IIE Transactions (2002) **34**, 627–636

Base stock versus WIP cap in single-stage make-to-stock production–inventory systems

GEORGE LIBEROPOULOS[1] and YVES DALLERY[2]

[1]*Department of Mechanical and Industrial Engineering, University of Thessaly, Pedion Areos, GR-38334 Volos, Greece*
E-mail: glib@mie.uth.gr
[2]*Laboratoire Productique et Logistique, Ecole Centrale de Paris, Grande Voie des Vignes FR-92295 Chatenay-Malabry Cedex, France*
E-mail: dallery@pl.ecp.fr

Received November 2000 and accepted October 2001

A single-stage production–inventory system produces parts in a make-to-stock mode, and unsatisfied demand is backordered. The system operates under a so-called *base stock with WIP cap* replenishment policy, which works as follows. Whenever the Work-In-Process (WIP) plus finished goods inventory falls below a specified level, called *base stock*, a replenishment order for the production of a new part is issued. If the WIP inventory is below a different specified level, called *WIP cap*, the order goes through and a new part is released for production; otherwise, the order is put on hold until the WIP inventory drops below the WIP cap. First, it is shown that the optimal base stock that minimizes the long-run, average, inventory holding cost for a given minimum customer service level, is a non-increasing function of the WIP cap that reaches a minimum value, called *minimum optimal base stock*, at a finite WIP cap value, called *critical WIP cap*. Then, it is shown that the optimal WIP cap is less than or equal to the critical WIP cap and therefore the optimal base stock is greater than or equal to the minimum optimal base stock. More interestingly, however, it is conjectured that the optimal WIP cap is in fact exactly equal to the critical WIP cap and therefore the optimal base stock is exactly equal to the minimum optimal base stock. The minimum optimal base stock is none other than the optimal base stock of the same system operating under a classical base stock policy (with no WIP cap). Finally, the optimal parameters of a system operating under a base stock with WIP cap policy are related to the optimal parameter of the same system operating under a make-to-stock CONWIP policy.

1. Introduction

We consider a single-stage production–inventory system that operates in a make-to-stock mode. The system consists of a manufacturing facility where parts are processed, and an output store where finished parts are stored. Parts in the manufacturing facility are called *Work-In-Process* or WIP, and parts in the output store are called *Finished Goods* or FG. We make the following assumptions:

1. There is an infinite supply of raw parts and a single type of parts produced.
2. Demands for FG inventory arrive randomly, one at a time.
3. Demands that are not satisfied immediately from FG inventory are backordered and are called *Backordered Demands* or BD.
4. There is no setup cost or setup time for placing an order to the manufacturing facility.
5. There is no limit on the number of orders that can be placed per unit time. FG inventory levels are followed continuously, and replenishments of FG inventory may be ordered from the manufacturing facility at any time.

Under the above assumptions there is no incentive to replenish FG inventory by anything other than a continuous review, one-for-one replenishment policy. For such a system, Rubio and Wein (1996) proposed and analyzed a classical base stock replenishment policy, according to which, whenever the inventory position of the output store (i.e., WIP + FG − BD) falls below a specified level, called base stock, a new part is released into the manufacturing facility. A replenishment policy that has attracted considerable attention is the make-to-stock analog to a CONWIP policy proposed by Spearman *et al.* (1990). According to this policy, whenever the inventory in the system (i.e., WIP + FG) falls below a specified level, a new part is released into the manufacturing facility (e.g., see Veach and Wein, 1994). Tardif and Maaseidvaag (2001) proposed and studied an adaptive make-to-stock CONWIP policy, according to which, the level of inventory in the system that triggers the release of a new part into the manufacturing facility is allowed to change depending on the inventory and backorder levels.

Motivated by the base stock and the make-to-stock CONWIP policies, both of which depend on a single parameter, we focus on a slightly different two-parameter replenishment policy, which we call "base stock with WIP cap" policy. According to this policy, whenever the inventory position of the output store (WIP + FG − BD) falls below a specified level, called *base stock*, an order for the release of a new part in the system is issued. If the WIP inventory is below a different specified level, called *WIP cap*, the order goes through and a new part is released into the manufacturing facility; otherwise, the order is put on hold until the WIP inventory drops below the WIP cap.

More specifically, in a single-stage production–inventory operating under a base stock with WIP cap policy, a base stock, S, and a WIP cap, K, are established for the output store and the manufacturing facility, respectively. The system starts with S parts in FG inventory and no backorders or outstanding orders for FG inventory. Whenever a demand arrives, it is either satisfied from FG inventory, or is backordered, if there is no FG inventory on hand. At the same time, a replenishment order is placed to the manufacturing facility. After a replenishment lead time, during which a stock-out may occur, the replenishment order is received at the output store. Until it is received, however, it is considered to be an *outstanding order*. As in the classical base stock policy, the inventory position of the output store (i.e., outstanding orders + FG − BD) remains always equal to S. The replenishment order is released immediately into the production floor if there are fewer than K outstanding orders ahead of it in the form of WIP. Otherwise, it is put on hold until the number of outstanding orders ahead of it drops below K. This way, the WIP inventory always remains at or below K. At any time, the number of outstanding orders is equal to WIP plus the number of *orders on hold*. Clearly, at any time, the number of outstanding orders depends only on K and not on S.

The base stock with WIP cap policy described above, which will henceforth be called "(K, S) policy" for short, is not new. In fact it is equivalent to the *generalized kanban* policy proposed by Buzacott (1989) and Zipkin (1989), where the WIP capping is achieved by requiring that every part entering the manufacturing facility be granted one of K kanbans or production authorization cards.

In the past, several methods have been employed to analyze and compare make-to-stock production–inventory systems operating under various control policies. Federgruen and Zipkin (1986a,b) studied the structure of the optimal policy for a single-stage production–inventory system. Spearman (1992) used stochastic ordering arguments to compare customer service in pull production systems. Veach and Wein (1994) employed dynamic programming to compute the optimal control policy for a make-to-stock production–inventory system consisting of two stations in tandem and compared it to several parameterized pull control policies. One of their results was that base stock policies are never optimal for such a system. Karaesmen and Dallery (2000) followed the same approach to analyze more complicated parameterized policies than those considered in Veach and Wein (1994), including the (K, S) policy. Bonvik *et al.* (1997) used simulation and compared classical parameterized policies with a hybrid kanban-CONWIP control policy in a four-machine tandem production line. Rubio and Wein (1996) followed on a long line of queuing theory-based research to analyze production–inventory systems that operate under a classical base stock policy. Frein *et al.* (1995) modeled production–inventory systems that operate under a (K, S) policy as open queueing networks with restricted capacity and employed approximate analysis to evaluate their performance. Duri *et al.* (2000) compared base stock, kanban and generalized kanban policies for production–inventory systems consisting of one, two, three and four stages, using approximation techniques. Liberopoulos and Koukoumialos (2001) used simulation to evaluate the performance of single-stage and two-stage production–inventory systems operating under production–inventory policies that are modified so as to benefit from advance demand information. Zipkin (2000, Sec. 8.8.2) compared base stock, kanban and generalized kanban policies applied to a two-stage production–inventory system, where each stage consists of a single server. Finally, Kapuscinski and Tayur (1999) reported on recent advances in discrete time, single part type, capacitated systems.

In all of the above works, the main emphasis was on developing methods for evaluating and comparing the performance of different pull control policies. In some of these works it was demonstrated numerically that the (K, S) policy outperforms the classical base stock policy and the make-to-stock CONWIP policy. This is not surprising since the classical base stock policy and the make-to-stock CONWIP policy are special cases of the (K, S) policy. What remain difficult to compute are the optimal parameters of the (K, S) policy because no exact analytical solution for the performance evaluation of the (K, S) policy exists.

Clearly, there is a trade-off between the WIP cap and the optimal base stock. Namely, the lower the WIP cap, the higher the loss in effective production capacity and therefore the higher the optimal base stock to compensate for this loss, if the goal is to achieve a given minimum customer service level (percentage of demands filled from stock) with minimum inventory holding costs. But what are the characteristics of the overall optimal WIP cap and base stock values, and how do these optimal values relate to the optimal parameters of the classical base stock and make-to-stock CONWIP policies? These are the questions that we set out to answer in this paper.

The rest of this paper is organized as follows. In Section 2 we describe the (K,S) policy, and we present some special cases of it as well as properties of its dynamics and steady-state measures. In Section 3 we present properties of the optimal parameters of the (K,S) policy and in particular how they relate to the optimal parameters of the classical base stock policy. We also state an important conjecture. The validity of this conjecture is further investigated in Section 4. In Section 5 we relate the optimal parameters of a system operating under a (K,S) policy to the optimal parameter of the same system operating under a make-to-stock CONWIP policy. Finally, we draw conclusions in Section 6. The proofs of the propositions stated in Sections 2–4 are given in the Appendix.

2. Single-stage production–inventory systems operating under a (K,S) policy

To describe a single-stage make-to-stock production–inventory system operating under a (K,S) policy, we use the following notation:

$I(t)$ = FG inventory at time t;
$D(t)$ = backordered demands or BD at time t;
$M(t)$ = outstanding orders in the form of WIP at time t;
$O(t)$ = outstanding orders on hold at time t;
$N(t)$ = outstanding orders at time t;
$A(t)$ = WIP-slack (or free kanbans) at time t.

It is easy to show that the following relations hold for $t \geq 0$ (Liberopoulos and Dallery, 2000):

$$M(t) + O(t) = N(t), \quad (1)$$

$$M(t) + A(t) = K, \quad (2)$$

$$A(t) \times O(t) = 0, \quad (3)$$

$$N(t) + I(t) - D(t) = S, \quad (4)$$

$$I(t) \times D(t) = 0. \quad (5)$$

The following special cases of a single-stage (K,S) policy are worth noting (Liberopoulos and Dallery, 1995, 2000):

1. When $K = \infty$, no order is ever put on hold; therefore $O(t) = 0$ and $N(t) = M(t)$. The resulting policy is equivalent to a classical single-stage *base stock* policy.
2. When $K = S$, an order is released in the manufacturing facility only if there is at least one part in FG inventory at the time of a demand arrival. This means that an order is released in the manufacturing facility only when a part from FG inventory is consumed. The resulting policy is equivalent to a classical single-stage *kanban* policy, which is also equivalent to a make-to-stock *CONWIP* policy (Spearman *et al.* 1990).
3. When $K < S$, an order is released in the manufacturing facility if there are more than $S - K$ parts in FG inventory at the time of a demand arrival. The resulting policy is equivalent to a single-stage *reserve-stock kanban* (Buzacott, 1989) or *local control* policy (Buzacott and Shanthikumar, 1993).
4. When $K > S$, an order is released in the manufacturing facility if there are fewer than $K - S$ backorders at the time of a demand arrival. The resulting policy is equivalent to a single-stage *backordered kanban* policy (Buzacott, 1989), which is also equivalent to a single-stage *extended kanban* policy (Dallery and Liberopoulos, 2000).

It is easy to see that in a single-stage production–inventory system operating under a (K,S) policy, $N(t)$, as well as its components, $M(t)$ and $O(t)$, only depend on K and not on S. This fact and Equations (1) – (5) imply that if the sample path of $N(t)$ for a given K, denoted $N_K(t)$, is known, then the sample paths of $M(t)$ and $O(t)$ for the same K, denoted $M_K(t)$ and $O_K(t)$, respectively, and the sample paths of $I(t)$ and $D(t)$ for the same K and any finite S, denoted $I_{K,S}(t)$ and $D_{K,S}(t)$, respectively, can be easily obtained by setting:

$$M_K(t) = \min[K, N_K(t)], \quad (6)$$

$$O_K(t) = \max[0, N_K(t) - K], \quad (7)$$

$$I_{K,S}(t) = \max[0, S - N_K(t)], \quad (8)$$

$$D_{K,S}(t) = \max[0, N_K(t) - S]. \quad (9)$$

Equations (6)–(9) imply that in a single-stage production–inventory system operating under a (K,S) policy, if the steady-state pdf of the process $\{N_K(t), t \geq 0\}$, $P(N_K = i)$, $i = 0, 1, \ldots$, exists and is known for a given value of K, then the steady-state pdfs of the processes $\{M_K(t), t \geq 0\}$, $\{I_{K,S}(t), t \geq 0\}$ and $\{D_{K,S}(t), t \geq 0\}$, for the same value of K and any bounded value of S, can be easily obtained by setting:

$$P(M_K = i) = P(N_K = i), \quad i = 0, 1, \ldots, K-1, \quad (10)$$

$$P(M_K = K) = P(N_K \geq K), \quad (11)$$

$$P(I_{K,S} = 0) = P(N_K \geq S), \quad (12)$$

$$P(I_{K,S} = i) = P(N_K = S - i), \quad i = 1, 2, \ldots, S, \quad (13)$$

$$P(D_{K,S} = 0) = P(N_K \leq S), \quad (14)$$

$$P(D_{K,S} = i) = P(N_K = S + i), \quad i = 1, 2, \ldots. \quad (15)$$

The significance of equalities (10) – (15) is that the computational burden of optimizing K and S rests on optimizing K only, as was first noted in Frein *et al.* (1995) and subsequently in Liberopoulos and Dallery (1995).

The *production capacity* of a single-stage production–inventory system operating under a (K,S) policy is defined as the throughput of the manufacturing facility when it is operated under a release policy that keeps the WIP in the manufacturing facility constant and equal to K (Frein *et al.* 1995). The production capacity, therefore depends only on K and not on S. Let TH_K denote the production capacity when the WIP cap is equal to K (Frein *et al.* 1995).

In the rest of this paper we make the following assumptions.

1. $N_{K+1}(t) \leq N_K(t), \text{ for } t \geq 0 \text{ and } K \geq 0;$ (16)
2. TH_K is an increasing, concave function of K, such that $\mathrm{TH}_0 = 0$ and $\mathrm{TH}_\infty < \infty$;
3. $\mathrm{TH}_K > \lambda$, for $K \geq K_{\min}$, where $K_{\min} < \infty$ and λ is the average demand rate.

The above assumptions are not very restrictive and hold under fairly general conditions (e.g., when the manufacturing facility exhibits "max-plus" behavior). The first assumption states that the number of outstanding orders at time t is a non-increasing function of K. The second assumption implies that the manufacturing facility has a finite production capacity. Finally, the third assumption guarantees that there exists a finite minimum value of K, $K_{\min}$, such that if $K \geq K_{\min}$, the system is stable, i.e., it has enough production capacity to meet the demands in the long-run. For example, if the manufacturing facility contains m machines in series having iid exponentially distributed processing times with identical processing rate μ, the production capacity of the system is given by $\mu/[1 + (m-1)/K]$, in which case $\mathrm{TH}_\infty = \mu$ (Frein *et al.* 1995).

The following proposition, which follows from (10)–(16), demonstrates the effect of K and S on the long-run average values of N_K, O_K, M_K, $I_{K,S}$ and $D_{K,S}$.

Proposition 1. *In a single-stage production–inventory system operating under a (K,S) policy, $E(N_K)$, $E(O_K)$ and $E(D_{K,S})$ are decreasing in K, whereas $E(M_K)$ and $E(I_{K,S})$ are increasing in K. Moreover, $E(D_{K,S})$ is decreasing in S, whereas $E(I_{K,S})$ is increasing in S.*

The proofs of Proposition 1 and all propositions that follow are in the Appendix. The insight behind Proposition 1 is that although the *total* replenishment time of an outstanding order and the *fraction* of the replenishment time that the order spends being on hold are decreasing in K, the *remaining fraction* of the replenishment time that the outstanding order spends in the form of WIP is increasing in K due to the increased congestion in the system at higher values of K. Proposition 1 is validated numerically in Frein *et al.* (1995).

In the following section we present several properties of the optimal values of K and S when the objective is to minimize the long-run average cost of holding inventory for a given minimum customer service level.

3. Properties of the optimal values of K and S in single-stage production–inventory systems operating under a (K, S) policy

In this section we show some important properties of the optimal WIP cap and base stock values of single-stage production–inventory systems operating under a (K,S) policy. More specifically, we show that there is a trade-off between the WIP cap and the optimal base stock and that this trade-off holds for up to a certain finite critical WIP cap value (Proposition 2). Then, we show that the optimal WIP cap and base stock lie in the region where this trade-off holds (Proposition 3). This means that the optimal WIP cap is less than or equal to the critical WIP cap. More interestingly, however, we conjecture that the optimal WIP cap is in fact exactly equal to the critical WIP cap (Conjecture 1). This implies that the optimal base stock is exactly equal to the optimal base stock of the same system operating under a classical base stock policy. But let us take things one at a time.

We consider a typical optimization problem where the objective is to find the values of K and S that minimize the long-run average cost of holding inventory, for a given customer service level (fraction of demands filled from stock) in a single-stage production–inventory system operating under a (K,S) policy. We choose this formulation because it has particular appeal to practitioners. The service level is defined as the probability that upon arrival, a demand will find non-zero FG inventory. Under Poisson demand this probability is equal to $P(I_{K,S} > 0)$ (by the PASTA property), which by (12) is equal to $P(N_K < S)$. More specifically, the optimization problem is to find the values of K and S which minimize the long-run average cost of holding inventory,

$$C(K,S) = h[E(M_K) + E(I_{K,S})], \tag{17}$$
$$\text{subject to } P(N_K < S) \geq P_{\mathrm{F}},$$

where

h = cost of holding one part in WIP or FG inventory per unit time;
P_{F} = customer *service level* or *fill rate* (fraction of demands filled from FG inventory).

Notice that the cost function $C(K,S)$ given by (17) is increasing in S since, by Proposition 1, $E(I_{K,S})$ is increasing in S. This means that the optimal value of S for a given K, S_K^*, is the smallest integer S that satisfies

$$P(N_K < S) \geq P_{\mathrm{F}}. \tag{18}$$

S_K^* is independent of h and may take any integer value in the interval $[1, \infty]$, depending on P_{F}. More specifically, $S_K^* = 1$, if $P_{\mathrm{F}} \leq P(N_K = 0)$, whereas $S_K^* \to \infty$ as $P_{\mathrm{F}} \to 1$.

Inequality (16) implies that

$$P(N_{K+1} < S) > P(N_K < S), \text{ for } K \geq K_{\min} \text{ and } S > 0, \tag{19}$$

which together with (18) further imply that

$$S_{K+1}^* \leq S_K^*, \text{ for } K \geq K_{\min}. \tag{20}$$

The insight behind inequality (20) is that in a single-stage production–inventory system operating under a (K,S) policy, S_K^* is non-increasing in K, i.e., there is a trade-off between the WIP cap and the optimal base stock. Proposition 2 that follows states that this trade-off holds for up to a certain finite critical WIP cap value. The optimal base stock corresponding to the critical WIP cap is equal to the optimal base stock of the same system operating under a classical base stock policy with no WIP cap.

Proposition 2. *For a production–inventory system operating under a (K, S) policy, there exists a finite WIP cap, called critical WIP cap and denoted K_c $(K_{min} \leq K_c < \infty)$, such that $S^*_K = S^*_\infty$, for $K \geq K_c$, where S^*_∞ $(S^*_\infty \geq 1)$ is called the minimum optimal base stock and is equal to the optimal base stock for the same system operating under a classical base stock policy (i.e., a (K, S) policy with $K = \infty$).*

The insight behind Proposition 2 is that increasing the WIP cap beyond a finite critical value, increases the congestion in the manufacturing facility but does not increase the average FG inventory enough to warrant a further decrease in the optimal base stock.

Proposition 3 that follows states that the optimal WIP cap is *less than or equal* to the critical WIP cap and therefore the optimal base stock is *greater than or equal* to the minimum optimal base stock, i.e., the optimal base stock for the same system operating under a classical base stock policy.

Proposition 3. *In a single-stage production–inventory system operating under a (K, S) policy, the optimal WIP cap, K^*, and the optimal base stock, S^*, satisfy $K_{min} \leq K^* \leq K_c$ and $S^* \geq S^*_\infty \geq 1$, respectively.*

The insight behind Proposition 3 is that in the region where there is no trade-off between WIP cap and optimal base stock, it is costly to increase the WIP cap. This is because increasing the WIP cap only increases the congestion in the manufacturing facility without causing a decrease in the optimal base stock; therefore, the optimal WIP cap and base stock must be in the region where there is a trade-off between WIP cap and optimal base stock.

Moreover, Proposition 3 implies that, although the optimal base stock of a classical base stock policy is a lower bound for the optimal base stock of the (K, S) policy, the classical base stock policy itself (i.e., a (K, S) policy with $K = \infty$) is never optimal, since the optimal WIP cap of the (K, S) policy is at or below K_c, which is finite. This was also shown in Veach and Wein (1994) for the particular case of a system consisting of two stations in tandem.

Proposition 3 is quite helpful because it provides an upper bound on the optimal WIP cap. The question that remains unanswered, however, is how much smaller is the optimal WIP cap K^* than K_c and therefore how much larger is the optimal base stock S^*, than S^*_∞? Numerical experimentation reported by Duri *et al.* (2000), Karaesmen and Dallery (2000), Zipkin (2000 Sec. 8.8.2) and Liberopoulos and Koukoumialos (2001) indicates that the optimal WIP cap is in fact *exactly equal* to the critical WIP cap and therefore the optimal base stock is *exactly equal* to the minimum optimal base stock. This result is stated in the following conjecture.

Conjecture 1. *In a single-stage production–inventory system operating under a (K, S) policy, the optimal WIP cap, K^*, and the optimal base stock, S^*, satisfy $K^* = K_c$ and $S^* = S^*_\infty$, respectively.*

The insight behind Conjecture 1 that in the region where there is trade-off between WIP cap and optimal base stock, it is profitable to increase the WIP cap in exchange for a decrease in the optimal base stock; therefore, the WIP cap should be increased as much as possible until it reaches the border between the region where there is a trade-off between WIP cap and optimal base stock and the region where there is no such trade-off. Increasing the WIP cap any further would be costly by Proposition 3.

Conjecture 1 brings to light the essence of the role of the base stock. Namely, in a production–inventory system that operates in a make-to-stock mode, the base stock of FG inventory represents finished parts that must be "blindly" produced ahead of time, i.e., before any demands have arrived to the system, in order to meet a required minimum customer service level. "Blindly" producing parts is a necessary but very costly thing to do to meet the required customer service level under uncertain demand. For this reason the optimal base stock should be as low as possible as long as the customer service level constraint is not violated. The minimum optimal base stock is attained when a part is released in the system immediately after the arrival of a customer demand to the system, i.e., when the system is operating under a classical base stock policy. Moreover, from Proposition 2 and Proposition 3, the same minimum optimal base stock is attained at a lower cost when the system is operating under a (K, S) policy and the WIP cap is equal to the critical WIP cap. Therefore, the optimal WIP cap should be equal to the critical WIP cap and the optimal base stock should be equal to the minimum optimal base stock. This means that, *although the classical base stock policy is not optimal*, *the optimal base stock of the classical base stock policy is optimal.*

The numerical results reported in Liberopoulos and Koukoumialos (2001) further suggest that the long-run average cost increases more dramatically as K decreases away from K_c than it does as K increases away from K_c. This means that it is more costly to underestimate the WIP cap relatively to its conjectured optimal value K_c than to overestimate it. Of course, as $K \to \infty$, the long-run average cost approaches $C(\infty, S^*_\infty)$, i.e., the cost of the optimal classical base stock policy.

The difficulty in proving Conjecture 1 comes from the fact that no analytical expression for the steady-state pdf of N_K exists, except for a very simple system in which the manufacturing facility consists of a single workstation with one or more parallel machines. For such a system it is trivial to show that the optimal WIP cap is equal to the number of machines in the workstation. Despite this difficulty, in the following section we will investigate the validity of Conjecture 1 for any system for which the general assumptions stated in Section 2 hold.

4. Investigation of Conjecture 1

To investigate Conjecture 1, we will examine the conditions under which the inequality

$$C(K_c, S^*_{K_c}) = C(K_c, S^*_\infty) \le C(K_c - 1, S^*_\infty + 1) = C(K_c - 1, S^*_{K_c-1}), \tag{21}$$

holds, where it is assumed that $S^*_{K_c-1} = S^*_\infty + 1$. Proving inequality (21) would imply that K_c and S^*_∞ are the optimal parameters since, by Proposition 3, $K^* \le K_c$ and $S^* \ge S^*_\infty$. To do this, we need to derive expressions for the cost differences $C(K, S+1) - C(K, S)$ and $C(K, S) - C(K-1, S)$, for any $K > K_{\min}$ and $S > 0$.

After substituting $P(M_K = i)$, $P(I_{K,S} = i)$ and $P(D_{K,S} = i)$ from (10) – (15), rearranging terms and performing some algebraic manipulations, $C(K,S)$ in (17) can be written as

$$\begin{aligned} C(K,S) &= h\left[\sum_{i=1}^{K} iP(M_K = i) + \sum_{i=1}^{S} iP(I_{K,S} = i)\right], \\ &= h\left[\sum_{i=1}^{K-1} iP(N_K = i) + KP(N_K \ge K)\right. \\ &\quad \left. + \sum_{i=0}^{S-1} (S-i)P(N_K = i)\right], \\ &= h\left[K + \sum_{i=0}^{S-1} P(N_K \le i) - \sum_{i=0}^{K-1} P(N_K \le i)\right]. \end{aligned} \tag{22}$$

Equation (22) implies that for any $K > K_{\min}$ and $S > 0$ the following holds:

$$C(K, S+1) - C(K,S) = hP(N_K \le S), \tag{23}$$

$$C(K,S) - C(K-1,S) = \begin{cases} h\left[P(N_{K-1} \ge K) + \sum_{i=K}^{S-1} P(N_K \le i) - P(N_{K-1} \le i)\right], & K < S, \\ h[P(N_{K-1} \ge K)], & K = S, \\ h\left[P(N_{K-1} \ge K) - \sum_{i=S}^{K-1} P(N_K \le i) - P(N_{K-1} \le i)\right], & K > S. \end{cases} \tag{24}$$

Equations (23) and (24) imply that for any $K > K_{\min}$ and $S > 0$ the following holds:

$$\begin{aligned} C(K-1, S+1) - C(K,S) &= C(K-1,S+1) - C(K-1,S) + C(K-1,S) - C(K,S), \\ &= \begin{cases} h\left[P(N_{K-1} \le S) - P(N_{K-1} \ge K) - \sum_{i=K}^{S-1} P(N_K \le i) - P(N_{K-1} \le i)\right], & K < S, \\ h[P(N_{K-1} \le S) - P(N_{K-1} \ge K)], & K = S, \\ h\left[P(N_{K-1} \le S) - P(N_{K-1} \ge K) + \sum_{i=S}^{K-1} P(N_K \le i) - P(N_{K-1} \le i)\right], & K > S, \end{cases} \end{aligned} \tag{25}$$

which, for $K = K_c$ and $S = S^*_\infty$, becomes (as seen in Equation (26)).

In order for inequality (21) to hold, we need to prove that the right-hand side of Equation (26) is non-negative. Unfortunately, as we already mentioned above, to do this we need to know the steady-state pdf of N_K, a formidable

$$C(K_c - 1, S^*_\infty + 1) - C(K_c, S^*_\infty) = C(K_c - 1, S^*_\infty + 1) - C(K_c - 1, S^*_\infty) + C(K_c - 1, S^*_\infty) - C(K_c, S^*_\infty)$$
$$= \begin{cases} h\left[P(N_{K_c-1} \leq S^*_\infty) - P(N_{K_c-1} \geq K_c - \sum_{i=K_c}^{S^*_\infty - 1} P(N_{K_c} \leq i) - P(N_{K_c-1} \leq i\right], & K_c < S^*_\infty, \\ h\left[P(N_{K_c-1} \leq S^*_\infty) - P(N_{K_c-1} \geq K_c)\right], & K_c = S^*_\infty, \\ h\left[P(N_{K_c-1} \leq S^*_\infty) - P(N_{K_c-1} \geq K_c) + \sum_{i=S^*_\infty}^{K_c - 1} P(N_{K_c} \leq i) - P(N_{K_c-1} \leq i\right], & K_c > S^*_\infty. \end{cases} \tag{26}$$

task. However, there is a fairly broad class of systems for which we are able to prove that the right-hand side of Equation (26) is non-negative without explicitly knowing the steady-state pdf of N_K. This is given by the following result.

Proposition 4. *In a single-stage production–inventory system operating under a* (K,S) *policy, if* $K_c > S^*_\infty$ *and* $P_F \geq 0.5$, *then* $C(K_c, S^*_\infty) \leq C(K_c - 1, S^*_\infty + 1)$.

The insight behind Proposition 4 is that in systems for which the optimal (K,S) policy is a backordered kanban or extended kanban policy (i.e., a (K,S) policy with $K > S$) and the customer service level is at least 50%, then inequality (21) holds, and this implies Conjecture 1. Although Proposition 4 is somewhat restrictive, it is still quite powerful because it holds for any system for which the general assumptions stated in Section 2 hold.

In this following and final section we will relate the optimal parameters of a (K,S) policy to the optimal parameter of a make-to-stock CONWIP policy.

5. A property of the optimal CONWIP level in single-stage production–inventory systems operating under a make-to-stock CONWIP policy

In Section 1 it was mentioned that in a production–inventory system operating under a make-to-stock CONWIP policy, whenever the total inventory in the system (WIP + FG) falls below a certain level, say G, a new part is immediately released into the manufacturing facility. This means that the total inventory in the system always remains constant and equal to G. In Section 2, it was further mentioned that a single-stage production–inventory system operating under a (K,S) policy with $K = S$ is equivalent to the same system operating under a make-to-stock CONWIP policy with constant (WIP + FG) level G, i.e., $K = S = G$. The following result relates the optimal constant (WIP + FG) level of the make-to-stock CONWIP policy, G^*, to the optimal parameters of the (K,S) policy.

Proposition 5. *Let* K^* *and* S^* *be the optimal WIP cap and base stock, respectively, in a single-stage production–inventory system operating under a* (K,S) *policy. Let* G^* *be the optimal constant (WIP + FG) level in the same system operating under a make-to-stock CONWIP policy (i.e., a* (K,S) *policy with* $K = S = G^*$*). We have the following results. If* $K^* > S^*$, *then* $S^* < G^* \leq K^*$. *If* $K^* < S^*$, *then* $K^* < G^* \leq S^*$. *Finally, if* $K^* = K_c$ *(and therefore* $S^* = S^*_\infty$*) and* $K_c < S^*_\infty$, *then* $G^* = S^*_\infty$.

The insight behind Proposition 5 is that the optimal constant (WIP + FG) level in a single-stage production–inventory system operating under a make-to-stock CONWIP policy is between the optimal WIP cap and the optimal base stock of the same system operating under a (K,S) policy. Moreover, if the optimal (K,S) policy for the same system is a reserve-stock or a classical kanban policy (i.e., a (K,S) policy with $K^* \leq S^*$) and Conjecture 1 holds, i.e., $K^* = K_c$ and $S^* = S^*_\infty$, then the optimal constant (WIP + FG) level of the make-to-stock CONWIP policy is equal to the optimal base stock of the classical base stock policy.

6. Conclusions

In this paper, first we showed that in a single-stage production–inventory system operating under a (K,S) policy, the optimal base stock that minimizes the long-run, average, inventory carrying cost for a given minimum customer service level, is a non-increasing function of the WIP cap that reaches a minimum value, called minimum optimal base stock, at a finite WIP cap value, called critical WIP cap. This means that there is a trade-off between the WIP cap and the optimal base stock and that this trade-off holds for up to a certain finite critical WIP cap value. Then, we showed that the optimal WIP cap is less than or equal to the critical WIP cap and therefore the optimal base stock is greater than or equal to the minimum optimal base stock. This means that the optimal WIP cap and base stock are in the region where there is a trade-off between WIP cap and optimal base stock.

More interestingly, we conjectured that the optimal WIP cap is in fact exactly equal to the critical WIP cap and therefore the optimal base stock is exactly equal to the minimum optimal base stock (Conjecture 1). This minimal optimal base stock is the optimal base stock of the same system operating under a classical base stock policy. Finally, we related the optimal parameters of a system operating under a base stock with WIP cap policy to the optimal parameter of the same system operating under a make-to-stock CONWIP policy.

The insight behind Conjecture 1 is that in the region where there is trade-off between WIP cap and optimal base stock, it is profitable to increase the WIP cap in exchange for a decrease in the optimal base stock; therefore, the WIP cap should be increased as much as possible until it reaches the border between the region where there is a trade-off between WIP cap and optimal base stock and the region where there is no such trade-off. Increasing the WIP cap any further would be costly.

We showed why this conjecture, which is supported by numerical results in Duri *et al.* (2000), Karaesmen and Dallery (2000), Zipkin (2000, Sec. 8.8.2) and Liberopoulos and Koukoumialos (2001) may hold for a broad class of systems in which the optimal (K, S) policy is a backordered kanban or extended kanban policy (i.e., a (K, S) policy with $K > S$) and the minimum customer service level is at least 50%. An issue that deserves further investigation is whether this conjecture also holds for production–inventory systems with more than one stage. Numerical results in Liberopoulos and Koukoumialos (2001) suggest that it does.

References

Bonvik, A.M., Couch, C.E. and Gershwin, S.B. (1997) A comparison of production-line control mechanisms. *International Journal of Production Research*, **35**(3), 789–804.

Buzacott, J.A. (1989) Queueing models of kanban and MRP controlled production systems. *Engineering Costs and Production Economics*, **17**, 3–20.

Buzacott, J.A. and Shanthikumar, J.G. (1993) *Stochastic Models of Manufacturing Systems*, Prentice-Hall, Englewood Cliffs, NJ.

Dallery, Y. and Liberopoulos, G. (2000) Extended kanban control system: combining kanban and base stock. *IIE Transactions*, **32**(4), 369–386.

Duri, C., Frein, Y. and Di Mascolo, M. (2000) Comparison among three pull control policies: kanban, base stock and generalized kanban. *Annals of Operations Research*, **93**, 41–69.

Ferergruen, A. and Zipkin, P. (1986a) An inventory model with limited production capacity and uncertain demands I: the average-cost criterion. *Mathematics of Operations Research*, **11**, 193–207.

Ferergruen, A. and Zipkin, P. (1986b) An inventory model with limited production capacity and uncertain demands II: the discounted-cost criterion. *Mathematics of Operations Research*, **11**, 208–215.

Frein, Y., Di Mascolo, M. and Dallery, Y. (1995) On the design of generalized kanban control systems. *International Journal of Operations and Production Management*, **15**(9), 158–184.

Kapuscinski, R. and Tayur, S. (1999) Optimal policies and simulation based optimization for capacitated production inventory systems, in *Quantitative Models for Supply Chain Management*, Tayur, S., Ganeshan, R. and Magazine, M. (eds.), Kluwer, Boston, MA, pp. 7–40.

Karaesmen, F. and Dallery, Y. (2000) A performance comparison of pull control mechanisms for multi-stage manufacturing systems. *International Journal of Production Economics*, **63**(1), 59–71.

Liberopoulos, G. and Dallery, Y. (1995) On the optimization of a single-stage generalized kanban control system in manufacturing, in *Proceedings of the 1995 INRIA/IEEE Symposium on Emerging Technologies and Factory Automation*, Paris, France, October 10–13. IEEE, pp. 437–444.

Liberopoulos, G. and Dallery, Y. (2000) A unified framework for pull control mechanisms in multistage manufacturing systems, *Annals of Operations Research*, **93**, 325–355.

Liberopoulos, G. and Koukoumialos, S. (2001) Numerical investigation of trade-offs between base stock levels, numbers of kanbans and production lead times in production–inventory supply chains with advance demand information. Working paper, Production Management Laboratory, Department of Mechanical and Industrial Engineering, University of Thessaly, Pedion Areos, GR-38334, Volos, Greece.

Rubio, R. and Wein, L.W. (1996) Setting base stock levels using product-form queueing networks. *Management Science*, **42**(2), 259–268.

Spearman, M.L. (1992) Customer service in pull production systems. *Operations Research*, **40**(5), 948–958.

Spearman, M.L., Woodruff, D.L. and Hopp, W.J. (1990) CONWIP: a pull alternative to kanban. *International Journal of Production Research*, **28**, 879–894.

Tardif, V. and Maaseidvaag, L. (2001) An adaptive approach to controlling kanban systems. *European Journal of Operations Research*, **132**(2), 411–424.

Veach, M.H. and Wein, L.M. (1994) Optimal control of a two-station tandem production–inventory system. *Operations Research*, **42**(2), 337–350.

Zipkin, P. (1989) A kanban-like production control system: analysis of simple models. Research working paper No. 89-1, Graduate School of Business, Columbia University, New York.

Zipkin, P. (2000) *Foundations of Inventory Management*, McGraw Hill, Boston, MA.

Appendix

Proof of Proposition 1. First, consider a single-stage production–inventory system operating under a (K, S) policy with WIP cap $K = k + 1$ and base stock S; call this *system A*. Then, consider the same physical system operating under the same (K, S) policy but this time with WIP cap $K = k$ and the same base stock S, where $k \geq K_{\min}$; call this *system B*. Inequality (16) implies that $E(N_{K+1}) < E(N_K)$. Moreover, (7) and (16) imply that

$$O_{k+1}(t) \overset{(7)}{=} \max[0, N_{k+1}(t) - (k+1)] \leq \max[0, N_{k+1}(t) - k] \overset{(16)}{\leq} \max[0, N_k(t) - k] \overset{(7)}{=} O_k(t), \text{ for } t \geq 0,$$

which further implies that $E(O_{K+1}) < E(O_K)$. Similarly, (8), (9) and (16) imply that $D_{k+1,S}(t) \leq D_{k,S}(t)$ and $I_{k+1,S}(t) \geq I_{k,S}(t)$, for $t \geq 0$, which implies that $E(D_{k+1,S}) < E(D_{k,S})$ and $E(I_{k+1,S}) > E(I_{k,S})$, respectively.

It is more difficult to show that $E(M_{k+1}) > E(M_k)$ because (6) and (16) do not imply that $M_{k+1}(t) \geq M_k(t)$, for

$t \geq 0$. Instead, in this case we note that although the nth order is released into the manufacturing facility and departs from it earlier in system A than in system B, the difference in the departure times of the nth order in the two systems is smaller than the difference in the release times of the same order in the two systems because of queueing delays in the manufacturing facility. This means that on the average an order will spend more time in the manufacturing facility in system A than in system B, which implies that $E(M_{k+1}) > E(M_k)$.

Finally, from (8) and (9) it follows directly that $I_{k,S+1}(t) \geq I_{k,S}(t)$ and $D_{k,S+1}(t) \leq D_{k,S}(t)$ for $t \geq 0$, which implies that $E(D_{k,S+1}) < E(D_{k,S})$ and $E(I_{k,S+1}) > E(I_{k,S})$. ■

Proof of Proposition 2. Inequality (19) implies that $P(N_\infty < S) > P(N_K < S)$ or, equivalently, that

$$P(N_\infty < S) = P(N_K < S) + \varepsilon_{K,S}, \quad \text{for any finite } K \geq K_{\min} \text{ and } S > 0, \tag{A1}$$

for some $\varepsilon_{K,S}$, where $\varepsilon_{K,S}$ satisfies $0 < \varepsilon_{K,S} < 1$, $\lim_{S\to\infty} \varepsilon_{K,S} = 0$ and $\lim_{K\to\infty} \varepsilon_{K,S} = 0$.

From (18) it follows that the optimal base stock for $K = \infty, S^*_\infty$, satisfies $P(N_\infty < S^*_\infty) \geq P_{\mathrm{F}}$ and $P(N_\infty < S^*_\infty - 1) < P_{\mathrm{F}}$. This implies that there exist two numbers, ε^+_∞ and ε^-_∞, where $0 \leq \varepsilon^+_\infty < 1$ and $0 < \varepsilon^-_\infty < 1$, such that $P(N_\infty < S^*_\infty) - P_{\mathrm{F}} = \varepsilon^+_\infty$ and $P(N_\infty < S^*_\infty - 1) - P_{\mathrm{F}} = -\varepsilon^-_\infty$. After substituting $P(N_\infty < S^*_\infty)$ and $P(N_\infty < S^*_\infty - 1)$ from (A1) and rearranging terms, the above expressions can be rewritten as $P(N_K < S^*_\infty) - P_{\mathrm{F}} = \varepsilon^+_\infty - \varepsilon_{K,S^*_\infty}$ and $P(N_K < S^*_\infty - 1) - P_{\mathrm{F}} = -(\varepsilon^-_\infty + \varepsilon_{K,S^*_\infty - 1}) < 0$.

Since $\lim_{K\to\infty} \varepsilon_{K,S} = 0$, for $S > 0$, there must exist a finite, critical value of K, K_{c}, such that $\varepsilon_{K,S^*_\infty} \leq \varepsilon^+_\infty$, for $K \geq K_{\mathrm{c}}$ (except in the degenerate case where $\varepsilon^+_\infty = 0$). This implies that $P(N_K < S^*_\infty) - P_{\mathrm{F}} = \varepsilon^+_\infty - \varepsilon_{K,S^*_\infty} \geq 0$ and $P(N_K < S^*_\infty - 1) - P_{\mathrm{F}} = -(\varepsilon^-_\infty + \varepsilon_{K,S^*_\infty - 1}) < 0$, for $K \geq K_{\mathrm{c}}$. This, in turn, implies that S^*_∞ is the smallest integer such that $P(N_K < S^*_\infty) \geq P_{\mathrm{F}}$, for $K \geq K_{\mathrm{c}}$, which means that S^*_∞ is the optimal base stock for $K \geq K_{\mathrm{c}}$, i.e., $S^*_K = S^*_\infty$, for $K \geq K_{\mathrm{c}}$. ■

Proof of Proposition 3. Proposition 1 and Proposition 2 imply that

$$C(K, S^*_K) \overset{\text{(Prop.2)}}{=} C(K, S^*_\infty) \overset{\text{(Prop.1)}}{>} C(K_{\mathrm{c}}, S^*_\infty) \overset{\text{(Prop.2)}}{=} C(K_{\mathrm{c}}, S^*_{K_{\mathrm{c}}}), \text{ for } K > K_{\mathrm{c}}.$$

Therefore, $K^* \leq K_{\mathrm{c}}$. From (20) this further implies that $S^* = S^*_{K^*} \overset{\text{(Prop.3)}}{\geq} S^*_{K_{\mathrm{c}}} = S^*_\infty$. ■

Proof of Proposition 4. Suppose that $K_{\mathrm{c}} = S^*_\infty + \Delta$, where Δ is some positive integer (e.g., 1, 2, ...). Then, (26) implies that

$$\begin{aligned}
&C(K_{\mathrm{c}} - 1, S^*_\infty + 1) - C(K_{\mathrm{c}}, S^*_\infty), \\
&= h\left[P(N_{K_{\mathrm{c}}-1} \leq S^*_\infty) - P(N_{K_{\mathrm{c}}-1} \geq S^*_\infty + \Delta) + \sum_{i=S^*_\infty}^{S^*_\infty+\Delta-1} P(N_{K_{\mathrm{c}}} \leq i) - P(N_{K_{\mathrm{c}}-1} \leq i)\right], \\
&= h\left[P(N_{K_{\mathrm{c}}-1} < S^*_\infty + 1) + P(N_{K_{\mathrm{c}}-1} < S^*_\infty + \Delta) - 1 + \sum_{i=S^*_\infty}^{S^*_\infty+\Delta-1} P(N_{K_{\mathrm{c}}} \leq i) - P(N_{K_{\mathrm{c}}-1} \leq i)\right], \\
&= h\left[2 \times P(N_{K_{\mathrm{c}}-1} < S^*_\infty + 1) + P(S^*_\infty + 1 \leq N_{K_{\mathrm{c}}-1} < S^*_\infty + \Delta) - 1 + \sum_{i=S^*_\infty}^{S^*_\infty+\Delta-1} P(N_{K_{\mathrm{c}}} \leq i) - P(N_{K_{\mathrm{c}}-1} \leq i)\right], \\
&= h\left[2 \times P(N_{K_{\mathrm{c}}-1} < S^*_{K_{\mathrm{c}}-1}) + P(S^*_\infty + 1 \leq N_{K_{\mathrm{c}}-1} < S^*_\infty + \Delta) - 1 + \sum_{i=S^*_\infty}^{S^*_\infty+\Delta-1} P(N_{K_{\mathrm{c}}} \leq i) - P(N_{K_{\mathrm{c}}-1} \leq i)\right], \\
&\geq h\left[2 \times P_{\mathrm{F}} + P(S^*_\infty + 1 \leq N_{K_{\mathrm{c}}-1} < S^*_\infty + \Delta) - 1 + \sum_{i=S^*_\infty}^{S^*_\infty+\Delta-1} P(N_{K_{\mathrm{c}}} \leq i) - P(N_{K_{\mathrm{c}}-1} \leq i)\right], \\
&\geq h\left[(2)(0.5) - 1 + P(S^*_\infty + 1 \leq N_{K_{\mathrm{c}}-1} < S^*_\infty + \Delta) + \sum_{i=S^*_\infty}^{S^*_\infty+\Delta-1} P(N_{K_{\mathrm{c}}} \leq i) - P(N_{K_{\mathrm{c}}-1} \leq i)\right], \\
&> 0,
\end{aligned}$$

where we have used the fact that $P(N_{K_c} \leq i) - P(N_{K_c - 1} \leq i) > 0$ from (19). ■

Proof of Proposition 5. A single-stage production–inventory system operating under a make-to-stock CONWIP policy with constant (WIP + FG) level G, is equivalent to the same system operating under a (K, S) policy with $K = S = G$. From (22), it follows that the long-run average cost of holding inventory in a single-stage production–inventory system operating under a make-to-stock CONWIP policy with constant (WIP +FG) level G is given by $C(G, G) = hG$. This means that the optimization problem related to the make-to-stock CONWIP policy is to minimize

$$C(G, G) = hG$$
$$\text{subject to } P(N_G < G) \geq P_F.$$

Clearly, since the cost function $C(G, G)$ is increasing in G, the optimal constant (WIP + FG) level G, G^*, is the smallest integer that satisfies

$$P(N_G < G) \geq P_F. \tag{28}$$

From (18) and (19), the optimal parameters K^* and S^* of the (K, S) policy satisfy

$$\begin{aligned} &P(N_K < S) \geq P_F, \text{for } K \geq K^* \text{ and } S \geq S^*, \\ &P(N_K < S) < P_F, \text{ for } K_{\min} \leq K \leq K^* \text{ and } S < S^*, \\ &\quad \text{and for } K_{\min} \leq K < K^* \text{ and } S \leq S^*. \end{aligned}$$

This means that if $K^* > S^*$, then $P(N_G < G) \geq P_F$, for $G \geq K^*$, and $P(N_G < G) < P_F$, for $K_{\min} \leq G \leq S^*$, which by (18) implies that G^* is an integer that satisfies $S^* < G^* \leq K^*$.

It also means that if $K^* < S^*$, then $P(N_G < G) \geq P_F$, for $G \geq S^*$, and $P(N_G < G) < P_F$, for $G \leq K^*$, which by (18) implies that G^* is an integer that satisfies $K^* < G^* \leq S^*$.

Finally, if $K^* = K_c$ and $S^* = S^*_\infty$, (18) and (19) and Proposition 2 imply that the optimal parameters K_c and S^*_∞ of the (K, S) policy satisfy

$$\begin{aligned} &P(N_K < S) \geq P_F, \text{ for } K \geq K_c \text{ and } S \geq S^*_\infty, \\ &P(N_K < S) < P_F, \text{ for } K_{\min} \leq K \text{ and } S < S^*_\infty \\ &\quad \text{and for } K_{\min} \leq K < K_c \text{ and } S \leq S^*_\infty. \end{aligned}$$

This means that if $K_c < S^*_\infty$, then $P(N_G < G) \geq P_F$, for $G \geq S^*_\infty$, and $P(N_G < G) < P_F$, for $G < S^*_\infty$, which by (18) implies that $G^* = S^*_\infty$. ■

Biographies

George Liberopoulos received his B.S. and M.Eng. from Cornell University in 1985 and 1986, respectively, and his Ph.D. from Boston University in 1993. In 1994–96 he was a visiting scientist at Universite Paris IV. He is currently an Assistant Professor of Production Management at the University of Thessaly. His research interests are in operations management with an emphasis on the analysis and control of production and inventory systems. He serves on the Editorial Boards of *IIE Transactions on Design* and *Manufacturing* and *OR Spectrum*.

Yves Dallery received his Ph.D. and the degree of Habilitation a Diriger des Recherches from the Institut National Polytechnique de Grenoble (INPG) in 1984 and 1989, respectively. He is currently Professor of Manufacturing and Logistics at Ecole Centrale de Paris. Before that, he was Directeur de Recherche at the Centre National de la Recherche Scientifique (CNRS). In 1984–1985, he was a post-doctoral fellow at Harvard University. In 1991–1992, he was a visiting scientist at MIT and in 1992–1993 he was an Associate Professor of Manufacturing Engineering at Boston University. His research interests are in operations management, supply chain management, and stochastic models.

Contributed by the Manufacturing Systems Control Department