbf{a}, \mathbf{v})$ is increasing in $\mathbf{v}$. Therefore, we have $y^*(\mathbf{a}, \mathbf{v} + \bar{\mathbf{e}}_{i+2}) - y^*(\mathbf{a}, \mathbf{v} + \bar{\mathbf{e}}_{i+3}) \geq 0$ for all $i = -1, \ldots, L-2$. Thus,

$$\begin{aligned} y^* (\mathbf{a}, \mathbf{v} + \bar{\mathbf{e}}_{i+2}) &- y^* (\mathbf{a}, \mathbf{v} + \bar{\mathbf{e}}_{i+3}) \\ &= y^* (\mathbf{a}, \mathbf{z} + \mathbf{e}_{i+2}) - y^* (\mathbf{a}, \mathbf{z} + \mathbf{e}_{i+3}) \\ &= y^* (\mathbf{a}, \mathbf{z} + \mathbf{e}_{i+2}) - y^* (\mathbf{a}, \mathbf{z}) + y^* (\mathbf{a}, \mathbf{z}) - y^* (\mathbf{a}, \mathbf{z} + \mathbf{e}_{i+3}) \\ &= \Delta_{z_i} y^*(\mathbf{a}, \mathbf{z}) - \Delta_{z_{i+1}} y^*(\mathbf{a}, \mathbf{z}) \geq 0, \end{aligned}$$

for all $l = 0, \ldots, L-2$. Similarly, we have $\Delta_x y^*(\mathbf{a}, \mathbf{z}) \geq \Delta_{z_0} y^*(\mathbf{a}, \mathbf{z})$.

□

A.2. Detailed results of the numerical experiment in Section 4.2

Table A.1. Effect of parameters on the value of imperfect ADI (case 1, $L \le \tau_l$).

					With return									Without return		
		All Cases			Low ($c_r = 2.5h$)			Med. ($c_r = 25h$)			High ($c_r = 125h$)			$c_r = \infty$		
Parameters	Values	Avg. (%)	Min. (%)	Max. (%)	Avg. (%)	Min. (%)	Max. (%)	Avg. (%)	Min. (%)	Max. (%)	Avg. (%)	Min. (%)	Max. (%)	Avg. (%)	Min. (%)	Max. (%)
All instances		30.06	0.03	87.64	41.48	1.09	87.64	33.36	0.11	85.01	25.26	0.03	82.06	20.12	0.03	79.33
L	1	29.47	0.03	87.64	40.78	0.72	87.64	32.76	0.15	85.03	24.74	0.03	82.09	19.60	0.03	79.36
	2	30.64	0.02	87.64	42.17	1.46	87.64	33.97	0.08	84.99	25.79	0.02	82.03	20.64	0.02	79.29
$[\tau_l, \tau_u]$	[2, 2]	30.95	0.01	88.77	43.37	1.13	88.77	34.20	0.14	86.15	25.72	0.01	83.04	20.52	0.01	80.31
	[2, 6]	29.16	0.04	86.50	39.59	1.06	86.50	32.53	0.08	83.87	24.81	0.04	81.08	19.72	0.04	78.35
λ	0.001	34.22	0.00	89.87	49.12	0.12	89.87	42.02	0.00	89.64	30.25	0.00	88.64	15.50	0.00	82.22
	0.005	28.29	0.06	88.68	40.20	0.76	88.68	32.32	0.12	86.45	20.67	0.06	80.79	19.99	0.06	79.01
	0.025	27.65	0.01	84.36	35.12	2.39	84.36	25.76	0.21	78.94	24.87	0.01	76.75	24.87	0.01	76.75
h	5	22.93	0.01	89.47	29.18	0.12	89.47	34.10	0.25	88.57	17.26	0.02	88.64	11.19	0.01	82.22
	50	29.68	0.00	89.87	41.39	10.39	89.87	35.12	0.12	87.47	23.95	0.04	49.24	18.26	0.00	79.36
	500	32.58	0.00	89.47	53.86	10.39	89.47	36.66	0.00	88.57	23.00	0.00	84.57	16.80	0.00	79.01
c_e	5000	31.15	0.00	89.87	48.61	2.11	89.87	34.10	0.00	89.64	23.33	0.00	88.64	18.56	0.00	82.22
	25 000	30.67	0.01	89.74	41.95	0.38	89.74	35.12	0.25	89.29	25.73	0.02	87.29	19.89	0.01	76.75
	125 000	28.35	0.03	89.47	33.88	0.12	89.47	30.87	0.12	88.57	26.74	0.11	84.57	21.91	0.03	79.01
p	0.5	24.01	0.00	89.49	38.78	0.12	89.49	27.15	0.00	87.47	17.26	0.00	78.51	12.84	0.00	47.57
	0.7	29.22	0.02	89.74	41.89	0.15	89.74	32.34	0.04	88.87	23.95	0.02	85.02	18.72	0.02	65.53
	0.9	36.94	0.04	89.87	43.77	0.19	89.87	40.61	0.13	89.64	34.59	0.04	88.64	28.80	0.04	82.22
q	0.5	17.75	0.00	49.92	25.45	0.12	49.92	19.93	0.00	49.79	14.45	0.00	49.24	11.19	0.00	47.48
	0.7	27.83	0.01	69.90	38.99	0.52	69.90	30.98	0.02	69.73	23.11	0.01	68.95	18.26	0.01	65.20
	0.9	44.58	0.00	89.87	60.00	4.08	89.87	49.19	0.02	89.64	38.22	0.00	88.64	30.91	0.00	82.22

Table A.2. Effect of parameters on the value of imperfect ADI (case 2, $L > \tau_l$).

		Cases (L, τ_l, τ_u)														
		All cases			*(1,0, 0)*			*(1,0, 4)*			*(2,0, 0)*			*(2,0, 4)*		
Parameters	*Values*	*Avg. (%)*	*Min. (%)*	*Max. (%)*	*Avg. (%)*	*Min. (%)*	*Max. (%)*	*Avg. (%)*	*Min. (%)*	*Max. (%)*	*Avg. (%)*	*Min. (%)*	*Max. (%)*	*Avg. (%)*	*Min. (%)*	*Max. (%)*
All instances		6.03	0.00	62.19	2.14	0.00	23.13	13.09	0.00	62.19	1.04	0.00	11.82	7.87	0.01	42.90
c_r	2.5h	9.76	0.00	67.84	3.27	0.00	28.93	20.85	0.00	67.84	1.78	0.00	17.32	13.13	0.00	48.07
	25h	6.40	0.00	67.12	2.01	0.00	28.35	14.37	0.00	67.12	0.90	0.00	15.79	8.31	0.00	46.91
	125h	4.35	0.00	63.93	1.65	0.00	26.03	9.58	0.00	63.93	0.74	0.00	15.78	5.43	0.00	41.75
	∞	3.63	0.00	52.04	1.62	0.00	26.03	7.56	0.00	52.04	0.73	0.00	15.78	4.61	0.00	35.58
λ	0.001	5.65	0.00	67.84	0.09	0.00	2.07	14.74	0.00	67.84	0.08	0.00	2.02	7.68	0.00	48.07
	0.005	5.63	0.01	66.13	2.89	0.01	28.93	12.18	0.02	66.13	0.64	0.01	8.19	6.82	0.02	46.52
	0.025	6.82	0.00	56.08	3.43	0.00	27.37	12.35	0.00	56.08	2.39	0.00	17.32	9.10	0.00	38.54
h	5	8.01	0.00	67.84	5.17	0.00	28.93	14.06	0.03	67.84	1.98	0.00	17.32	10.82	0.04	48.07
	50	5.29	0.00	67.64	1.00	0.00	17.11	12.46	0.00	67.64	0.90	0.00	13.82	6.82	0.00	47.90
	500	4.80	0.00	67.22	0.25	0.00	9.04	12.75	0.00	67.22	0.24	0.00	7.89	5.96	0.00	47.49
c_e	5000	5.76	0.00	67.84	1.87	0.00	27.37	13.56	0.00	67.84	0.18	0.00	2.96	7.44	0.00	48.07
	25 000	5.50	0.00	67.64	1.12	0.00	17.11	12.30	0.00	67.64	1.19	0.00	13.82	7.39	0.00	47.90
	125 000	6.84	0.00	67.22	3.42	0.00	28.93	13.41	0.00	67.22	1.75	0.00	17.32	8.78	0.00	47.49
p	0.5	3.14	0.00	43.07	1.57	0.00	28.55	7.02	0.00	43.07	0.68	0.00	16.92	3.29	0.00	24.39
	0.7	5.73	0.00	57.24	2.01	0.00	28.79	12.67	0.00	57.24	1.01	0.00	17.18	7.24	0.00	34.58
	0.9	9.23	0.00	67.84	2.84	0.00	28.93	19.58	0.00	67.84	1.43	0.00	17.32	13.08	0.00	48.07
q	0.5	4.20	0.00	37.68	1.69	0.00	20.01	8.86	0.00	37.68	0.82	0.00	10.55	5.44	0.00	26.68
	0.7	5.90	0.00	52.77	2.05	0.00	23.80	12.78	0.00	52.77	0.97	0.00	12.95	7.80	0.00	37.38
	0.9	8.00	0.00	67.84	2.67	0.00	28.93	17.63	0.00	67.84	1.33	0.00	17.32	10.37	0.00	48.07