European Journal of Operational Research 250 (2016) 143–154

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing and Logistics

Control and enforcement in order to increase supplier inventory in a JIT contract

Matan Shnaiderman*, Liron Ben-Baruch

Department of Management, Bar-Ilan University, Ramat-Gan 52900, Israel

ARTICLE INFO

Article history:
Received 4 November 2014
Accepted 24 October 2015
Available online 30 October 2015

Keywords:
Just-In-Time
Customer-led supply chain
Replenishment on-time enforcement

ABSTRACT

Prompt response to customer demand has long been a point of major concern in supply chains. “Inventory wars” between suppliers and their customers are common, owing to cases in which one supply chain party attempts to decrease its stock at the expense of the other party. In order to ensure that suppliers meet their commitments to fulfill orders on time, customers must formulate incentives or, alternatively, enforce penalties. This paper deals with a customer organization that has a contract with a supplier, based on Just-In-Time strategy. Initiating a policy of sanctions, the customer becomes the lead player in a Stackelberg game and forces the supplier to hold inventory, which is made available to the customer in real-time. Using a class of sanctioning functions, we show that the customer can force the supplier to hold inventory up to some maximal value, rendering actual enforcement of sanctions unnecessary. However, contrary to expectations, escalation of the enforcement level can in fact reduce the capacity of the supplier to replenish on time. Consequently, the customer must sanction meticulously in order to receive his inventory on time. Having the possibility to devote a few hours each day to sanctioning activity significantly reduces the customer's expected cost. In particular, numerical examples show that the customer's costs under an enforcement level may be only 2 percent higher than his costs in a situation in which all inventory is necessarily replenished on time.

1. Introduction

Prompt response to customer demand has long been a point of major concern in supply chains and has given rise to such inventory management strategies as *Continuous Replenishment Program, Efficient Consumer Response, Just-In-Time (JIT) Supply, Ship-to-Order* and *Demand-Driven Supply* (see, for example, Barnes et al., 2000; Harris, Swatman, & Kurnia, 1999; Raghunathan & Yeh, 2001; Ayers, 2001). In the electronics industry, for example, original equipment manufacturers (OEMs) frequently contract out their manufacturing to electronics manufacturing services, and the latter are contractually obligated to meet the OEMs' demands on a continuous basis in a JIT mode, with little or no advance notice (Barnes et al., 2000). Intra-supply-chain competition constitutes a main barrier to the implementation of such inventory management approaches. Indeed, “inventory wars” between suppliers (e.g., manufacturers) and their customers (e.g., retailers) are a common occurrence, owing to cases in which one supply chain party attempts to decrease its stock at the expense of another party (Cachon, 2001). As a result of such competition, the likelihood of stockouts grows, and the replenishment lead-time becomes uncertain.

Management literature suggests various coordination approaches to overcome intra-supply-chain competition. These approaches are based on specially designed incentives, penalties and cost sharing. Grout and Christy (1999), for example, examine how a supplier, committed to a long-term contract with a customer based on a fixed selling price, responds to incentives for supplying JIT shipments on time, as well as to penalties for failure to fulfill demand on time. The authors show how increasing the incentive or penalties increases the inventory capacities that the supplier holds. If there is no incentive or, alternatively, the penalties are not enforced, the supplier is led to reduce his inventory capacities as well as his service level.

Vendor-managed inventories (VMI) are another successful approach to preventing the uncertain lead-times and low service levels associated with intra-supply-chain competition. With VMI, suppliers generate orders based on mutually agreed-upon objectives for inventory levels, fill rates and transaction costs, in addition to demand information sent by their distributor customers. The supplier shares the customer's inventory-related costs and monitors the inventory status information to make sure that the distributor customer always has the appropriate amount of stock on hand (see, for example, De Toni & Zamolo, 2005; Lee, So, & Tang, 2000; Vigtil, 2007; Yonghui & Raiesh, 2004 for deliberations on information sharing between parties employing VMI). It has been shown that, in promoting information sharing between the customer and the supplier, vendor-managed systems enable the customer (distributor) to lower his inventory

* Corresponding author. Tel.: +972 505528951.
E-mail address: shnidem@biu.ac.il (M. Shnaiderman).

http://dx.doi.org/10.1016/j.ejor.2015.10.047
0377-2217/

levels, thereby leading to carrying-cost savings (Cachon & Fisher, 1997; Schenk & McInerny, 1998). Moreover, vendor-managed systems provide the supplier with flexibility in its production operations (Fry, Kapuscinski, & Olsen, 2001; Savaşaneril & Erkip, 2010). Given that the firms maintain their independence, however, they can exercise discretion over the extent of information sharing, which may have complex consequences. Notably, Lee et al. (2000) show that the supplier's on-hand inventory level may be affected by the level of information shared with the customer. In particular, the authors find that if the customer does not inform the supplier of the realized value of the demand in each period, the supplier ends up holding an inventory level that is almost two times greater than that under information sharing.

Another approach is to incentivize the supplier to increase its inventory level, thereby reducing the lead-times and the likelihood of stockouts on the customer's side. This approach is realized through so-called option contracts, in which the customer pays for the option to obtain additional supplies when needed. Specifically, in Wang and Liu (2007) the customer orders a basic level of inventory, and the supplier necessarily holds that level. In addition, the customer pays an option cost for every additional unit of inventory that the supplier chooses to hold for him. If, according to the realized demand, the customer needs to exercise an additional purchase of inventory, he pays an exercise cost. Consequently, it is in the supplier's interest to hold a greater level of inventory than that of the customer's basic order (see also Zhao, Wang, Cheng, Yang, & Huang, 2010). Fang and Whinston (2007) consider an option contract in which the supplier is dominant and sets the option and exercise costs. If a customer buys options in advance (before the demand is realized), he receives priority over other customers in receiving the inventory. The authors show that the inventory level that the supplier holds in this case is higher than that under no option contract.

In this paper we examine the case of a customer who employs the JIT management strategy when contracting with a supplier. The JIT management strategy implies that lead-times are short. The customer's goal is to set an optimal enforcement policy in order to prevent breach of contract by the supplier, and to minimize the associated costs.

Our study is motivated by a real-life supply chain involving the Israel Police (customer) and a supplier of security products. The arrangement between the customer and the supplier is based on a standard contract, according to which, at the beginning of each period, the customer sets an order quantity to be supplied and pays for the order. The supplier then ships the products over the course of the period in response to the customer's ongoing requests. When demand during a given period is lower than expected, the customer might not request that the entire order quantity be shipped during that period; in such a case, the supplier will still fulfill the entire prepaid order by the end of the period. Similarly, if, at any point in time over the course of the current period, the supplier does not have sufficient stock to fulfill a specific order, the unshipped quantity will be supplied by the end of the period.

A key point of concern in the scenario described is that inventory shortage can lead to deadly consequences. As a result, both the police (customer) and the supplier accumulate excessive inventories, thereby consuming vast resources. The large order quantities dealt with imply that high inventory costs are involved, even if the unit holding cost is not high. In order to reduce his own holding costs, the supplier attempts to reduce the quantity of stock he holds at any given time. This implies that the supplier ships steadily in response to the customer's requests (i.e., his shipment costs are not affected) but is not always able to completely fill the orders on time. As noted above, in cases in which the supplier lacks sufficient stock to fulfill an order, he ships the remaining quantity at the end of the period. This behavior induces the customer to hold greater stocks. To prevent the supplier from engaging in such behavior, the customer employs sanctions against the supplier when the latter does not provide timely shipments. The supplier is then charged by the purchasing department, which takes all the complaints into account and issues them in the form of a monetary charge to the supplier. Note that the customer imposes sanctions only on days in which it requests products (which have been paid for as part of the prepaid order made at the beginning of the current period) and the supplier is not able to ship the products on those same days. Enforcement of sanctions inflicts costs on the customer: the act of sanctioning is time-consuming and is carried out in addition to the standard logistic functions executed by customer's management department. Thus, the time invested in sanctioning, referred to as the *enforcement level*, is a decision variable. That is, the customer's effort level is measured with the time spent on sanction-related activities, which is a common practice. The customer's goal is to find the optimal trade-off between the total investment in sanctioning and the inventory holding cost.

We model the competition in the described two-echelon supply chain with a Stackelberg game such that the customer is the leader, whereas the supplier is the follower. Our results show that the customer can force the supplier to hold inventory up to some maximal value. This value depends on the total time that the customer can actually spend enforcing penalties on each day, on the rigorousness of the punishment toward the supplier due to not replenishing on time (see below), on the holding cost and on demand distribution. Moreover, we find that when the customer escalates the enforcement level, the supplier does not necessarily increase the inventory level that he holds throughout the period (and can thereby replenish on time) and may even reduce it.

The rest of this paper is organized as follows. Section 2 presents a description of the problem and the corresponding model. Section 3 describes the decision of the supplier (follower player) regarding the inventory level he holds for the customer. This decision is made in response to the decisions of the customer (leader player) about the order quantity and enforcement level; the customer's model is described in Section 4. Both Sections 3 and 4 include analytical models as well as numerical illustrations. Section 5 concludes the paper.

2. Problem formulation

We consider a multi-period two-echelon supply chain model consisting of one supplier and one customer. During each period, n, of τ days, the customer's inventory has to satisfy a total periodic demand of D_n. This demand is stochastic with realization d_n, probability density function $f_n(d_n)$ and cumulative distribution function $F_n(\cdot)$. As in many studies, including Khmelnitsky and Caramanis (1998), Kogan and Lou (2002) and Kogan and Tell (2009), the demand rate D_n/τ within a given period n is assumed to be constant. Let N be the total number of periods. At the beginning of each period n, before the demand is realized, the customer orders q_n units of inventory based on his initial inventory I_n. According to the contract, the supplier has to replenish that quantity of inventory over the course of period n, in shipments of quantities that correspond to the customer's needs. The supplier is a distributor or a wholesaler that can deliver in no time if his stocks are sufficient. Moreover, in the type of environment that we consider (i.e., a small country), distances are small, so the delivery process is quick and efficient. Therefore, we assume the supplier's lead time is negligible. On the other hand, it takes time to his subcontractors to manufacture and deliver the products. Therefore, the subcontractors' lead-time is not zero, that is, the supplier cannot wait until the last moment for the demand to realize as he might not be able to meet the demand and thus incur penalties. However, in an attempt to avoid holding costs, the supplier may choose not to hold the entire quantity q_n in his warehouse during the period and instead to hold only some fraction α_n of q_n. We refer to $\alpha_n q_n$ as the "held quantity" at period n. Thus, the quantity of inventory that is available for the customer during period n is $I_n + \alpha_n q_n$. The customer will necessarily receive the rest of the order, $(1-\alpha_n)q_n$, at the end of that period. On the basis of the value of α_n (as well as the values of q_n and D_n; see below), the customer chooses how many hours he will spend each day engaging in enforcement activity; this number of hours is denoted by x_n and is referred to as the enforcement level. Enforcement activity can include, for example, submitting reports and complaints to the customer's purchasing department regarding the supplier. The value of x_n, which is a function of α_n, must be lower than or equal to a maximal capacity $x^{\max}$. The customer pays cost c for every hour he spends engaging in sanctioning activity. At the end of period n, the supplier is charged by the purchasing department. The cost, or *penalty*, that the supplier pays as a result of the customer's enforced sanctions, is denoted $p(x_n)$; this function is increasing, convex and satisfies $p(0) = 0$. Once the customer orders a product quantity (which naturally accounts for his inventory on hand, I_n), the customer's intention is to penalize the supplier not fulfilling the order on time, even if the customer will have no shortage in his stock by the time the shipment is required. Therefore, the supplier's inventory policy does not depend on the stock that the customer holds.

Previous research shows that, in games such as that described above, the customer can gain a significant advantage by being the leader rather than the follower (Dukes, Geylani, & Srinivasan, 2009; Matsui, 2010; Xue, Demirag, & Niu, 2014). In order to fortify his leadership in the game, at the beginning of period n and after determining the order quantity q_n, the customer defines the enforcement level function $x_n(\alpha_n)$ and shares it with the supplier. (Note that, since this function's structure is affected by q_n as well as by the demand D_n, its complete notation is $x_n^{q_n,D_n}(\alpha_n)$; however, we only use this symbol when it is essential.) The supplier responds by setting α_n. Depending on the demand realization d_n, the enforcement level function is a random variable (with realization $x_n^{q_n,d_n}(\alpha_n)$) for every value of α_n. This function is decreasing in the interval [0,1], and satisfies $x_n(0) \le x^{\max}$ as well as

$$x_n^{q_n,D_n}(\alpha_n) = 0 \text{ for } \alpha_n \ge \min\left\{\frac{D_n}{q_n}, 1\right\}. \tag{1}$$

The reasoning for (1) is as follows. If $\alpha_n = 1$, then no sanctions are required. Moreover, if the held quantity $\alpha_n q_n$ is greater than the realized demand d_n, then during period n the supplier holds inventory that meets the whole periodic demand. Thus, the customer does not impose any sanctions. Note that as the demand grows, the value of x_n becomes higher for every value of α_n. This situation is fair, since the supplier knows that higher demand is associated with greater shortage—and thereby greater damage to the customer—in the event that the supplier has failed to hold the quantity of inventory that the customer requires. The customer's and supplier's unit holding costs during the whole period of duration τ are denoted by h_c and h_s, respectively. The customer pays cost w for every unit of inventory he receives, and the supplier pays cost b for every unit that he orders. We assume that all of the unit costs and demand distributions are known to both of the players. As stated above, the customer sets the optimal order quantity and the corresponding enforcement level before the periodic demand is realized, based on its distribution. On the basis of Stackelberg equilibrium, the customer is able to calculate the supplier's decision regarding the level of held quantity (which is also set at the beginning of the period, before realization of the periodic demand).

Given an order quantity q_n, there are numerous different types of functions that can be used to set the optimal enforcement level x_n. In this paper we consider the following type of function:

$$x_n^{q_n,D_n,K}(\alpha) = \begin{cases} x^{\max}\left(1 - \left[\max\left(\frac{q_n}{D_n}, 1\right)\cdot\alpha\right]^K\right), \text{ if } 0 \\ \quad \le \alpha < \min\left(\frac{D_n}{q_n}, 1\right) \\ 0, \text{ if } \min\left(\frac{D_n}{q_n}, 1\right) \le \alpha \le 1 \end{cases} \tag{2}$$

for $K > 0$, where K is a parameter representing the *rigorousness level* of the enforcement policy, i.e., the extent to which sanctions are actually enforced.

Remark 1. If $D_n \ge q_n$, then (2) becomes $x_n^{q_n,D_n,K}(\alpha) = x^{\max}(1 - \alpha^K)$ for every $0 \le \alpha \le 1$.

As we show in Section 4.3.1 below, under policy (2) the customer does not actually have to implement the sanctions. He only needs to have the possibility to do so for the sake of deterrence. Since higher values of $x^{\max}$ imply more effective deterrence, it is preferable for the customer to set the whole $x^{\max}$ as a coefficient in the first term of (2), and increase the held quantity. Even though this choice may increase his sanctioning cost, this cost remains theoretical and does not actually occur.

For every realized demand d_n, the corresponding function $x_n^{q_n,d_n,K}(\alpha_n)$ is continuous in [0,1]; twice differentiable in $[0, \min(\frac{d_n}{q_n}, 1))$; and vanishes in $[\min(\frac{d_n}{q_n}, 1), 1]$. For every fixed value of α_n, the value $x_n^{q_n,d_n,K}(\alpha_n)$ is non-decreasing in K. If $K > 1$, then $x_n^{q_n,d_n,K}(\alpha_n)$ is concave in $[0, \min(\frac{d_n}{q_n}, 1))$, and if $0 < K < 1$, then it is convex in that interval. Moreover, as K increases and tends to infinity, $x_n^{q_n,d_n,K}(\alpha_n)$ becomes more rigorous and converges to the limit function $x_n^{q_n,d_n}(\alpha_n) = x^{\max}$ for $0 \le \alpha_n < \min(d_n/q_n, 1)$. On the other hand, as K decreases and tends to zero, the enforcement level tends to the limit function $x_n^{q_n,d_n}(\alpha) = 0$ for $0 < \alpha \le \min(d_n/q_n, 1)$. As we show below, the quantity of inventory that the supplier holds for the customer, $\alpha_n q_n$, does not necessarily increase in K.

Using three-stage backward dynamic programming, we start by calculating the supplier's response to the customer's decisions regarding q_n and x_n.

3. The supplier's model

3.1. Model setup

The supplier's (stochastic) cost at period n, denoted by $C_n^s(q_n, x_n, \alpha_n, D_n)$, is calculated as follows.

If $\alpha_n q_n \le D_n$, then the supplier's cost at period n is

$$C_n^s(q_n, x_n, \alpha_n, D_n) = p\left(\left(1 - \frac{\alpha_n q_n}{D_n}\right)\tau x_n^{q_n,D_n}(\alpha_n)\right) + (b - w)q_n + \frac{\alpha_n^2 q_n^2}{2D_n}h_s. \tag{3}$$

The first term on the right-hand side of (3) is equal to the penalty inflicted on the supplier, i.e., the cost the supplier pays as a result of the customer's enforced sanctions. Assuming that the demand rate is fixed during the period, the supplier provides new inventory during a fraction of $\frac{\alpha_n q_n}{D_n}$ out of the total period duration, τ days. During the remaining fraction of the period $(1 - \frac{\alpha_n q_n}{D_n})$, the customer spends on average $x_n^{q_n,D_n}(\alpha_n)$ hours on sanctioning each day. Therefore, the total number of hours spent on sanctions in period n is equal to $(1 - \frac{\alpha_n q_n}{D_n})\tau x_n^{q_n,D_n}(\alpha_n)$. Note that the supplier may hold zero inventory and supply at the beginning of the next time period rather than during the current period, but then he will incur excessive penalties. The second and third terms on the right-hand side of (3) are the supplier's order and holding costs, respectively.

On the other hand, if $\alpha_n q_n \ge D_n$, then the supplier's periodic cost is

$$C_n^s(q_n, x_n, \alpha_n, D_n) = (b - w)q_n + \left(\alpha_n q_n - \frac{D_n}{2}\right)h_s, \tag{4}$$

containing only the order and holding costs.

The supplier's goal at period n is to minimize the total expected cost-to-go. Let

$$B_n^s(q_n, x_n) = \min_{(\alpha_n,\ldots,\alpha_N)} \mathrm{E}\left(\sum_{m=n}^{N} C_m^s(q_m, x_m, \alpha_m, D_m)\right), \tag{5}$$

and let $(\mathrm{E}C_n^s)(q_n, x_n, \alpha_n)$ denote the supplier's expected value of (4).

According to the Bellman equation (Bellman, 1957; Kogan & Shnaiderman, 2010), the minimal cost (5) recursively satisfies the following:

$$\begin{cases} B_N^s(q_N, x_N) = \min\limits_{0\le\alpha_N\le 1} (EC_N^s)(q_N, x_N, \alpha_N) \\ B_n^s(q_n, x_n) = \min\limits_{0\le\alpha_n\le 1} \left((EC_n^s)(q_n, x_n, \alpha_n) + E\left[B_{n+1}^s(q_{n+1}, x_{n+1})\right]\right), \\ \quad n < N. \end{cases}$$

Consequently, we recursively define the supplier's objective function

$$\begin{cases} J_N^s(q_N, x_N, \alpha_N) = (EC_N^s)(q_N, x_N, \alpha_N) \\ J_n^s(q_n, x_n, \alpha_n) = (EC_n^s)(q_n, x_n, \alpha_n) + E\left[B_{n+1}^s(q_{n+1}, x_{n+1})\right], \quad n < N, \end{cases} \tag{6}$$

and obtain the following:

$$B_n^s(q_n, x_n) = \min_{0\le\alpha_n\le 1} J_n^s(q_n, x_n, \alpha_n).$$

The customer's subsequent decisions q_{n+1} and x_{n+1} depend on his inventory level I_{n+1}, which fulfills the following conditions.

Remark 2. The initial inventory at period $n+1$ satisfies $I_{n+1}=I_n+q_n-D_n$, and depends neither on the fraction α_n nor on the enforcement level x_n.

Taken together, the second term of (6) and Remark 2 lead to the following proposition, which substantially simplifies the supplier's calculations.

Proposition 1. The only component of $J_n{}^s$ that depends on the decision variable α_n is $EC_n{}^s$, that is, the expected cost of the current period. Thus, the supplier can myopically set the value of α_n to be that one that minimizes $EC_n{}^s$, namely, the supplier deals with the following periodic problem:

$$\min_{0\le\alpha_n\le 1} \quad (EC_n^s)(q_n, x_n, \alpha_n). \tag{7}$$

By (3) and (4), the supplier's expected cost at period n, as a function of α_n (given q_n and x_n), is

$$(EC_n^s)(q_n, x_n, \alpha_n) = (b-w)q_n + \int_0^{\alpha_n q_n} \left(\alpha_n q_n - \frac{d_n}{2}\right) h_s f_n(d_n) dd_n + \int_{\alpha_n q_n}^{\infty} \left[p\left(\left(1-\frac{\alpha_n q_n}{d_n}\right)\tau x_n^{q_n,d_n}(\alpha_n)\right) + \frac{\alpha_n^2 q_n^2}{2d_n} h_s\right] f_n(d_n) dd_n. \tag{8}$$

In order to solve (7), we differentiate (8) and obtain

$$\frac{\partial (EC_n^s)}{\partial \alpha_n} = \int_0^{\alpha_n q_n} q_n h_s f_n(d_n) dd_n + \int_{\alpha_n q_n}^{\infty} \left[p'\left(\left(1-\frac{\alpha_n q_n}{d_n}\right)\tau x_n^{q_n,d_n}(\alpha_n)\right)\left(-\frac{q_n}{d_n}\tau x_n^{q_n,d_n}(\alpha_n) + \left(1-\frac{\alpha_n q_n}{d_n}\right)\tau\cdot(x_n^{q_n,d_n})'(\alpha_n)\right) + \frac{\alpha_n q_n^2}{d_n} h_s\right] f_n(d_n) dd_n \tag{9}$$

as well as

$$\frac{\partial^2 (EC_n^s)}{\partial \alpha_n^2} = \int_{\alpha_n q_n}^{\infty} \left[\begin{array}{l} p''\left(\left(1-\frac{\alpha_n q_n}{d_n}\right)\tau x_n^{q_n,d_n}(\alpha_n)\right)\left(-\frac{q_n}{d_n}\tau x_n^{q_n,d_n}(\alpha_n) + (1-\frac{\alpha_n q_n}{d_n})\tau\cdot(x_n^{q_n,d_n})'(\alpha_n)\right)^2 \\ +p'\left(\left(1-\frac{\alpha_n q_n}{d_n}\right)\tau x_n^{q_n,d_n}(\alpha_n)\right)\left((1-\frac{\alpha_n q_n}{d_n})\tau\cdot(x_n^{q_n,d_n})''(\alpha_n) - \frac{2q_n}{d_n}\tau\cdot(x_n^{q_n,d_n})'(\alpha_n)\right) + \frac{q_n^2}{d_n} h_s \end{array}\right] f_n(d_n) dd_n + p'(0)\frac{q_n}{\alpha_n}\tau x_n^{q_n,\alpha_n q_n}(\alpha_n) f_n(\alpha_n q_n). \tag{10}$$

Note that, though the conditions for the supplier's objective function to be convex are very awkward and therefore not useful, the second derivative (10) of the objective function contains five terms, only one of which, $(1-\frac{\alpha_n q_n}{d_n})\tau x_n{}''(\alpha_n)$, may be negative, and this is the situation if $x_n(\alpha_n)$ is strictly concave. Consequently, the second derivative is likely to remain positive even if the enforcement level is rigorous. Our numerical experiments also show that the objective function is either convex and monotonically decreasing or quasi-convex (with a unique optimal value $0 < \alpha_n{}^* < 1$). Therefore, we next assume that the supplier's objective function (8) is quasi-convex in α_n. Also, substituting $\alpha_n = 0$ in (9) leads to

$$\int_0^{\infty} \left[p'\left(\tau x_n^{q_n,d_n}(0)\right)\left(-\frac{q_n}{d_n}\tau x_n^{q_n,d_n}(0) + \tau\cdot(x_n^{q_n,d_n})'(0)\right)\right] f_n(d_n) dd_n < 0, \tag{11}$$

and substituting $\alpha_n = 1$ in (9) leads to

$$\int_0^{q_n} q_n h_s f_n(d_n) dd_n + \int_{q_n}^{\infty} \left[p'(0)\left((1-\frac{q_n}{d_n})\tau\cdot(x_n^{q_n,d_n})'(1)\right) + \frac{q_n^2}{d_n} h_s\right] f_n(d_n) dd_n. \tag{12}$$

We obtain the following proposition.

Proposition 2. Let the supplier's objective function (8) be quasi-convex. According to (11), if $x_n(0) > 0$ then the supplier will always choose $\alpha_n{}^* > 0$. If the value of (12) is greater than or equal to 0, then there exists a unique optimal value $0 < \alpha_n{}^* < 1$ that solves the following equation:

$$\int_0^{\alpha_n^* q_n} q_n h_s f_n(d_n) dd_n + \int_{\alpha_n^* q_n}^{\infty} \times \left[\begin{array}{l} p'\left(\left(1-\frac{\alpha_n^* q_n}{d_n}\right)\tau x_n^{q_n,d_n}(\alpha_n^*)\right)\times \\ \left(-\frac{q_n}{d_n}\tau x_n^{q_n,d_n}(\alpha_n^*) + (1-\frac{\alpha_n^* q_n}{d_n})\tau\cdot(x_n^{q_n,d_n})'(\alpha_n^*)\right) + \frac{\alpha_n^* q_n^2}{d_n} h_s \end{array}\right] \times f_n(d_n) dd_n = 0. \tag{13}$$

Otherwise, $\alpha_n{}^*=1$. □

Proposition 2 determines that the supplier will always hold some positive level of inventory for the customer (while the order quantity is positive and the enforcement function does not identically vanish) during the period. Furthermore, if the supplier's holding cost is high while the penalty cost p is not rigorous (e.g., in the case of $p'(0) = 0$), then he will hold only a partial quantity of the order. Otherwise, he will prefer to hold the total quantity. However, below we show numerically that even in the former case, if the order quantity q_n is lower than or equal to an upper bound Q_n, then the customer can set an enforcement level that leads to an optimal value of α_n, that is approximately equal to 1. Once q_n exceeds Q_n, then $\alpha_n{}^*$ necessarily becomes lower than 1. In other words, when the customer's order quantity does not exceed the maximal value Q_n, the supplier is forced to hold that entire quantity during the current period. On the other hand, the supplier will never hold any quantity that is greater than Q_n (even if q_n is much higher than that quantity).

In what follows we numerically examine these theoretical results.

3.2. Numerical examples and sensitivity analysis

In our numerical example, we assume that the duration of each period is 90 days (i.e. $\tau = 90$), that is, one quarter. Let also $x^{\max} = 3$,

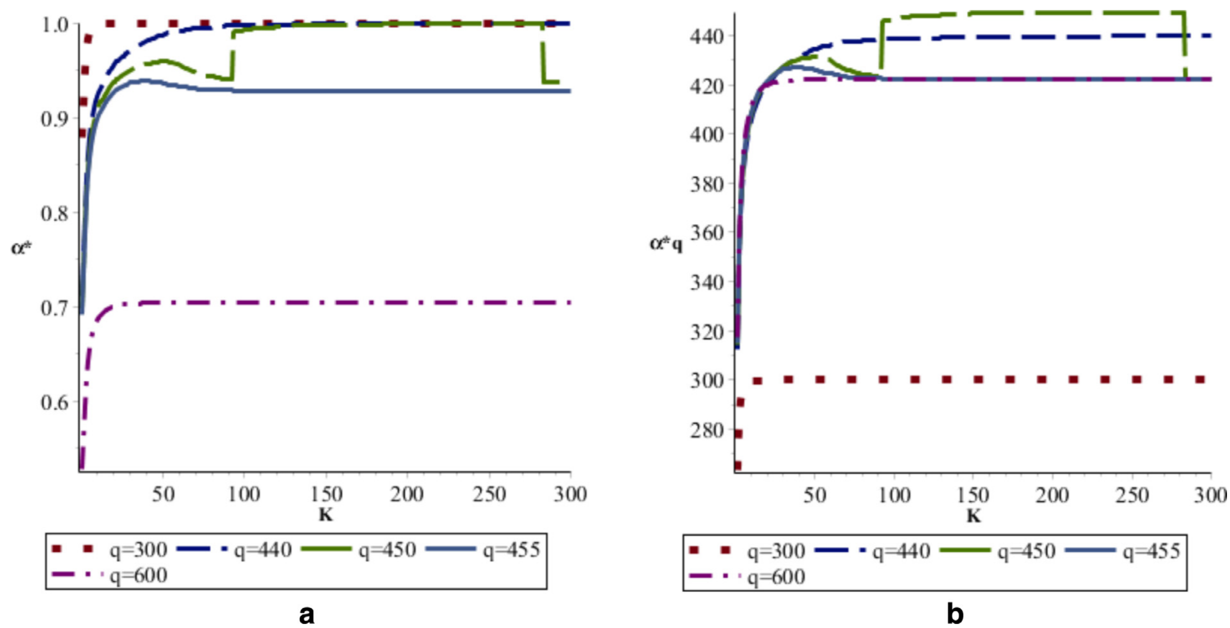

Fig. 1. Optimal fractions of order quantities (a) and actual held quantities (b) for different values of K and q_n.

and assume that the periodic demand is uniformly distributed in the interval [0, 500]. Let the penalty cost inflicted by the purchasing department be of the following quadratic type

$$p(x_n) \ = \ a \cdot {x_n}^2 \tag{14}$$

(with $a > 0$), which becomes more rigorous as a increases. Given (2) and (14), we find that the supplier's objective function (8) is either strictly convex or decreasing in [0, 1], namely, it is quasi-convex in this interval. We first consider how the parameter K affects the held quantity for fixed orders q_n. As expected, $\lim\limits_{K\to 0} \alpha_n^*(q_n, x_n^{q_n,d_n,K})q_n = 0$, and in addition, there exists a limit $\ell_n(q_n) = \lim\limits_{K\to\infty} \alpha_n^*(q_n, x_n^{q_n,d_n,K})q_n$. The value of ℓ_n is equal to q_n, if q_n is lower than or equal to some quantity ${q_n}^0$;otherwise, it is equal to a limit L_n (which does not depend on q_n).

Let $h_s = 2$ and $a = 0.5$. We then obtain $Q_n = 453$ (the maximal possible held quantity) as well as $L_n = 422$. The fraction ${\alpha_n}^*(q_n, x_n^{q_n,d_n,K})$ and the held quantity are presented in Fig. 1 as functions of K, for several values of q_n.

The jumps in the figures show the existence of an upper bound on the supplier's held quantity. Up to this bound, the supplier maintains sufficient stock to fill the customer's order, thereby avoiding sanctions. Beyond that bound the supplier incurs sanctions that are almost quantity-invariant sanctions. As a result, in cases in which the supplier is understocked, i.e., incurs sanctions, the optimal level of inventory held by the supplier is lower than the upper bound. Therefore we observe jumps in the figures from the upper bound (no sanction) to the new optimal level (quantity-invariant sanctions).

For values of q_n that are lower than or equal to 453, the customer can choose values of K that lead to a value of ${\alpha_n}^*$ that is very close to 1. As $p'(0) = 0$, these values of K must be lower than 1 (see (12)). However, they are found to be closer to 1 than to 0.999. If $q_n = 300$ or $q_n = 440$, then the held quantity is monotonically increasing in K. Once K exceeds 23 (in the former case) and 230 (in the latter), ${\alpha_n}^*$ becomes approximately 1 and the held quantity is equal to q_n. This situation does not change when K grows and tends to infinity. If the value of q_n is close to Q_n, then the customer can still force the supplier to hold the entire ordered quantity during the current period. However, once q_n exceeds Q_n, the maximal held quantity decreases. For instance, if $q_n = 450$, then the whole order quantity can be held by the supplier during the current period. Nevertheless, once q_n becomes greater than Q_n, then the maximal held quantity, which is lower than q_n, is even smaller than Q_n. When $q_n = 455$ (just slightly greater than Q_n), the maximal held quantity drops to 427 (but is still higher than L_n). If q_n increases further, then the held quantity monotonically tends to L_n, that is, it will never be greater than that limit (see $q_n = 600$ in Fig. 1).

The decrease of the maximal held quantity for values of q_n that exceed Q_n is due to the following. If the order quantity q_n is a little higher than L_n (up to the value of Q_n), then the supplier prefers to hold the whole quantity during period n, preventing sanctions and a penalty (in spite of paying higher holding costs). Once q_n exceeds the value of Q_n, the supplier knows he will not hold the whole customer's order quantity due to his holding cost. Hence, he expects to pay some penalty cost anyway (under (2)), and chooses to reduce the held quantity (to even a lower value than Q_n), as doing so does not substantially increase his penalty cost, whereas it does decrease his holding cost.

Moreover, the value of ${\alpha_n}^*$ is not necessarily increasing in K, and may even decrease (see the cases of $q_n = 450$ and $q_n = 455$ in Fig. 1). This situation is explained as follows. If the value of K is high, but not huge, then x_n is strictly concave, and there exists an optimal value ${\alpha_n}^*$ that is extremely close to 1, and $x_n({\alpha_n}^*)$ almost vanishes. This value is optimal, since reducing α_n beyond this value may significantly increase the total time that the customer spends on sanctioning activity. As a result, the supplier's penalty may grow substantially. On the other hand, once K grows even more and becomes very large, x_n becomes piecewise linear; it is constant (and equal to $x^{\max}$) in an interval $[0,\, 1{-}\delta)$ for some very small value δ, which includes ${\alpha_n}^*$. Now $x_n({\alpha_n}^*)$ becomes approximately $x^{\max}$, and in order to reduce his penalty cost, the supplier may increase the value of α_n. However, there is an alternative choice: the supplier may significantly reduce α_n, such that the penalty cost remains the same (but does not grow), but the holding cost decreases significantly. As our example shows, the latter possibility may be preferable for the supplier.

Fig. 2 below presents how the maximal held quantities (as functions of q_n) are affected by the parameters h_s and a. As expected, these quantities become lower as the supplier's unit holding cost h_s grows, and higher when the penalty cost p becomes more rigorous (i.e., a increases). If h_s=2 and $a = 0.5$, then Q_n is equal to 453 and $L_n = 422$,

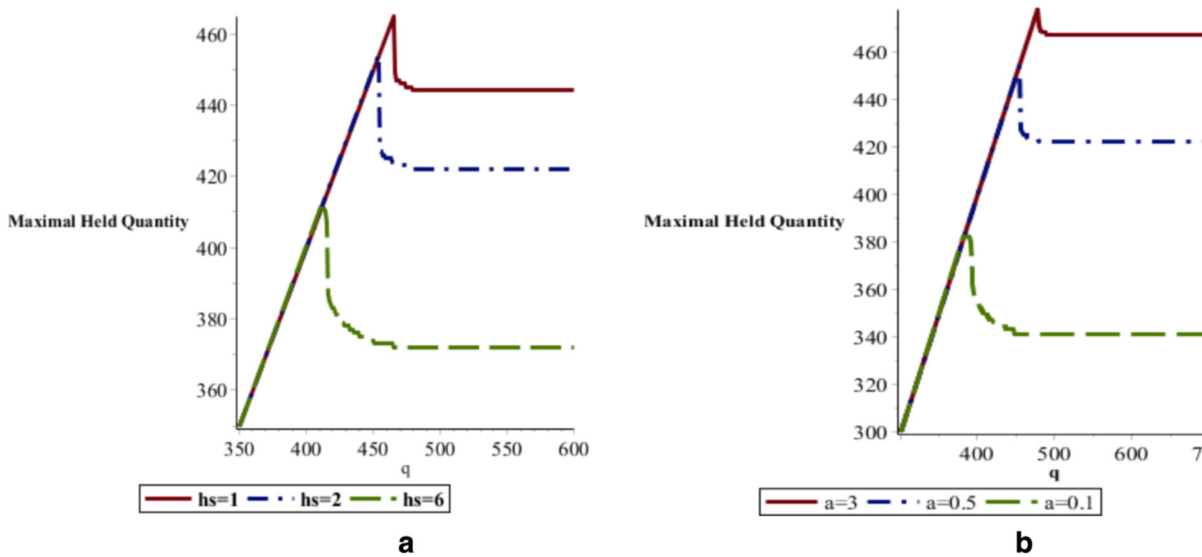

Fig. 2. The impact of the supplier's holding cost (a) and punishment level (b) on held quantities.

as we found above. When h_s decreases to 1 or increases to 6, then Q_n increases to 463 (and $L_n = 446$) or decreases to 410 (and $L_n = 372$), respectively. If a decreases to 0.1 or increases to 3, the maximal quantity Q_n becomes 381 or 478, respectively. The corresponding limit value L_n is equal to 340 or to 467, respectively.

We now go through the customer's periodic problem.

4. The customer's problem

4.1. Model setup

In order to satisfy his periodic demand, the customer makes all decisions such that $I_n + \alpha_n q_n \geq D_n$ necessarily holds (see below). The customer's periodic cost is

$$\begin{aligned} &C_n^c\left(I_n, q_n, x_n^{q_n,D_n}, D_n\right) \\ &= c\cdot\left(1-\frac{\alpha_n^*(q_n,x_n^{q_n,D_n})q_n}{D_n}\right)\tau x_n^{q_n,D_n}\left(\alpha_n^*(q_n,x_n^{q_n,D_n})\right)+wq_n \\ &\quad + \left(I_n - \frac{D_n}{2}+\alpha_n^*(q_n,x_n^{q_n,D_n})q_n - \frac{\left(\alpha_n^*(q_n,x_n^{q_n,D_n})\right)^2 q_n^2}{2D_n}\right)h_c \end{aligned} \quad (15)$$

while $\alpha_n q_n \leq D_n \leq I_n + \alpha_n q_n$, or

$$C_n^c\left(I_n,\ q_n,\ x_n^{q_n,d_n},\ D_n\right) = wq_n + I_n h_c, \quad (16)$$

if $D_n \leq \alpha_n q_n$. The first term on the right-hand side of (15) is the cost that the customer pays due to sanctioning (recall that the expression $(1-\frac{\alpha_n^*(q_n,x_n^{q_n,D_n})q_n}{D_n})\tau x_n^{q_n,D_n}(\alpha_n^*(q_n,x_n^{q_n,D_n}))$ represents the total time that the customer devotes to sanctions, as described in Section 3.1 above). The third term on the right-hand side of (15) refers to the customer's holding cost, which is obtained as follows. The customer first uses the inventory that is replenished by the supplier. If this inventory does not satisfy the periodic demand (i.e. $D_n > \alpha_n q_n$), the customer then uses the inventory from his own warehouse (and may also implement sanctions at the same time).

Based on (15) and (16) above, the customer's expected periodic cost (before the periodic realized demand is known), as a function of I_n, q_n and x_n, is equal to

$$\begin{aligned} &(EC_n^c)(I_n, q_n, x_n^{q_n,D_n}) \\ &= wq_n + I_n h_c + \int_{\alpha_n q_n}^{\infty}\left(c\cdot\left(1-\frac{\alpha_n^*(q_n,x_n^{q_n,d_n})q_n}{d_n}\right)\tau x_n^{(q_n,d_n)}(\alpha_n)\right. \\ &\quad \left. + \left(\alpha_n^*(q_n,x_n^{q_n,d_n})q_n - \frac{d_n}{2} - \frac{\left(\alpha_n^*(q_n,x_n^{q_n,d_n})\right)^2 q_n^2}{2d_n}\right)h_c\right) \\ &\quad \times f_n(d_n)\mathrm{d}d_n. \end{aligned} \quad (17)$$

Similarly to the supplier, the customer's goal is to minimize his total expected cost-to-go, and we denote

$$B_n^c(I_n) = \min_{\left(q_n,\ldots q_N, x_n^{q_n,D_n},\ldots, x_N^{q_N,D_N}\right)} \mathrm{E}\left(\sum_{m=n}^{N} C_m^c(I_m, q_m, x_m^{q_m,D_m}, D_m)\right). \quad (18)$$

According to the Bellman equation,

$$\begin{cases} B_N^c(I_N) = \min\limits_{q_N, x_N^{q_N,D_N}} (EC_N^c)(I_N, q_N, x_N^{q_N,D_N}) \\ B_n^c(I_n) = \min\limits_{q_n, x_n^{q_n,D_n}} \left((EC_n^c)(I_n, q_n, x_n^{q_n,D_n}) + \mathrm{E}\left[B_{n+1}^c(I_{n+1})\right]\right), n < N, \end{cases} \quad (19)$$

such that $I_{n+1} = I_n + q_n - D_n$.

Based on (19), we define the following objective function:

$$\begin{cases} J_N^c(I_N, q_N, x_N^{q_N,D_N}) = (EC_N^c)(I_N, q_N, x_N^{q_N,D_N}) \\ J_n^c(I_n, q_n, x_n^{q_n,D_n}) = (EC_n^c)(I_n, q_n, x_n^{q_n,D_n}) \\ \qquad + \mathrm{E}\left[B_{n+1}^c(I_n + q_n - D_n)\right], n < N, \end{cases} \quad (20)$$

receiving

$$B_n^c(I_n) = \min_{q_n, x_n^{q_n,D_n}} J_n^c\left(I_n, q_n, x_n^{q_n,D_n}\right).$$

In order to prevent inventory shortages during period n, the customer must ensure that his total available inventory during this period, $I_n + \alpha_n q_n$, meets the demand D_n. In other words, the customer has to determine his order quantity and enforcement level such that

$$\Pr\left(D_n > I_n + \alpha_n^*(q_n, x_n)\cdot q_n\right) \leq \varepsilon \quad (21)$$

for a small positive number ε. Constraint (21) implies that, theoretically, a situation of a shortage is possible for every small $\varepsilon > 0$ if the demand probability distribution is not bounded. Since the probability for such shortage is very small, this approximation does not affect the main results. Let $d_n^{\max} = F_n^{-1}(1-\varepsilon)$; then inequality (21) is satisfied if and only if

$$I_n + \alpha_n^*(q_n, x_n)\cdot q_n \geq d_n^{\max}. \quad (22)$$

According to the previous section, the left-hand side of (22) is bounded from above by $I_n + Q_n$, so a necessary condition for the existence of a solution that satisfies (22) is

$$I_n \geq {d_n}^{max} - Q_n \tag{23}$$

(it should be noted that as the supplier knows how D_n is distributed, the value of Q_n never exceeds ${d_n}^{\max}$). Furthermore, the initial inventory of the next period, $n+1$, must be taken into account, and the current order quantity q_n must be high enough such that I_{n+1} will necessarily be greater than or equal to ${d_{n+1}}^{\max} - Q_{n+1}$. That is,

$$\Pr(I_{n+1} < {d_{n+1}}^{max} - Q_{n+1}) \leq \varepsilon \tag{24}$$

(at the beginning of period n, the inventory I_{n+1} is a random variable that depends on D_n). Thus, q_n must satisfy

$$q_n \geq \ {d_n}^{max} - I_n + d_{n+1}^{max} - \ Q_{n+1}. \tag{25}$$

If the initial inventory I_n satisfies (23) but is very close to ${d_n}^{\max} - Q_n$, then the customer is in need of a high quantity held by the supplier which is close to the maximal possible capacity Q_n. As we saw in Section 3 above, the customer must order exactly that quantity, and no more. However, from (25) we find that q_n has to be higher than ${d_n}^{\max} - I_n$, which is very close to Q_n (while I_n is approximately ${d_n}^{\max} - Q_n$). That is, q_n must be greater than Q_n. Consequently there may be no value of q_n that satisfies both (22) and (25). As a result, a more rigorous constraint regarding the initial inventory is required, and, accordingly, based on the previous section, we replace Q_n with L_n in constraints (23) and (25), for $n < N-1$. In this situation, the customer needs the supplier to hold for him L_n units of inventory at most during the current period. The customer can increase his order q_n as much as required, provided that the quantity being held does not drop lower than L_n.

We now formulate the customer's periodic problem. Assuming that

$$I_n > {d_n}^{max} - L_n, \tag{26}$$

the customer solves the following problem:

$$\min_{q_n, x_n^{q_n, D_n}} J_n^c(q_n, x_n^{q_n, D_n}) \tag{27}$$

$$\text{s.t. } \alpha_n^*(q_n, x_n^{q_n, D_n}) \cdot q_n \geq {d_n}^{\max} - I_n \tag{28}$$

$$q_n \geq {d_n}^{max} - I_n + d_{n+1}^{max} - L_{n+1}. \tag{29}$$

In the last period, the customer only has to ensure that his available inventory satisfies constraint (28). Consequently, if the initial inventory I_N is lower than ${d_N}^{\max}$, the customer orders up to that bound. Otherwise, the customer orders nothing, as indicated in the following proposition.

Proposition 3. The customer's optimal order quantity in the last period is

$$q_N{}^*(I_N) = \left({d_N}^{max} - I_N\right)^+. \tag{32}$$

Proof. To find an optimal order quantity, we differentiate objective function (31).

$$\frac{\partial J_N^c(I_N,\ q_N,\ 0)}{\partial q_N} = A + \int_{q_N}^{\infty} \left(1 - \frac{q_N}{d_N}\right) h_c f_N(d_N) \mathrm{d}d_N > 0.$$

Since the derivative is positive, the objective function is increasing in q_n. Therefore, the optimal order quantity is the minimal non-negative value that meets constraint (28). □

Substituting the optimal quantity (32) in (31), the minimal expected cost at period N is as follows. If $I_N < {d_N}^{\max}$, then

$$\begin{aligned} B_N^c(I_N) = {} & w({d_N}^{\max} - I_N) + I_N h_c + \int_{{d_N}^{\max} - I_N}^{\infty} \\ & \times \left({d_N}^{\max} - I_N - \frac{d_N}{2} - \frac{({d_N}^{\max} - I_N)^2}{2 d_N} \right) h_c f_N(d_N) \mathrm{d}d_N. \end{aligned} \tag{33}$$

Otherwise,

$$B_N^c(I_N) = \left(I_N - \frac{\mathrm{E}[D_N]}{2} \right) h_c. \tag{34}$$

Next, we consider period $N-1$.

Let $n = N-1$. According to (30), constraint (29) is necessarily satisfied for every order quantity $q_{N-1} \geq 0$. The customer has to replenish his inventory up to ${d_{N-1}}^{\max}$, and if the initial inventory I_{N-1} is sufficiently large to meet (23), then this is possible. Moreover, as mentioned in the previous section, when $0 \leq q_{N-1} \leq Q_{N-1}$, the customer is able to force the supplier to hold his entire order quantity during that period, that is, $\alpha_{N-1}=1$, preventing any sanction cost. Thus, the customer's objective function (obtained by (17), (20), (33) and (34)) is

$$\begin{aligned} J_{N-1}^c(q_{N-1}, x_{N-1}) = {} & w q_{N-1} + I_{N-1} h_c + \int_{q_{N-1}}^{\infty} \left(\left(q_{N-1} - \frac{d_{N-1}}{2} - \frac{q_{N-1}^2}{2 d_{N-1}} \right) h_c \right) f_{N-1}(d_{N-1}) \mathrm{d}d_{N-1} \\ & + \int_0^{I_{N-1} + q_{N-1} - {d_N}^{\max}} \left(I_{N-1} + q_{N-1} - d_{N-1} - \frac{\mathrm{E}[D_N]}{2} \right) h_c f_{N-1}(d_{N-1}) \mathrm{d}d_{N-1} + \int_{I_{N-1} + q_{N-1} - {d_N}^{\max}}^{\infty} \\ & \times \left[\begin{array}{l} w({d_N}^{\max} - I_{N-1} - q_{N-1} + d_{N-1}) + (I_{N-1} + q_{N-1} - d_{N-1}) h_c + \\ \int_{{d_N}^{\max} - I_{N-1} - q_{N-1} + d_{N-1}}^{\infty} \left({d_N}^{\max} - I_{N-1} - q_{N-1} + d_{N-1} - \frac{d_N}{2} - \frac{({d_N}^{\max} - I_{N-1} - q_{N-1} + d_{N-1})^2}{2 d_N} \right) h_c f_N(d_N) \mathrm{d}d_N \end{array} \right] f_{N-1}(d_{N-1}) \mathrm{d}d_{N-1}. \end{aligned} \tag{35}$$

In Sections 4.2 and 4.3 we solve problem (27)–(29). First, we find the optimal solution in the last two periods.

4.2. The last two periods

Let $n = N$. During the current period, the supplier must hold the customer's entire order quantity, since there is no "next period" towards which he can supply the rest of the inventory. Hence,

$$Q_N = L_N = \infty, \tag{30}$$

and $x_n \equiv 0$, and according to (17) and (20), the customer's objective function is equal to

$$J_N^c(I_N,\ q_N,\ 0) = w q_N + I_N h_c + \int_{q_N}^{\infty} \left(q_N - \frac{d_N}{2} - \frac{q_N^2}{2 d_N} \right) h_c f_N(d_N) \mathrm{d}d_N. \tag{31}$$

If the initial inventory is lower than ${d_{N-1}}^{\max}$, then the customer may order ${d_{N-1}}^{\max} - I_{N-1}$ units of inventory. He does not need to order a greater quantity than this, since at the beginning of the next period, after the demand of the current period is realized, he can observe his initial inventory I_N and order (as well as receive) as much as required. Indeed, the objective function (35)) is increasing, and the following proposition is obtained.

Proposition 4. Assume that I_{N-1} satisfies (23). Then the customer's optimal order quantity at period $N-1$ is

$$q_{N-1}^*(I_{N-1}) = ({d_{N-1}}^{max} - I_{N-1})^+, \tag{36}$$

and the optimal enforcement level ${x_{N-1}}^*$ (for $I_{N-1} < {d_{N-1}}^{\max}$) is one that leads to $\alpha_{N-1}^*((d_{N-1}^{\max} - I_{N-1})_+, x_{N-1}^*) = 1$. □

We now examine the customer's problem in earlier periods.

4.3. *Period* n, n < N−1

At the beginning of the current period, the customer's initial inventory I_n is assumed to meet (26). The ordered quantity q_n and the enforcement level x_n must satisfy constraints (28) and (29). We first calculate the customer's optimal enforcement level given the order quantity.

4.3.1. *Setting the optimal enforcement level*

Assume that the order quantity q_n has been set, satisfying (29). According to (20), we obtain the following remark.

Remark 3. Given the order quantity q_n, the only component of the objective function (20) that is affected by the enforcement level is the expected cost of the current period, i.e., $(\mathrm{EC}_n^{\mathrm{c}})(I_n, q_n, x_n)$. □

Given the order quantity q_n, the enforcement level x_n must be rigorous enough so that constraint (28) can be met. Additional rigorousness increases the supplier's held quantity $\alpha_n^*(q_n, x_n)\cdot q_n$ up to either Q_n (for $q_n \le Q_n$) or L_n (otherwise). Generally, a more rigorous policy leads to the supplier holding a larger inventory. This may reduce the customer's sanctioning cost but leads to higher holding costs. The customer needs the supplier to hold $(d_n{}^{\max}-I_n)^+$ units for him during the current period. As shown in Section 3, there exists a nonempty set of enforcement levels, which we denote by $S(I_n, q_n)$, such that $\alpha_n^*(q_n,x_n)\cdot q_n \ge (d_n^{\max}-I_n)^+$ for every $x_n \in S(I_n,q_n)$. As a result, the customer is able to set the following policy (for $q_n > 0$):

$$\tilde{x}_n(\alpha_n) = \begin{cases} x_n(\alpha_n), & \text{if } 0 \le \alpha_n < \frac{(d_n^{\max}-I_n)^+}{q_n} \\ 0, & \text{otherwise}, \end{cases} \tag{37}$$

for any $x_n \in S(I_n, q_n)$, receiving the following:

Proposition 5. Let $x_n \in S(I_n,q_n)$ such that $(\mathrm{EC}_n{}^{\mathrm{s}})(\alpha_n)$ is quasi-convex. Under policy (37), the supplier sets $\alpha_n^* = \frac{(d_n^{\max}-I_n)^+}{q_n}$, that is, the held quantity is equal to exactly $(d_n{}^{\max}-I_n)^+$. Furthermore, $\tilde{x}_n$ is the customer's optimal enforcement level.

Proof. Let $\alpha_n{}^{**}$ be the value that minimizes $\mathrm{EC}_n{}^{\mathrm{s}}$ under q_n and x_n; then $\alpha_n{}^{**} \ge \alpha_n{}^*$. As $\mathrm{EC}_n{}^{\mathrm{s}}$ is quasi-convex, it is decreasing in the interval $[0, \alpha_n{}^{**}]$, and in particular in $[0, \alpha_n{}^*]$. As a result, $\mathrm{EC}_n{}^{\mathrm{s}}$ is decreasing in $[0,\alpha_n{}^*)$ under $\tilde{x}_n$ as well. Since $\tilde{x}_n$ is integrable in $[0,\alpha_n{}^*]$ (piecewise continuous), $\mathrm{EC}_n{}^{\mathrm{s}}$ is continuous in that interval. Namely, for every $0 \le \alpha_n < \alpha_n{}^*$,

$$(\mathrm{EC}_n^{\mathrm{s}})(q_n,\ \tilde{x}_n,\ \alpha_n) \ge (\mathrm{EC}_n^{\mathrm{s}})(q_n,\ \tilde{x}_n,\ \alpha_n^*). \tag{38}$$

The sanction policy $\tilde{x}_n$ vanishes in the interval $[\alpha_n{}^*, 1]$, and by substituting it in (8) we have

$$\frac{\partial(\mathrm{EC}_n^{\mathrm{s}})}{\partial\alpha_n} = \int_0^{\alpha_n q_n} q_n h_s f_n(d_n)\mathrm{d}d_n + \int_{\alpha_n q_n}^{\infty} \frac{\alpha_n q_n^2}{d_n} h_s f_n(d_n)\mathrm{d}d_n > 0$$

there. From the continuity of $\mathrm{EC}_n{}^{\mathrm{s}}$ we find that (38) is satisfied for every $\alpha_n{}^* \le \alpha_n \le 1$ as well.

Consequently, under $\tilde{x}_n$, $(\mathrm{EC}_n^{\mathrm{s}})(q_n,\tilde{x}_n,\alpha_n^*) = \min\limits_{0\le\alpha_n\le 1}\{(\mathrm{EC}_n^{\mathrm{s}})(q_n,\tilde{x}_n,\alpha_n)\}$.

Now consider the customer's expected cost under policy (37). Let x^n be an enforcement level such that constraint (28) is met, and let $\hat{\alpha}_n$ be the supplier's response. Then $\hat{\alpha}_n \cdot q_n \ge (\,d_n{}^{\max} - I_n)^+$. According to (17),

$$\begin{aligned}(\mathrm{EC}_n^{\mathrm{c}})(I_n,q_n,\hat{x}_n) &= wq_n + I_n h_{\mathrm{c}} + \int_{\hat{\alpha}_n q_n}^{\infty}\left(c\cdot\left(1-\frac{\hat{\alpha}_n q_n}{d_n}\right)\tau\hat{x}(\hat{\alpha}_n) + \left(\hat{\alpha}_n q_n - \frac{d_n}{2} - \frac{(\hat{\alpha}_n q_n)^2}{2d_n}\right)h_{\mathrm{c}}\right)f_n(d_n)\mathrm{d}d_n \\ &\ge wq_n + I_n h_{\mathrm{c}} + \int_{\hat{\alpha}_n q_n}^{\infty}\left(\hat{\alpha}_n q_n - \frac{d_n}{2} - \frac{(\hat{\alpha}_n q_n)^2}{2d_n}\right)h_{\mathrm{c}} f_n(d_n)\mathrm{d}d_n \\ &\ge wq_n + I_n h_{\mathrm{c}} + \int_{(D_n-I_n)^+}^{\infty}\left(\left((d_n^{\max}-I_n)^+ - \frac{d_n}{2} - \frac{\left((d_n^{\max}-I_n)^+\right)^2}{2d_n}\right)h_{\mathrm{c}}\right)f_n(d_n)\mathrm{d}d_n = (\mathrm{EC}_n^{\mathrm{c}})(I_n,q_n,\tilde{x}_n)\end{aligned}$$

(the second inequality is based on increasing the function $g(y) = \int_y^\infty (y - \frac{t}{2} - \frac{y^2}{2t})h_{\mathrm{c}} f(t)\mathrm{d}t$). □

We now calculate the optimal order quantity.

4.3.2. *Setting the optimal order quantity*

Unlike the optimal enforcement level, the optimal value of q_n is not calculated myopically, but dynamically, taking into account the derived expected costs of the next periods. Indeed, both the components of the right-hand side of the second term of (20) depend on q_n. The order quantity must lead to an enforcement level which satisfies constraint (28), and according to policy (37) and Proposition 5, constraint (28) is met as well. We now show that the optimal order quantity is

$$q_n{}^* = \left(d_n{}^{max} - I_n + d_{n+1}{}^{max} - L_{n+1}\right)^+. \tag{39}$$

The optimality of (39) is formulated in Proposition 6 below, and its proof is based on the following lemmas.

Lemma 1. The expected cost of the current period is increasing in q_n such that $\frac{\partial[(\mathrm{EC}_n^{\mathrm{c}})(I_n,q_n,\tilde{x}_n)]}{\partial q_n} = w$.

Proof. According to (37) and Proposition 5, (17) becomes

$$\mathrm{EC}_n^{\mathrm{c}} = wq_n + I_n h_{\mathrm{c}} + \int_{(d_n^{\max}-I_n)^+}^{\infty} \times\left((d_n^{\max}-I_n)^+ - \frac{d_n}{2} - \frac{\left((d_n^{\max}-I_n)^+\right)^2}{2d_n}\right)h_{\mathrm{c}} f_n(d_n)\mathrm{d}d_n.$$

□

Lemma 2. Let $n \le N-2$. If the Bellman function $B^{\mathrm{c}}{}_{n+1}(I_{n+1})$ satisfies $\frac{\mathrm{d}B_{n+1}^{\mathrm{c}}}{\mathrm{d}I_{n+1}} > -w$, then the objective function $J^{\mathrm{c}}{}_n(q_n)$ is increasing. Moreover, the Bellman function $B^{\mathrm{c}}{}_n(I_n)$ satisfies $\frac{\mathrm{d}B_n^{\mathrm{c}}}{\mathrm{d}I_n} > -w$ as well.

Proof. According to Remark 2, $B^{\mathrm{c}}{}_{n+1}(I_{n+1}) = B^{\mathrm{c}}{}_{n+1}\ (I_n+q_n-D_n)$, and given a value of D_n,

$$\frac{\partial B_{n+1}^{\mathrm{c}}}{\partial q_n} = \frac{\mathrm{d}B_{n+1}^{\mathrm{c}}}{\mathrm{d}I_{n+1}}\frac{\partial I_{n+1}}{\partial q_n} = \frac{\mathrm{d}B_{n+1}^{\mathrm{c}}}{\mathrm{d}I_{n+1}}\cdot 1 > -w. \tag{40}$$

As (40) is satisfied for every value of D_n, then the expected value of that derivative is greater than –w, and by the Leibniz Integral Rule we obtain

$$\frac{\partial \mathrm{E}[B_{n+1}^{\mathrm{c}}]}{\partial q_n} = \mathrm{E}\left[\frac{\partial B_{n+1}^{\mathrm{c}}}{\partial q_n}\right] > -w.$$

Based on this result, as well as on Lemma 1, the objective function satisfies

$$\frac{\partial J_n^{\mathrm{c}}}{\partial q_n} = \frac{\partial(\mathrm{EC}_n^{\mathrm{c}})}{\partial q_n} + \frac{\partial E[B_{n+1}^{\mathrm{c}}]}{\partial q_n} = w + \frac{\partial E[B_{n+1}^{\mathrm{c}}]}{\partial q_n} > w - w = 0. \tag{41}$$

As a result, $J^{\mathrm{c}}{}_n$ is increasing, and the optimal order quantity is (39).

Now consider $B^{\mathrm{c}}{}_n$. Since $B_n^{\mathrm{c}}(I_n) = J_n^{\mathrm{c}}(I_n, q_n^*(I_n), \tilde{x}_n(I_n, q_n^*))$, then according to the chain rule,

$$\frac{dB_n^c}{dI_n} = \left(\frac{\partial J_n^c}{\partial I_n} + \frac{\partial J_n^c}{\partial q_n}\frac{dq_n^*}{dI_n} + \frac{\partial J_n^c}{\partial \tilde{x}_n}\frac{d\tilde{x}_n}{dI_n}\right)\Bigg|_{q_n=q_n^*}$$
$$= \left(\frac{\partial J_n^c}{\partial I_n} + \frac{\partial J_n^c}{\partial q_n}\frac{dq_n^*}{dI_n} + \frac{\partial J_n^c}{\partial \tilde{x}_n}\cdot 0\right)\Bigg|_{q_n=q_n^*}$$
$$= \left(\frac{\partial J_n^c}{\partial I_n} + \frac{\partial J_n^c}{\partial q_n}\frac{dq_n^*}{dI_n}\right)_{q_n=q_n^*}. \tag{42}$$

Define $\varphi_n(I_n,q_n)=E[B^c{}_{n+1}(I_n+q_n-D_n)]$. Then according to (20) and Lemma 1

$$\frac{\partial J_n^c}{\partial q_n} = \frac{\partial(EC_n^c)}{\partial q_n} + \frac{\partial\varphi_n}{\partial q_n} = w + \frac{\partial\varphi_n}{\partial q_n}.$$

If $I_n \le d_n{}^{\max}$, then

$$\frac{\partial J_n^c}{\partial I_n} = \frac{\partial(EC_n^c)}{\partial I_n} + \frac{\partial\varphi_n}{\partial I_n}$$
$$= \left[1 - \int_{d_n{}^{\max}-I_n}^{\infty}\left(1 - \frac{d_n{}^{\max}-I_n}{d_n}\right)f_n(d_n)dd_n\right]h_c + \frac{\partial\varphi_n}{\partial I_n}$$

Note that

$$\frac{\partial\varphi_n}{\partial q_n} = \frac{\partial\varphi_n}{\partial(I_n+q_n)}\frac{\partial(I_n+q_n)}{\partial q_n} = \frac{\partial\varphi_n}{\partial(I_n+q_n)}$$

as well as

$$\frac{\partial\varphi_n}{\partial I_n} = \frac{\partial\varphi_n}{\partial(I_n+q_n)}\frac{\partial(I_n+q_n)}{\partial I_n} = \frac{\partial\varphi_n}{\partial(I_n+q_n)},$$

and therefore $\frac{\partial\varphi_n}{\partial q_n} = \frac{\partial\varphi_n}{\partial I_n}$.

Also, according to (39), $\frac{dq_n^*}{dI_n} = -1$. Thus, (42) becomes

$$\frac{dB_n^c}{dI_n} = \left(\frac{\partial J_n^c}{\partial I_n} - \frac{\partial J_n^c}{\partial q_n}\right)\Big|_{q_n=q_n^*}$$
$$= \left[1 - \int_{d_n{}^{\max}-I_n}^{\infty}\left(1 - \frac{d_n{}^{\max}-I_n}{d_n}\right)f_n(d_n)dd_n\right]h_c + \frac{\partial\varphi_n}{\partial I_n}\Big|_{q_n=q_n^*}$$
$$- w - \frac{\partial\varphi_n}{\partial q_n}\Big|_{q_n=q_n^*}$$
$$> \left[1 - \int_{d_n{}^{\max}-I_n}^{\infty} f_n(d_n)dd_n\right]h_c - w$$
$$> \left[1 - \int_0^{\infty} f_n(d_n)dd_n\right]h_c - w = -w.$$

If $d_n{}^{\max} \le I_n \le d_n{}^{\max} + d_{n+1}{}^{\max} - L_{n+1}$, then

$$\frac{\partial J_n^c}{\partial I_n} = \frac{\partial(EC_n^c)}{\partial I_n} + \frac{\partial\varphi_n}{\partial I_n} = h_c + \frac{\partial\varphi_n}{\partial I_n} \tag{43}$$

and $\frac{dq_n^*}{dI_n} = -1$. Thus, (42) is equal to

$$\frac{dB_n^c}{dI_n} = \left(\frac{\partial J_n^c}{\partial I_n} - \frac{\partial J_n^c}{\partial q_n}\right)\Big|_{q_n=q_n^*}$$
$$= h_c + \frac{\partial\varphi_n}{\partial I_n}\Bigg|_{q_n=q_n^*} - w - \frac{\partial\varphi_n}{\partial q_n}_{q_n=q_n^*} > -w.$$

If $I_n > d_n{}^{\max} + d_{n+1}{}^{\max} - L_{n+1}$, then (43) is valid and $\frac{dq_n^*}{dI_n} = 0$. Thus, (42) becomes

$$\frac{dB_n^c}{dI_n} = \frac{\partial J_n^c}{\partial I_n}\Big|_{q_n=q_n^*} = h_c + \frac{\partial\varphi_n}{\partial I_n}\Big|_{q_n=q_n^*}$$
$$= h_c + \left(\frac{\partial}{\partial I_n}E\left[B_{n+1}^c(I_n+q_n-D_n)\right]\right)\Big|_{q_n=q_n^*}$$
$$= h_c + E\left[\frac{\partial}{\partial I_n}B_{n+1}^c(I_n+q_n-D_n)\right]\Big|_{q_n=q_n^*}$$
$$= h_c + E\left[\frac{d}{dI_{n+1}}B_{n+1}^c(I_{n+1})\frac{\partial I_{n+1}}{\partial I_n}\right]\Big|_{q_n=q_n^*}$$
$$= h_c + E\left[\frac{dB_{n+1}^c(I_{n+1})}{dI_{n+1}}\right]\Big|_{q_n=q_n^*} > h_c - w > -w.$$

Corollary 1. If $\frac{dB_{n_0}^c}{dI_{n_0}} > -w$ for $n_0 \le N-1$, then for every $n < n_0$ the objective function $J^c{}_n(q_n)$ is increasing.

Proof. According to the first result of Lemma 2, our assumption leads to an increase of $J_{n_0-1}^c(q_{n_0-1})$. Moreover, from the second result of that lemma we have $\frac{dB_{n_0-1}^c}{dI_{n_0-1}} > -w$. Satisfying Lemma 2 again, we find that $J_{n_0-2}^c(q_{n_0-2})$ is increasing as well, and that $\frac{dB_{n_0-2}}{dI_{n_0-2}} > -w$, and so on. □

We are now ready to formulate the main result of the current subsection. The optimal order quantity is the minimal one that ensures that the demand of the current period is satisfied and that the next period begins with a basic initial inventory.

Proposition 6. The optimal order quantity at period $n < N-1$ is (39).

Proof. Let $n_0 = N-1$. Then according to (35) and (36), if $I_{N-1} < d_{N-1}{}^{\max}$ then

$$B_{N-1}^c(I_{N-1}) = w(d_{N-1}^{\max} - I_{N-1}) + I_{N-1}h_c$$
$$+ \int_{d_{N-1}^{\max}-I_{N-1}}^{\infty}\left(\left(d_{N-1}^{\max} - I_{N-1} - \frac{d_{N-1}}{2} - \frac{(d_{N-1}^{\max}-I_{N-1})^2}{2d_{N-1}}\right)h_c\right)f_{N-1}(d_{N-1})dd_{N-1} + \int_0^{d_{N-1}^{\max}-d_N^{\max}}\left(d_{N-1}^{\max} - d_{N-1} - \frac{E[D_N]}{2}\right)h_c f_{N-1}(d_{N-1})dd_{N-1}$$
$$+ \int_{d_{N-1}^{\max}-d_N^{\max}}^{\infty}\left[\begin{array}{l} A(d_N^{\max} - d_{N-1}^{\max} + d_{N-1}) + (d_{N-1}^{\max} - d_{N-1})h_c + \\ \int_{d_N^{\max}-d_{N-1}^{\max}+d_{N-1}}^{\infty}\left(d_N^{\max} - d_{N-1}^{\max} + d_{N-1} - \frac{d_N}{2} - \frac{(d_N^{\max}-d_{N-1}^{\max}+d_{N-1})^2}{2d_N}\right)h_c f_N(d_N)dd_N \end{array}\right]f_{N-1}(d_{N-1})dd_{N-1}$$

that is

$$\frac{dB_{N-1}^c(I_{N-1})}{dI_{N-1}} = -w + \left[1 - \int_0^{\infty}\left(1 - \frac{d_{N-1}^{\max}-I_{N-1}}{d_{N-1}}\right)f_{N-1}(d_{N-1})dd_{N-1}\right]h_c > -w + \left[1 - \int_0^{\infty} f_{N-1}(d_{N-1})dd_{N-1}\right]h_c = -w.$$

On the other hand, if $I_{N-1} > d_{N-1}{}^{\max}$ then

$$B_{N-1}^c(I_{N-1}) = I_{N-1}h_c - \frac{E[D_{N-1}]}{2}h_c + \int_0^{I_{N-1}-d_N^{\max}}\left(I_{N-1} - d_{N-1} - \frac{E[D_N]}{2}\right)h_c f_{N-1}(t_{N-1})dd_{N-1}$$
$$+ \int_{I_{N-1}-d_N^{\max}}^{\infty}\left[\begin{array}{l} w(d_N^{\max} - I_{N-1} + d_{N-1}) + (I_{N-1} - d_{N-1})h_c + \\ \int_{d_N^{\max}-I_{N-1}+d_{N-1}}^{\infty}\left(d_N^{\max} - I_{N-1} + d_{N-1} - \frac{d_N}{2} - \frac{(d_N^{\max}-I_{N-1}+d_{N-1})^2}{2d_N}\right)h_c f_N(d_N)dd_N \end{array}\right]f_{N-1}(d_{N-1})dd_{N-1}$$

namely,

$$\begin{aligned}
\frac{\mathrm{d}B^{c}_{N-1}(I_{N-1})}{\mathrm{d}I_{N-1}} &= h_c + \int_0^{I_{N-1}-d_N^{\max}} h_c f_{N-1}(d_{N-1})\mathrm{d}d_{N-1} + \int_{I_{N-1}-d_N^{\max}}^{\infty}\left[-w + h_c - \int_{d_N^{\max}-I_{N-1}+d_{N-1}}^{\infty}\left(1-\frac{d_N^{\max}-I_{N-1}+d_{N-1}}{d_N}\right)h_c f_N(d_N)\mathrm{d}d_N\right] \\
&\quad \times f_{N-1}(d_{N-1})\mathrm{d}d_{N-1} \\
&= -w(1-F_{N-1}(I_{N-1}-d_N^{\max})) + 2h_c - \int_{I_{N-1}-d_N^{\max}}^{\infty}\left[\int_{d_N^{\max}-I_{N-1}+d_{N-1}}^{\infty}\left(1-\frac{d_N^{\max}-I_{N-1}+d_{N-1}}{t_N}\right)h_c f_N(d_N)\mathrm{d}d_N\right]f_{N-1}(d_{N-1})\mathrm{d}d_{N-1} \\
&> -w + 2h_c - \int_{I_{N-1}-d_N^{\max}}^{\infty}\left[\int_{d_N^{\max}-I_{N-1}+d_{N-1}}^{\infty} h_c f_N(d_N)\mathrm{d}d_N\right]f_{N-1}(t_{N-1})\mathrm{d}d_{N-1} > -w + 2h_c - \int_0^{\infty}\left[\int_0^{\infty} h_c f_N(d_N)\mathrm{d}d_N\right]f_{N-1}(d_{N-1})\mathrm{d}d_{N-1} \\
&= -w + h_c > -w.
\end{aligned}$$

According to Corollary 1, the solution (39) is optimal for every $n < N-1$. □

Having determined the customer's periodic optimal solution, we present numerical examples in the next subsection.

4.4. Numerical illustrations

In our example, we consider a one-year contract such that every quarter is a "period", namely, $N = 4$ and $\tau = 90$. As in our previous examples, the customer's enforcement level is according to (2), and the penalty that the customer inflicts on the supplier is calculated as in (14). Let w= 10, $c = 50$, $x^{\max} = 3$, $a = 0.5$ and $\varepsilon = 0.025$, and assume that the demands of the four quarters are uniformly distributed such that $D_1 \sim U[0,200]$, $D_2 \sim U[0,500]$, $D_3 \sim U[0,800]$ and $D_4 \sim U[0,300]$. Showing how the customer's expected total cost is affected by the model's various parameters, we compare the results of our model to two benchmarks. The first one ("Benchmark 1") is based on the "ideal" case. The supplier always holds the whole ordered quantity q_n during period n. The second one ("Benchmark 2") corresponds to the "worst" case. The supplier does not hold any inventory during period n, and the customer receives his entire order at the end of that period. Fig. 3 presents the impact of the parameters h_c, h_s, $x^{\max}$ and a on the customer's optimal cost, $B^{c}{}_{1}$ (the rest of the parameters in each graph are assumed to be as given above). To be precise, this cost includes the order cost of the initial inventory, i.e., $w \cdot I_1$, since as h_s increases or alternatively, $x^{\max}$ or a decreases, the supplier reduces the held quantity (and the customer's costs grow). As a result, the customer must hold greater initial inventories in order to meet probable demand. Therefore, the initial inventories are assumed to be the minimal values that ensure that this demand is met.

As the unit holding costs h_c and h_s grow, a situation in which the supplier holds the entire order quantity becomes more effective for the customer. According to Fig. 3a, the customer's expected cost under our model is only slightly higher than that of Benchmark 1 (but significantly lower than that of Benchmark 2). If h_c=h_s=2, then this cost is equal to 14,272 in Benchmark 1 and to 14,567 in our model. That is to say, the deterrence directed toward the supplier is effective, such that the customer's actual cost is only 2 percent higher than that of the ideal case. Moreover, the actual cost is reduced by 22 percent compared to Benchmark 2 (the latter is 18719). If h_c and h_s increase to 5, then the total expected costs under Benchmark 1, our model and Benchmark 2 are, respectively 16,480, 17,817, and 25,365. The cost under our model is 8 percent higher than Benchmark 1, and 30 percent lower than Benchmark 2.

The maximal number of hours devoted to sanctions each day affects the customer's costs in our model (in contrast to Benchmarks 1 and 2, where the costs are fixed at 14,272 and 18,719, respectively), as shown in Fig. 3b. If $x^{\max} = 3$, then the customer's expected total cost is equal to 14,567, as before. If $x^{\max}$ increases to 5, then the cost decreases to 14,439 (i.e., it is only 1 percent higher than that of Benchmark 1). Even if $x^{\max}$ is lower and is equal to 0.5, the cost, 15,806 in our model, is reduced by 16 percent, compared with that of Benchmark 2. Therefore, the ability to devote even 1 hour each day to sanctions may constitute a significant deterrence measure. Similarly to $x^{\max}$, the purchasing department's punishment level, a in our example, affects the customer's expected cost, as shown in Fig. 3c.

In our research, we assume that shortages are unacceptable to the customer, and that he orders sufficient inventory to ensure that the entire demand will be satisfied. Consequently, if demand uncertainty is high, then the total quantity that the customer orders ($\sum_n q_n$) is supposed to significantly exceed the actual required capacity ($\sum_n D_n$). As a result, the customer's total cost is expected to grow. Let the customer's initial inventory be $I_1 = 80$; then Fig. 4 illustrates how the customer's total order as well as his cost depends on variance in demand. Values on the horizontal axis represent the levels of variance compared with those of the distributions given above ("100 percent"

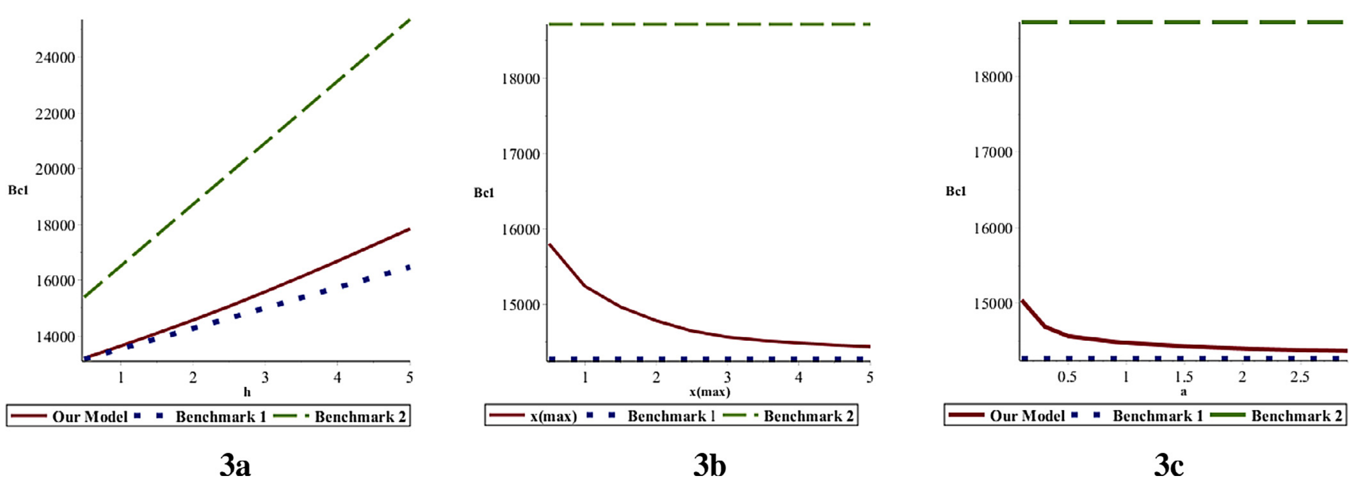

Fig. 3. Customer's expected total costs as function of the parameters h_c and h_s (3a), $x^{\max}$ (3b) and a (3c).

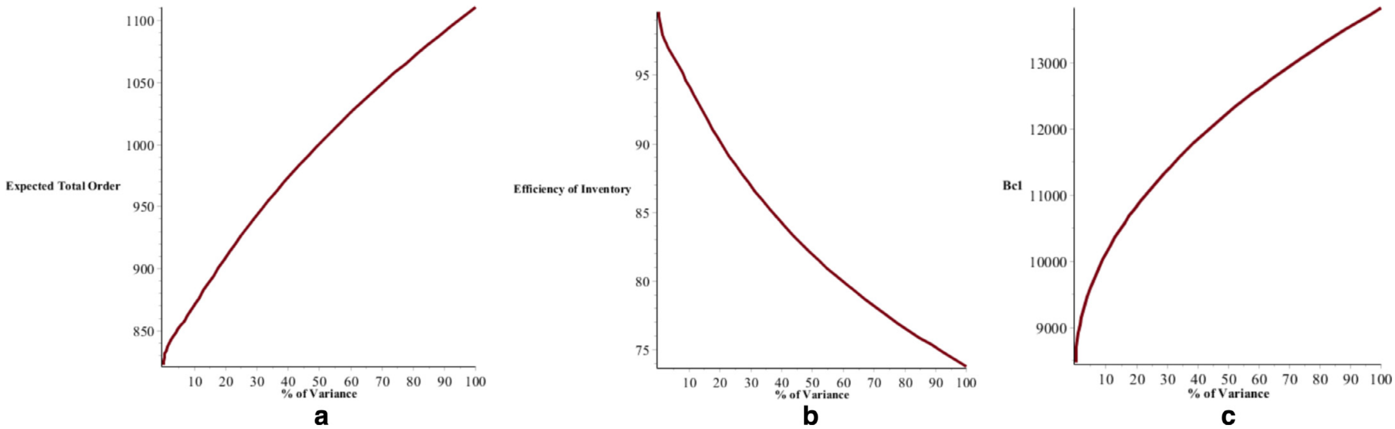

Fig. 4. Influence of demand uncertainty on the customer's total order quantities (a), percentage of used inventory (b) and total expected cost (c).

means that they are the same as those in the original examples). All the means remain the same as in the original distributions.

If the possible ranges of the demands are reduced by half, then the demand distributions are $D_1 \sim U[50,150]$, $D_2 \sim U[125,375]$, $D_3 \sim U[200,600]$ and $D_4 \sim U[75,225]$, and the corresponding variances constitute 25 percent of the original ones. In this case, the total ordered quantity is reduced by 17 percent, from 1111 to 927 (see Fig. 4a). The percentage of used inventory (out of the total ordered inventory) grows from 74 percent (820 of 1111) to 88 percent (820 of 927), as shown in Fig. 4b. According to Fig. 4c, the customer's total expected cost is reduced by 20 percent, from 13,832 to 11,123. As the variances tend to zero, the total ordered quantity decreases by 26 percent, to 820 (that is, the customer complements the initial inventory 80 to the exact quantity of the total deterministic demand 900). The percentage of inventory used tends to 100 percent, and the total expected cost is reduced by almost 40 percent, to 8372.

5. Conclusions

In order to ensure short replenishment lead times by suppliers, organizations have to enforce contracts and impose penalties. In this paper we show how a customer can effectively enforce on-time replenishment, by imposing sanctions on the supplier. Our model is based on a Stackelberg competition game. Initiating a enforcement level, the customer becomes the lead player and forces the supplier to hold inventory, which is made available to the customer in real time, according to the JIT approach. This policy, which is analytically formulated in a mathematical function referred to as the *enforcement level*, takes into account the stochastic demand distribution as well; the higher the demands are supposed to be, the greater the inventory the supplier holds for the customer. Using a class of sanctioning functions, we show that the customer can force the supplier to hold inventory up to some maximal value, such that actual enforcement of sanctions is unnecessary. This value depends on the total time that the customer can actually spend on enforcement each day; on the rigorousness of the penalty that the purchasing department inflicts on the supplier when he does not fulfill orders on time; on the holding cost; and on the demand distribution. Moreover, we find that, contrary to expectations, escalation of the customer's enforcement level may decrease the level of inventory that the supplier holds, thereby diminishing his capacity to replenish inventory on time.

The problem is solved using backward dynamic programming such that the total expected costs-to-go are minimized. The supplier's decision regarding the periodic held quantity (i.e., the inventory quantity made available for the customer), as well as the customer's decision regarding the periodic enforcement level, can be set myopically, taking into account only the expected cost of the current period. The customer's minimal periodic order quantity, which ensures that the customer is able to satisfy the current demand as well as to receive basic initial inventory for the next period, is shown to be optimal. Having the possibility to spend a few hours each day on sanctioning activity substantially reduces the customer's expected cost. In particular, in a numerical example based on the Israel Police's data, the customer's cost under an enforcement level was less than 10 percent higher than the cost in an "ideal" situation, in which all inventory was necessarily replenished on time. Under certain parameter values, the difference was reduced to less than 2 percent. Moreover, as the customer must avoid any shortages, high uncertainty regarding demand significantly increases the inventory quantity that he orders. In our examples, a decrease of 50 percent in the possible range of demand (for the same mean) was shown to reduce by almost 20 percent both the total order quantity and expected cost. When demand variance tended to zero, the customer's expected cost was reduced by 40 percent.

We propose several important yet challenging directions for future research. In this paper, we considered one organization that is committed to a unique supplier. It would be very interesting to research scenarios in which several organizations work with a single supplier. Furthermore, the possibility of adding an additional supplier for backup, as suggested by Kouvelis and Li (2008), may be integrated into our model. In this paper we made several assumptions, including (i) a specific class of sanctioning functions; (ii) random total demand in each period, but with constant demand rates over the course of the period; and (iii) full transparency between the supply chain parties, i.e., all unit costs and demand distributions were known to both the supplier and the customer. These assumptions could be relaxed. Likewise, some limited shortages could be allowed.

References

Ayers, J. B. (2001). *Handbook of supply chain management*. Boca Raton, FL: Auerbach Publication/Taylor and Francis Group.

Barnes, E., Dai, J., Deng, S., Down, D., Goh, M., Lau, H. C., & Sharafali, M. (2000). *Electronics manufacturing service industry*. Singapore: The Logistics Institute-Asia Pacific, National University of Singapore TLI-AP white paper.

Bellman, R. (1957). *Dynamic programming*. Princeton N.J: Princeton University Press.

Cachon, G., & Fisher, M. (1997). Campbell soup's continuous replenishment program: evaluation and enhanced inventory decision rules. *Production and Operations Management, 6*, 266–276.

Cachon, G. (2001). Stock wars: inventory competition in a two echelon supply chain with multiple retailers. *Operations Research, 49*(5), 658–674.

De Toni, A. F., & Zamolo, E. (2005). From a traditional replenishment system to vendor-managed inventory. *International Journal of Production Economics, 96*(1), 63–79.

Dukes, A. J., Geylani, T., & Srinivasan, K. (2009). Strategic assortment reduction by a dominant retailer. *Marketing Science, 28*, 309–319.

Fang, F., & Whinston, A. (2007). Option contracts and capacity management – enabling price discrimination under demand uncertainty. *Production and Operations Management, 16*(1), 1059–1478.

Fry, M. J., Kapuscinski, R., & Olsen, T. L. (2001). Coordination production and delivery under a (z,z)-type vendor managed inventory contract. *Manufacturing & Service Operations Management, 3*, 151–173.

Grout, J. R., & Christy, D. P. (1999). *A model of supplier responses to Just in Time delivery requirements*. Springer.

Harris, J. K., Swatman, P. M. C., & Kurnia, S. (1999). Efficient consumer response (ECR): a survey of the Australian grocery industry. *Supply Chain Management: An International Journal, 4*(1), 35–42.

Khmelnitsky, E., & Caramanis, M. (1998). One machine N-part-type optimal set-up scheduling: analytical characterization of switching surfaces. *IEEE Transactions on Automatic Control, 43*, 1584–1588.

Kogan, K., & Lou, S. (2002). Optimal production control: analytical solution for the limit cycles. *IIE Transactions, 34*(4), 363–374.

Kogan, K., & Tell, H. (2009). Production smoothing by balancing capacity utilization and advance orders. *IIE Transactions, 41*(3), 223–231.

Kogan, K., & Shnaiderman, M. (2010). Continuous-time replenishment under intermittent observability (Technical Note). *IEEE Transactions on Automatic Control, 55*(6)), 1460–1465.

Kouvelis, P., & Li, J. (2008). Flexible backup supply and the management of lead-time uncertainty. *Production and Operations Management, 17*(2), 184–199.

Lee, H. L., So, K. V., & Tang, C. S. (2000). The value of information sharing in a two-level supply chain. *Management Science, 46*(5), 626–643.

Matsui, K. (2010). Returns policy, new model introduction, and consumer welfare. *International Journal of Production Economics, 124*, 299–309.

Raghunathan, S., & Yeh, A. B. (2001). Beyond EDI: impact of continuous replenishment program (CRP) between a manufacturer and its retailers. *Information Systems Research, 12*(4), 406–419.

Savaşaneril, S., & Erkip, N. (2010). An analysis of manufacturer benefits under vendor-management-systems. *IEE Transactions, 42*, 455–477.

Schenk, J., & McInerny, J. (1998). Applying vendor managed inventory to the apparel industry. *Automatic I.D. News, 14*, 239–252.

Vigtil, A. (2007). Information exchange in vendor managed inventory. *International Journal of Physical Distribution & Logistics Management, 37*(2), 131–147.

Wang, X., & Liu, L. (2007). Coordination in a retailer-led supply chain through option contract. *International Journal of Production Economics, 110*, 115–127.

Xue, W., Demirag, O. C., & Niu, B. (2014). Supply chain performance and consumer surplus under alternative structures of channel dominance. *European Journal of Operational Research, 239*, 130–145.

Yonghui, F., & Raiesh, P. (2004). Supply-side collaboration and its value in supply chains. *European Journal of Operational Researc, 152*(1), 281–298.

Zhao, Y., Wang, S., Cheng, T. C. E., Yang, X., & Huang, Z. (2010). Coordination of supply chains by option contracts: a cooperative game theory approach. *European Journal of Operational Research, 207*, 668–675.