Stocking Decisions in a Multiple Location Periodic Review Inventory System with Order Crossover

Yahel Giat
Department of Industrial Engineering, Jerusalem College of Technology, Jerusalem, Israel
Department of Medicine, Hebrew University, Jerusalem, Israel
yahel@jct.ac.il

Abstract

Deliveries in global supply chains are often made through lengthy shipping routes that are subject to many delays such as border crossings, inspections and so forth. Consequently, orders frequently crossover, that is, their order of arrival is not the same as the order that they were issued. In this paper we consider the window fill rate of a multiple location inventory system with order crossover. Demand arrives to each location according to a Poisson arrival process and inventory in each location is reviewed periodically. The system's window fill rate is the probability that a customer arriving to the system is served within her tolerable wait. We consider two stocking optimization problems. The first is the spares allocation problem, which is defined as finding the optimal allocation that maximizes the window fill rate given a budget of spares. The second problem is the optimal budget problem, which is defined as the optimal allocation that minimizes the total budget of spares subject to a required window fill rate level. We show that when resources are scarce it is optimal to cluster spares to only few locations. In contrast, when resources are abundant, then an equitable solution is optimal.

Keywords
Spares allocation; window fill rate; customer patience; order crossover; optimization

1. Introduction

Inventory system models regularly assume that order lead times are stochastic and independent. Unless very strict assumptions are imposed on the lead times distribution, the independence of the lead times implies that orders may crossover, that is, orders may be delivered in a sequence that differs from the sequence that they were issued. The problem with order crossover is that it makes the tracking of the deliveries very difficult, and therefore, many studies analyzing these systems make an approximation assumption that orders do not crossover. This assumption results in a reasonable approximation when the lead times are not too volatile, so that the probability for crossovers is low. Unfortunately, real-life lead times distributions of international supply chains are in fact very volatile (see Disney et al. 2016) and the assumption of no crossovers results in poor estimations.

In this paper, we consider a multiple location inventory system that may be either an exchangeable-item repair system or a make-to-order (MTO) inventory system. In the former, customers arrive with a malfunctioning item and leave the system once they receive an operational item either from the location's stock or through repair. In the latter, customers wait until their order is satisfied either by a spare from the location's stock or through the arrival of a new item from an external source. Each location in the system is a warehouse that operates under a periodic review policy. Customer arrive to each location following a Poisson process with a demand for a single item and are served by order of arrival. While demand is backlogged, it is assumed that customers have a level of patience, i.e., they will tolerate a certain amount of wait that we name the tolerable wait. It is the objective of managers that customers are served within their tolerable wait and accordingly, the location's performance is measured as the location's window fill rate, the probability that a customer is served within the tolerable wait.

© IEOM Society International

In the multiple location system, we define the system's window fill rate as the average of the locations' window fill rates weighted by the arrival rate to each location. We consider two typical problems that managers face in such a system. The first is the spares allocation problem (SAP), defined as finding the allocation of spares that maximizes the window fill rate for a given budget of spares. The second is the optimal budget problem (OBP) that is defined as the optimal allocation of spares that minimizes the budget for a required window fill rate level.

For the location's window fill rate, van der Heijden and de Kok (1992) develop a formula that makes the approximation assumption that orders do not crossover. However, since that formula may result with considerable approximation error, we use the exact formulas developed in Giat and Dreyfuss (2020). The solution to the two optimization problems show that when budget is scarce it is optimal for managers to concentrate spares in few locations whereas if the budget is sufficiently large then a more equitable solution is optimal. Furthermore, we find a near linear negative tradeoff between customer patience and the number of spares needed to meet a target window fill rate. This result is helpful in designing time-dependent service contracts.

2. Literature Review

Our study is part of the rich literature of multiple location inventory systems. Inventory in each location in our model is monitored periodically. Wensing (2011) provides a summary of the main periodic review policies. Most periodic review policies can be categorized into either the "fixed order size" category or the "order-up-to" category. In the "fixed order size" policies every r periods the inventory is reviewed and if it falls below the level s then an order of q (or multiples of q) is issued. Studies of these systems include Tempelmeier and Fischer (2010) and van Donselaar and Broekmeulen (2013). With order-up-to policies, the order size q is set so that the current inventory level will be a fixed target level, S. Studies of these systems include Avinadav and Henig (2015), Disney et al. (2015) and Dreyfuss and Giat (2019). In our model, each location follows the basic order-up-to review policy (r, S). For this policy, inventory is reviewed every r periods at which point an order is made to bring the inventory level to S.

The (r, S) policy has been investigated by van der Heijden and de Kok (1992) who developed for it the waiting time distribution when customer arrivals follow a compound Poisson process using the approximating assumption that orders do not crossover. Dreyfuss and Giat (2019) show that when demand follow a Poisson process and lead times are deterministic that the window fill rate (i.e., the waiting time distribution) is initial convex and then concave in the number of spares. Chen and Zheng (1992) develop the waiting time distribution when lead times are dependent and do not cross over. In all these papers, orders are issued at the same times that the inventory is reviewed. Prak et al. (2015) considers an alternative modelling approach in which orders are not issued at the time of review.

The approximating assumption that orders do not crossover is often used by researchers to facilitate the derivation of the system's performance measures. See, for example, van der Heijden and de Kok (1992) Tempelmeier and Fischer (2010), Tempelmeier and Fischer (2019) and Johansen (2019). This assumption is justified particularly when there is a single supplier since "if the dispatch of a particular replenishment is delayed considerably, in practice it is never crossed, at most it is combined with the next replenishment order" (van der Heijden and de Kok, 1992, p. 318). Hadley and Whitin (1963, p. 203), and more recently by Tempelmeier and Fischer (2019) explain that the approximation is justified when r is large compared to the span of the lead times. Moreover, it has been argued by van der Heijden and de Kok (1992, p. 326) that "lead time variation is not very large in practice" and therefore the no-crossovers approximation is reasonable. It appears, however, that these arguments are untenable in current global supply chains. These supply networks favor supplier diversification over relying on a single supplier in order to reduce risks, create value and preempt competition (Balakrishnan and Chakravarty 2008; Mukherjee and Tsai 2013; Heese 2015; Namdar et al. 2018). With multiple suppliers, order crossovers are much more likely to happen. Furthermore, even when orders are issued to a single supplier, they often crossover due to clerical errors, customs inspections and other shipment delays (Wensing and Kuhn, 2015). For example, Disney et al. (2016) examine shipping routes between cities in China and the USA and show how shipment times exhibit a high variability. For the shipping route between Shenzhen and Colorado, in particular, they found that 40% of the orders crossed over.

Studies that examine systems with order crossover include Bischak et al. (2014) who develop a cost model in a periodic order-up-to (r, S) inventory system and Svoronos and Zipkin (1991) in a continuous one-for-one (base-stock) setting. Wensing and Kuhn (2015) derive formulas for different service measures in a periodic model similar to our model. To show how crossovers affect performance they use the effective lead times distribution for the model with no crossovers and the actual lead times for the model with crossover. Srivastav and Agrawal (2018) analyze a periodic review multi-objective inventory model with crossovers and Kouki et al. (2019) consider lost sales in a base-stock inventory system. Our study differs from these studies in that we use the window fill rate as the system's performance measure. This service measure is optimized in Dreyfuss and Giat (2018) in a multiple-echelon continuous base-stock model with two-echelons and multiple locations.

The tolerable wait in our model may be the result of contractual agreements about service times, see Caggiano et al. (2009, p.744) or the result of customer patience or agreement to wait. Although it is seldom modeled in inventory models, it is very common in studies in the service industry. Examples include Katz et al., (1991) who name it "reasonable duration", and Demoulin and Djelassi (2013) who term it "wait acceptability". Our study enhances our understanding of the importance of customer patience and its effect on inventory stocking decisions.

3. The Model

The inventory system of this model comprises N warehouses, with each warehouse operating similarly to the model in Giat and Dreyfuss (2020). Our model description begins with a description of a single warehouse and then makes the extension to a multiple location system.

3.1 Single Location

Consider an inventory warehouse that manages a single item-type. Customers' arrivals follow a Poisson process with rate λ. We assume that each customer arrives to the facility with a demand for exactly one item, implying that the demand follows a simple Poisson process, too. The warehouse serves customers according to a first come first serve (FCFS) policy. Orders are backordered, nevertheless, customers expect to be served within their tolerable wait, w, and it is the system's objective to meet their expectation. Accordingly, for a single location, the system's performance level is the *location* window fill rate ($LWFR$), i.e., the probability that a customer is served within w units of time,

Inventory management in the warehouse follows a periodic (r, S) review policy, meaning that every r units of time the warehouse issues an order to replenish its stock back to its initial inventory level, S. Therefore, S denotes the number of spares in the warehouse.

It is assumed that the order lead times are stochastic and independent. We let L denote their cumulative distribution function and assume that if $x < 0$ then $L(x) = 0$. Furthermore, we assume that the support of L is a bounded interval, and therefore there exists a $\hat{x}$ such that for all $x > \hat{x}$, $L(x) = 1$. These assumptions imply that order deliveries may crossover. This phenomenon is very common in global supply chains in which the supplier is distant from the ordering warehouse and therefore transit times are long compared to the cycle review time. In these cases, orders sent by the supplier may experience sufficiently long transit delays to crossover.

3.2 The Location Window Fill Rate

To derive the location's window fill rate, $LWFR$, we need to consider a random customer's arrival and compute the probability that the supply of available items after w units of time from her arrival is greater than the demand for items upon her arrival. The resulting formula, developed in Giat and Dreyfuss (2020) makes repeated use of Skellam random variables (Skellam, 1948), reflecting the need to find the *difference* between Poisson random variables. We refer the reader to that study and present here only a simplified version that assumes $w < r$.

Let K be such that the probability for an order to be fulfilled within Kr units of time is one. Such K exists because we assume L has a finite support. Let m be any integer between 0 and $2^K - 1$. Notice that the binary representation of m comprises K digits. Let $m^{(n)}$ denotes the n'th digit of the binary representation of m, where the units digit is counted as digit 0. Let $Sum(m)$ denote the sum of the digits of the binary representation of m. The location's window fill rate, $LWFR$, is given by the following formula

$$LWFR(S) = \frac{1}{r}\left[\int_0^{r-w} \sum_{m=0}^{2^K-1} \Pr[M=m]\Pr[Y(m) \le S-1]\,dt + \int_{r-w}^{r}\left(\sum_{m\text{ Odd}} \Pr[M=m]\Pr[Y(m-1) \le S-1] + \sum_{m\text{ Even}} \Pr[M=m]\Pr[Z(m) \le S]\right)dt\right], \tag{LWFR}$$

where M, $Y(m)$ and $Z(m)$ are random variables with the distribution

$$\Pr[M=m] = \begin{cases} \prod_{n=0}^{K-1} \hat{L}\left(t+w+nr, m^{(n)}\right) & \text{if } t < r-w \\ \prod_{n=0}^{K-1} \hat{L}\left(t+w+nr-r, m^{(n)}\right) & \text{if } t \ge r-w, \end{cases} \tag{MDist}$$

$$Y(m) \sim \text{Poisson}\big(\lambda(rSum(m)+t)\big), \tag{YDist}$$

$$Z(m) \sim \text{Skellam}\big(\lambda rSum(m), \lambda(r-t)\big), \tag{ZDist}$$

and where we use the notation

$$\hat{L}(x,j) = \begin{cases} L(x) & \text{if } j=0 \\ 1-L(x) & \text{if } j=1. \end{cases} \tag{Lhat}$$

3.3 Multiple Locations

Let us now return to the multiple location inventory system. It comprises N warehouses, each operating in a similar manner to the warehouse described earlier. Locations may be characterized by their order cycle time, demand arrivals rate, order lead times and tolerable wait. This last assumption, that customer patience is location-dependent, requires further clarification. It could be because customers are assigned to specific locations and therefore the customer base of each location may differ in its tolerable wait than the customer base of other locations. Alternatively, customers may be allowed to arrive at any location, but their expectations change according to the location to which they arrived. For example, a certain location may be considered a "premium" location in which service is expected to be rendered sooner than the standard location.

Let r_n, λ_n, L_n and w_n denote the order cycle time, demand arrivals rate, their order lead times and tolerable wait of location n, respectively. Additionally, we index the locations' window fill rates to denote the location. The *system's* window fill rate is the probability that a random customer in the system is served within her tolerable wait. This is given as the average of the locations' window fill rates weighted by their arrival rates. Let $\lambda \coloneqq \sum_{n=1}^{N} \lambda_n$ denote the total arrival to the system, let $\vec{S} \coloneqq (S_1, \dots, S_N)$ denote a spares allocation. The system's window fill rate is given by the following formula

$$SWFR(\vec{S}) \coloneqq \sum_{n=1}^{N} \frac{\lambda_n}{\lambda} LWFR_n(S_n) \tag{SWFR}$$

The formulation of the system's window fill rate emphasizes the system over the individual locations. In the next section, we demonstrate how this focus could result with certain locations performing very poorly whereas the overall system performing reasonably well.

4. Optimization

4.1 The Spares Allocation Problem

Given a budget of spares B, the system's managers' objective is to find the allocation of spares that maximizes the window fill rate. Formally, the spares allocation problem is

$$\max_{\vec{S}} SWFR(\vec{S}) \quad \text{subject to:} \quad B = \sum_{n=1}^{N} S_n \tag{SAP}$$

Consider the case that each of the locations' window fill rates is concave in the number of spares. This, in fact, happens when w_n or S_n are large. In this case, a greedy algorithm solves the spares allocation problem in computation time $O(B)$. When not all the locations are concave in S, then SAP is an NP hard problem. The algorithm that we describe is taken from Dreyfuss and Giat (2017) and has a running time of $O(B)$ and generally results in a near-optimal solution. It assumes that whenever a particular $LWPR$ is not concave, then it is initially convex and then concave in the number of spares (see Figure 1). While this assumption is not formally proven, in all our experiments we were unable to produce parametric values that are inconsistent with this assumption. The algorithm comprises the following steps:

1. For each location n, find the tangent point of $LWFR_n(S_n)$, p_n. p_n is the first integer such that the slope of the line connecting $LWFR_n(0)$ and $LWFR_n(p_n)$ is greater or equal to the slope of $LWFR_n$ at p_n. Note that if $LWFR_n(S_n)$ is concave then $p_n = 1$.
2. For each location n, construct $H_n(S_n)$, the concave covering of $LWFR_n(S_n)$, by replacing $LWFR_n(S_n)$ left of the tangent point with the line connecting $LWFR_n(0)$, and $LWFR_n(p_n)$.
3. Define $H(\vec{S}) \coloneqq \sum_{n=1}^{N} \frac{\lambda_n}{\lambda} H_n(S_n)$
4. Solve the problem
$$\max_{\vec{S}} H(\vec{S}) \quad \text{subject to:} \quad B = \sum_{n=1}^{N} S_n \tag{CCP}$$
using a greedy algorithm. We denote the solution as $\vec{S}^H$ and point out that it is the optimal solution to (CCP) since $H(\vec{S})$ is a separable sum of concave functions.
5. Algorithm output: $H(\vec{S}^H)$ and $SWFR(\vec{S}^H)$, are the upper and lower bounds of (SAP), respectively.

The shape of each H_n is linear between 0 and the location's tangent point (p_n), and concave beyond it. Therefore, the greedy algorithm in Step 4 assigns spares in the following manner: Locations are sorted according to the slope of their linear region. Spares are allocated to the first sorted location (i.e., the location with the steepest slope) until it has reached its tangent point. Next, the second steepest location receives spares until it has reached its tangent point and so forth. Note that between switching from one location to the next, it is possible that locations that have already reached their tangent point will receive additional spares. This will happen if the (concave) slope at a location that has already reached its tangent point is steeper than the linear slope of the next sorted location. At any event, however, once a location starts receiving spares, it will be the only one to receive spares until it has reached its tangent point. The difference between the optimal solution and the solution of our algorithm is the difference between $H(\vec{S}^H)$ and $SWFR(\vec{S}^H)$. By construction, for any n, if $S_n = 0$ or $S_n = p_n$ then $H_n(S_n) = LWFR_n(S_n)$. Therefore, if all the locations are allotted either zero spares or at least their tangent point then the solution is optimal. An important property of $\vec{S}^H$ is that at most one location, n^*, will be in the intermediate position with $0 \leq S_{n^*}^H \leq p_{n^*}$. This happens when the greedy algorithm has begun allotting spares to location n^* and the budget was depleted before location n^* has reached its tangent point. When this happens, the solution may be suboptimal and the *a posteriori* distance from optimality is $\frac{\lambda_{n^*}}{\lambda}\left(H_{n^*}(S_{n^*}^H) - LWFR_{n^*}(S_{n^*}^H)\right)$. The *a priori* distance from optimality is the maximal difference between H and $SWFR$, when there is at most one "intermediate" location. It is given by $\max_{n=1,\dots,N} \max_{S_n=1,\dots,p_n} \frac{\lambda_n}{\lambda}(H_n(S_n^H) - LWFR_n(S_n^H))$. Observe that when the number of locations in the system increases, the ratio λ_n/λ decreases. Therefore, the larger the inventory system in terms of the number of locations, the smaller the *a priori* and *a posteriori* distances from optimality.

4.2 The Optimal Budget Problem

In the spares allocation problem (SAP) we assume that the budget is determined exogenously. This assumption is appropriate, for example, when it is difficult to purchase additional spares due to, for example, lack of capital resources or scarcity of components. In many realistic situations, however, managers are given a target performance level (F^{req}), and they need to determine the minimal budget (and its allocation) that is needed to meet that target. This problem is known as the optimal budget problem and is given by

$$\max_{\vec{S}} B = \sum_{n=1}^{N} S_n \quad \text{subject to:} \quad SWFR(\vec{S}) \geq F^{req} \tag{OBP}$$

The solution to OBP, B, ensures that the system's window fill rate meets the required performance level. The solution is an upper bound and it is optimal only when the solution has no intermediate location. The lower bound is determined by replacing $SWFR(\vec{S})$ with $H(\vec{S}^H)$ in (OBP).

4. Numerical Illustration

4.2 Single Location

We will use the following parameter values throughout the numerical illustration: The review cycle time is $r = 14$ days, the customer arrivals rate is $\lambda = 1$ customers per day and the lead times are distributed uniformly $L \sim U(10,50)$ days. The tolerable wait values are w = 0, 5 and 10 days.

In Figure 1 we display the location window fill rate as a function of the number of spares when the tolerable wait is $w = 0$. The concave covering function is also depicted. In Figure 2 we display the location window fill rate for different customer patience levels. Notice that the figure shifts to the left as the tolerable wait increases, reflecting the fact that to maintain a fixed performance level, less spares are needed when w increases. When w is 0, 5 and 10 then the tangent points are 51, 44 and 38, respectively. These values are important for the optimization procedure in the multiple location system in the next section.

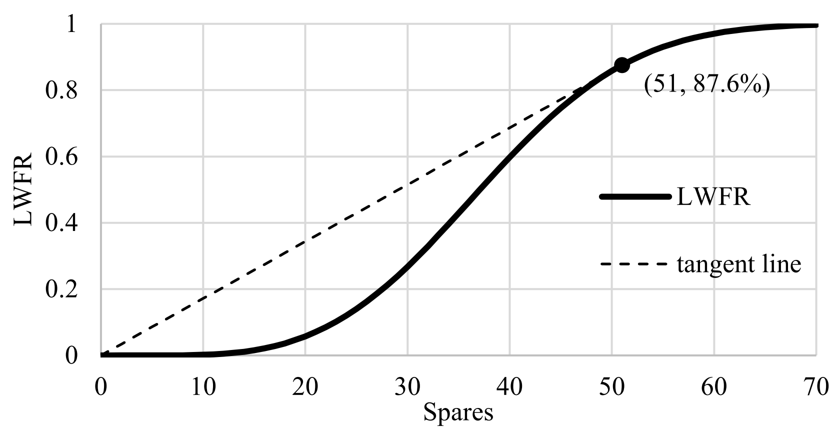

Figure 1. $LWFR(B)$ and its concave covering when $w = 0$

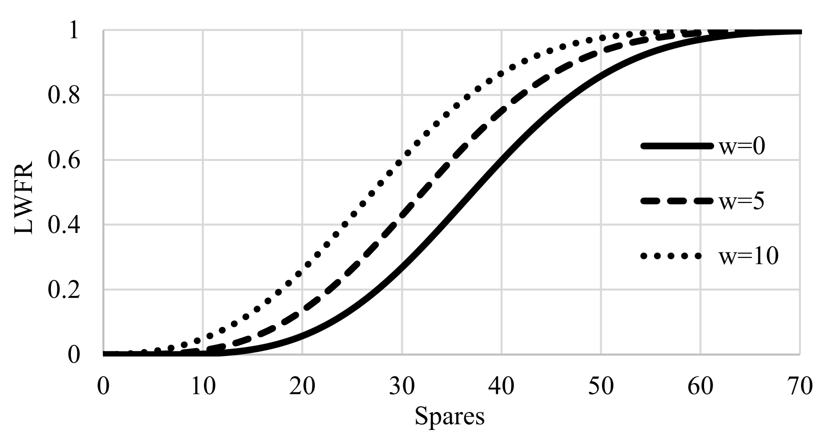

Figure 2. $LWFR(B)$ for different values of tolerable wait

© IEOM Society International

4.2 Multiple Location System

We now consider a system with $N = 10$ locations, each identical to the warehouse described in the previous section. Our choice to limit the illustration to identical locations is intentional as it highlights the asymmetry that is induced by the structure of the window fill rate. In Figure 3 we solve SAP for different budget sizes when $w = 0$, and report the upper bound and lower bound of the optimal $SWFR$ values. Recall, when $w = 0$, the tangent point of each location is 51 spares. Therefore, when the budget is $B = 510$ and higher, then each of the locations has reached its tangent point and the solution is guaranteed to be optimal. For a smaller budget, then whenever it is not a multiple of 51, then there is an intermediate location and the solution is not necessarily optimal.

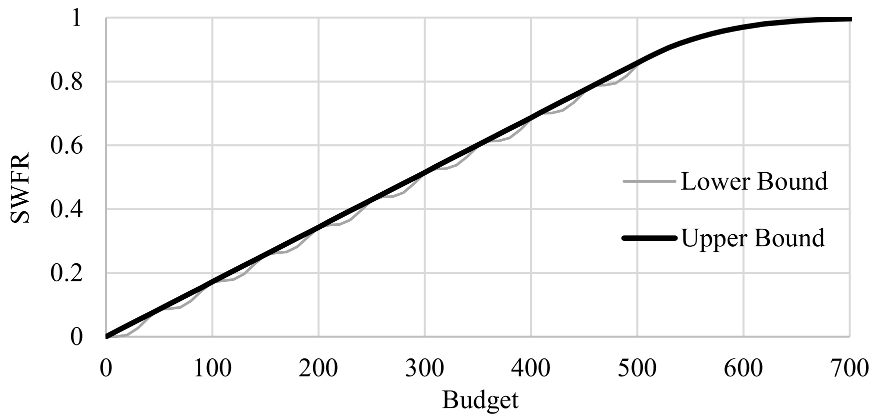

Figure 3. The SWFR as a function of the budget (S) when $w = 0$

We solve SAP for different budget sizes and different tolerable waits. The optimal allocation, system window fill rate and distance between bound are reported in Table 1. Although the locations are identical, the optimal allocation is very asymmetric for small budgets. For example, when $w = 0$, since the tangent point is 51, then when $B = 200$ then 3 locations will receive 51 spares each and the fourth location receives the remaining 47 spares. The other locations receive zero spares. Since there is an "intermediate" location (the fourth location), then the solution is not necessarily optimal. The effect of customer patience is greatest for intermediate budget values. For example, when spares are scarce such that $B = 100$ then increasing customer patience from zero to 10 will only result in a (20.5%-17.1%) 3.4% improvement in $SWFR$. In contrast, when $B = 400$ the same increase in customer patience will result in a (86.6%-68.2%) 18.4% improvement in $SWFR$. Finally, when $B = 600$ the improvement in $SWFR$ is less than 2%.

Table 1. The solution to SAP

	$w = 0$	$w = 5$	$w = 10$
B	$(\vec{S}) = (S_1, S_2, \dots, S_N)$ $SWFR(\vec{S})$, (dist. between bounds)	$(\vec{S}) = (S_1, S_2, \dots, S_N)$ SWFR$(\vec{S})$, (dist. between bounds)	$(\vec{S}) = (S_1, S_2, \dots, S_N)$ SWFR$(\vec{S})$, (dist. between bounds)
100	(51,49,0,0,0,0,0,0,0,0) 17.1% (0.02%)	(44,44,12,0,0,0,0,0,0,0) 17.1% (2.05%)	(38,38,24,0,0,0,0,0,0,0) 20.5% (1.30%)
200	(51,51,51,47,0,0,0,0,0,0) 34.2% (0.11%)	(44,44,44,44,24,0,0,0,0,0) 36.1% (2.25%)	(38,38,38,38,38,10,0,0,0,0) 41.8% (1.68%)
300	(51,51,51,51,51,45,0,0,0,0) 51.2% (0.26%)	(44,44,44,44,44,44,36,0,0,0) 56.9% (0.58%)	(38,38,38,38,38,38,38,34,0,0) 65.1% (0.13%)
400	(51,51,51,51,51,51,51,43,0,0) 68.2% (0.48%)	(44,44,44,44,44,44,44,44,44,4) 75.9% (0.76%)	(40,40,40,40,40,40,40,40,40,40) 86.6% (optimal)
500	(51,51,51,51,51,51,51,51,51,41) 85.1% (0.74%)	(50,50,50,50,50,50,50,50,50,50) 93.4% (optimal)	(50,50,50,50,50,50,50,50,50,50) 97.5% (optimal)
600	(60,60,60,60,60,60,60,60,60,60) 97.1% (optimal)	(60,60,60,60,60,60,60,60,60,60) 99.1% (optimal)	(60,60,60,60,60,60,60,60,60,60) 99.8% (optimal)

© IEOM Society International

The SAP algorithm dictates that if there are insufficiently many spares to bring all the locations to their tangent points, then spares should be concentrated into few locations. This prescribes that some locations are neglected (see Table 1), a prescription that many managers are reluctant to follow. In Table 2 we consider the case that managers insist on allocating spares equally among all the locations and compare $SWFR$ for this symmetric solution and $SWFR$ for the SAP algorithm's solution (we report the lower bound). The advantage of the concentrated solution is obvious when the budget is limited. There is only one instance in which the symmetric solution is better than the algorithm's solution, due to the suboptimality of the solution (see the distance between bounds for this case in Table 1). Finally, when there are sufficiently many spares then the algorithm's solution and the symmetric solution converge and perform identically.

Table 2. Performance under the algorithm's solution and the symmetric solution

	$w = 0$		$w = 5$		$w = 10$	
B	SAP solution	Symmetric sol.	SAP solution	Symmetric sol.	SAP solution	Symmetric sol.
100	17.1%	0.2%	17.1%	1.3%	20.5%	4.9%
200	34.2%	5.7%	36.1%	13.5%	41.8%	26.0%
300	51.2%	26.8%	56.9%	42.8%	65.1%	60.1%
400	68.2%	59.7%	75.9%	75.0%	86.6%	86.6%
500	85.1%	85.8%	93.4%	93.4%	97.5%	97.5%
600	97.1%	97.1%	99.1%	99.1%	99.8%	99.8%

When the system must meet a performance level, the tolerable wait determines the number of spares that must be purchased. This is quite common in many service contracts that dictate the time in which service must be rendered (e.g. Caggiano et al., 2009). We solved OBP for different values of required performance (F^{req}) and customer patience. Fill rate values of 80% and above are common in the inventory industry and the 50% threshold is presented for comparison. The results are reported in Table 3. Our experiments (of which only three are reported in Table 3) reveal a near linear negative relationship between the required budget and the tolerable wait. A day increase in the tolerable wait results in approximately 11.3 less spares that are needed to maintain a $F^{req} = 80\%$. This ratio slightly decreases with F^{req}. Thus, when $F^{req} = 95\%$, approximately 10.7 less spares are required for a day increase in the tolerable wait.

Table 3. The solution to OBP

	$w = 0$	$w = 5$	$w = 10$
F^{req}	B (dist. between bounds)	B (dist. between bounds)	B (dist. between bounds)
50%	296 (4)	262 (1)	238 (8)
80%	484 (18)	426 (8)	371 (3)
90%	526 (optimal)	474 (optimal)	421 (optimal)
95%	571 (optimal)	517 (optimal)	464 (optimal)

4. Conclusion

In this study we describe a multiple location inventory system with periodic review and solve two typical problems that managers must solve to reduce costs or maximize performance. We show that when there is a limited number of spares, the optimal solution is asymmetric even when the locations are identical. Thus, when resources are scarce, it is optimal to concentrate them in few locations rather than distribute them equitably. Customer patience play a significant role in the system's performance and reduces the number of spares needed to meet a required performance level. This observation has a practical implication when writing service contracts. System managers may offer customers reduced premiums if they agree to longer waits. Understanding how the tolerable wait affects the budget size is critical in the design of these contracts.

Our multiple location system's structure is a flat structure with all locations situated in a single tier and without the option of lateral transshipment of spares between locations. Extending the model to more complex structure such as multiple-echelon systems is a challenge left for future studies.

© IEOM Society International

References

Avinadav, T., and Henig, M. I., Exact accounting of inventory costs in stochastic periodic review models, *International Journal of Production Economics*, vol. 169, pp. 89-98, 2015.

Balakrishnan, N. R., and Chakravarty, A. K., Product design with multiple suppliers for component variants, *International Journal of Production Economics*, vol. 112, no. 2, 723-741, 2008.

Bischak, D. P., Robb, D. J., Silver, E. A., and Blackburn, J. D., Analysis and management of periodic review, order-up-to level inventory systems with order crossover, *Production and Operations Management*, vol. 23, no. 5, pp. 762-772, 2014.

Caggiano, K., Jackson, L., Muckstadt, A., and Rappold, A., Efficient computation of time-based customer service levels in a multi-item, multi-echelon supply chain: A practical approach for inventory optimization, *European Journal of Operational Research*, vol. 199, no. 3, pp. 744-749, 2009.

Chen, F., and Zheng, Y. S., Waiting time distribution in (T, S) inventory systems, *Operations Research Letters*, vol. 12, no. 3, pp. 145-151, 1992.

Demoulin, N. T., and Djelassi, S., Customer responses to waits for online banking service delivery, *International Journal of Retail & Distribution Management*, vol. 41, no. 6, pp. 442-460, 2013.

Disney, S. M., Gaalman, G. J., Hedenstierna, C. P. T., and Hosoda, T., Fill rate in a periodic review order-up-to policy under auto-correlated normally distributed, possibly negative, demand, *International Journal of Production Economics*, vol. 170, pp. 501-512, 2015.

Disney, S. M., Maltz, A., Wang, X., and Warburton, R. D., Inventory management for stochastic lead times with order crossovers, *European Journal of Operational Research*, vol. 248, no. 2, pp. 473-486, 2016.

Dreyfuss, M., and Giat, Y., Optimal spares allocation to an exchangeable-item repair system with tolerable wait, *European Journal of Operational Research*, vol. 261, no. 2, pp. 584-594, 2017.

Dreyfuss, M., and Giat, Y., Optimal allocation of spares to maximize the window fill rate in a two-echelon exchangeable-item repair system, *European Journal of Operational Research*, vol. 270, no. 3, pp. 1053-1062, 2018.

Dreyfuss, M., and Giat, Y., Allocating spares to maximize the window fill rate in a periodic review inventory system, *International Journal of Production Economics*, vol. 214, pp. 151-162, 2019.

Giat, Y., and Dreyfuss, M., The window fill rate in a periodic review inventory system with order crossover, working paper, Jerusalem College of Technology, 2020.

Hadley, G., and Whitin, T. M., *Analysis of Inventory Systems*, Prentice Hall, New Jersey. 1963.

Heese, H. S., Single versus multiple sourcing and the evolution of bargaining positions, *Omega*, vol. 54, pp. 125-133, 2015.

Johansen, S. G., Emergency orders in the periodic-review inventory system with fixed ordering costs and stochastic lead times for normal orders, *International Journal of Production Economics*, vol. 209, pp. 205-214, 2019.

Katz K., Larson B., and Larson, R., Prescriptions for the waiting in line blues: entertain, enlighten and engage, *Sloan Management Review*, vol. Winter, pp. 44-53, 1991.

Prak, D., Teunter, R., and Riezebos, J., Periodic review and continuous ordering, *European Journal of Operational Research*, vol. 242, no. 3, pp. 820-827, 2015.

Kouki, C., Babai, M. Z., Jemai, Z., and Minner, S., Solution procedures for lost sales base-stock inventory systems with compound Poisson demand, *International Journal of Production Economics*, vol. 209, pp. 172-182, 2019.

Mukherjee, A., and Tsai, Y., Multi-sourcing as an entry deterrence strategy, *International Review of Economics & Finance*, vol. 25, pp. 108-112, 2013.

Namdar, J., Li, X., Sawhney, R., and Pradhan, N., Supply chain resilience for single and multiple sourcing in the presence of disruption risks, *International Journal of Production Research*, vol. 56, no. 6, pp. 2339-2360, 2018.

Skellam, J. G., The frequency distribution of the difference between two Poisson variates belonging to different populations, *Journal of the Royal Statistical Society*, Series A (General) vol.109(Pt 3), p. 296, 1946.

Srivastav, A., and Agrawal, S., Multi-objective optimization of mixture inventory system experiencing order crossover, *Annals of Operations Research,* 2018. doi:10.1007/s10479-017-2744-4

Svoronos, A., and Zipkin, P., Evaluation of one-for-one replenishment policies for multi-echelon inventory systems, *Management Science*, vol. 37, no. 1, pp. 68-83, 1991.

Tempelmeier, H., and Fischer, L., Approximation of the probability distribution of the customer waiting time under an (r,s,q) inventory policy in discrete time, *International Journal of Production Research*, vol. 48, no. 21, pp. 6275-6291, 2010.

© IEOM Society International

Tempelmeier, H., and Fischer, L., A procedure for the approximation of the waiting time distribution in a discrete-time (r,S) inventory system, *International Journal of Production Research*, vol. 57, no. 5, pp. 1413-1426, 2019.
van der Heijden, M. C., and De Kok, A. G., Customer waiting times in an (R,S) inventory system with compound Poisson demand, *Zeitschrift fur Operations Research*, vol. 36, no. 4, pp. 315-332, 1992.
van Donselaar, K. H., and Broekmeulen, R. A., Determination of safety stocks in a lost sales inventory system with periodic review, positive lead-time, lot-sizing and a target fill rate, *International Journal of Production Economics*, vol. 143, no. 2, pp. 440-448, 2013.
Wensing, T., *Periodic Review Inventory Models*, Springer, New York, 2011.
Wensing, T., and Kuhn, H., Analysis of production and inventory systems when orders may cross over, *Annals of Operations Research*, vol. 231, no. 1, pp. 265-281, 2015.

Biography

Yahel Giat is a faculty member in the Department of Industrial Engineering and Management in the Jerusalem College of Technology. He holds a Ph.D. and an MSc. in Industrial Engineering from the Georgia Institute of Technology, an MSc. in Economics, a B.Sc. in Electrical Engineering and a B.A. in Computer Sciences from the Israel Institute of Technology. His current research efforts are inventory and operations management in the field of medicine and to that end he is currently pursuing a degree in Medicine from the Hebrew University

© IEOM Society International