Manuscript Details

Manuscript number OMEGA_2016_871

Title Basestock Policies with Reservations

Article type Research Paper

Abstract

All intensively studied and widely applied inventory control policies satisfy demand in accordance with the First-Come-First-Served (FCFS) rule, whether this demand is in backorder or not. Interestingly, this rule is sub-optimal when the fill-rate is constrained or when the backorder cost structure includes fixed costs per backorder \emph{and} costs per backorder per unit time. In this paper we study the degree of sub-optimality of the FCFS rule for inventory systems controlled by the well-known base-stock policy. As an alternative to the FCFS rule, we propose and analyze a class of generalized base-stock policies that reserve some maximum number of items in stock for future demands, even if backorders exist. Our analytic results and numerical investigations show that such alternative stock reservation policies are indeed very simple and considerably improve either the fillrate or reduce the total cost, without having much effect on the backorder level.

Keywords Inventory Theory and Control; base-stock models; stock reservation policies

Manuscript category Research Paper- Supply Chain Management or Business Analytics Applications

Corresponding Author Nicky van Foreest

Order of Authors Nicky van Foreest, Ruud Teunter, Aris Syntetos

Submission Files Included in this PDF

File Name [File Type]

answer_on_review1.pdf [Response to Reviewers]

answer_on_referee.pdf [Response to Reviewers]

highlights.pdf [Highlights]

fts.pdf [Manuscript File]

Response to reviewer 1 on the article 'Basestock Policies with Reservations'

Nicky van Foreest, Ruud Teunter, Aris Syntetos

July 29, 2017

We thank the referee for the comments on our paper. To discuss these comments we summarize them for convience and then state how we dealt with it.

a) Typo on Page 5, line 55.

   Repaired

b) page 10, Figure 1: It is not clear for me, which situation for the delivery lead-time is illustrated in Figure 1. I had an exponentially distributed replenishment lead- time in mind, but this does not fit to the rates in the Figure. Please explain which assumption for the replenishment lead-time is made for the Figure.

   It was indeed unclear, and also incomplete. We repaired the figure itself and changed the caption to indicate that replenishment times are exponentially distributed.

c) Page 13, line 257: Please explain, why demands and replenishments over the time interval $(t, t+L]$ are uniformly distributed. Alternatively, give a reference. (Similar on page 15, line 284)

   We now provide a reference to the book of Tijms, as we already refered to this book at other places in the article. The result is actually very interesting, and it states that for a Poisson process, given that n arrivals occured in some (finite) interval, the arrival epochs are uniformly distributed. The result can also be found in other books, such as Ross [1], as 'order statistics'.

d) Page 14: At the moment I could not follow the derivation for the formula for $Q^{b,i}(k,n)$. I am missing a clear definition of $p^{b,i}(k,n)$ and $q^{b,i}(k,n)$. Further, more explanation is needed. Perhaps a small example can also help as an illustration.

   After much efforts to satisfy your request, it turned out that our analysis was too simple and, in fact, not correct. We tried to rederive our results and discovered that we overlooked an elementary, but very important, detail. The process $\{B(t), I(t)\}$ for constant leadtimes is not a Markov process. As a result, our analysis provides a good approximation (we checked this with simulation), but was not entirely correct. Thus, no wonder that it was hard to follow. We are grateful for your critical reading. We decided to remove this part from the paper.

   The consequences are not really severe in our opinion because most of our numerical work relied on the exponential leadtimes; the analysis of this case was (and is) correct. Thus, all managerial insights remain the same.

   The case with constant L and $r = 1$ was also correct, so this part is not affected. In fact, we expanded our analytical work for this case, because while trying to address your above comment, we realized that a few extra steps were easy to achieve when $r = 1$.

   We decided to handle the cases with constant L and $r > 1$ with simulation. This enabled us to check all our other cases too. All models are now consistent.

   Finally, we changed Figure 3 into Figures 3 and 4, also in view of your point (f) below.

   All in all we hope that new text is much clearer now.

e) Page 14, line 274: I think the last equation in (16) is only correct, if all paths in the lattice have the same probability to be chosen. Why is this correct?

Given the changes to the updated paper, this point is no longer relevant.

f) Figure 2: Why is the average number of backorders increasing with increasing reser- vation level? I thought that the reservation level should help to prevent additional backorders and should only lead to an increase of the duration of a backorder. Can you please explain this?

We address this explicitly now in Eq. 11, which states that for the basestock policy and reservation policies alike,

$$\mathsf{E}\{I\} = S - \lambda \mathsf{E}\{L\} + \mathsf{E}\{B\},$$

that is, the expected on-hand inventory is equal to the order-up-to level S minus the expected demand during the leadtime plus the expected number of backlogged demand. Thus, if the reservation level $r > 0$ makes the inventory level higher (which we expect to be the case), then $\mathsf{E}\{B\}$ must also increase, provided we don't change S and $\lambda \mathsf{E}\{L\}$.

It took us some time to realize that the above generic equation was the easiest key to seeing this, perhaps, paradoxical relation.

References

[1] S.M. Ross. *Stochastic Processes*. John Wiley & Sons, 2nd edition, 1996.

Response to referee on the article 'Basestock Policies with Reservations'

Nicky van Foreest, Ruud Teunter, Aris Syntetos

July 29, 2017

We thank the referee for his/her comments on the paper. To discuss these comments we state them for convience and then explain how we dealt with it.

1. Typo on Page 5, line 55.

   Repaired

2. P.11, l. 206: The authors note 'Under the assumption that the lead-time L is exponentially distributed, it is clear that the joint process $\{B(t), I(t); t \geq 0\}$ is an ergodic Markov process.' Just claiming 'it is clear' is a little too vague. Either, the authors give a few more details why it is 'clear', or they should give at least a good reference where the reader can find some theory.

   We provide a specific reference, i.e., Theorem 4.3.1, in Tijms' book (a reference we already included in the paper), and comment on the existence on the stationary distribution p in Section 4.

3. P.12, l.231: Should that be $C_r^* = \min_S\{C_{S,r}\}$?

   Sure. We repaired it.

4. P.13, l. 257 and p.15, l. 285: In both paragraphs, it is stated that demands arrive 'uniformly' distributed. How is this related to the assumption on p. 8, l. 168 that the demand arrives in single units in accordance to a Poisson process with rate l?

   We now provide a reference to theorem 1.1.5 in Tijms' book. The result is known as the 'order statistic', that is, given the number of Poisson arrivals in some interval, then the arrival epochs are uniformly distributed on the interval.

5. In summary, the referee found some notations in the paper slightly confusing.

   To meet this point we decided to give this aspect a rather rigorous overhaul. However, after much efforts to satisfy this point, it turned out that our analysis was too simple and, in fact, not correct. We overlooked an elementary, but very important, detail: the process $\{B(t), I(t)\}$ for constant leadtimes is simply not a Markov process. As a result, our analysis provides a good approximation (we checked this with simulation), but was not entirely correct. We decided to remove this part from the paper. We like to thank you for your critical reading on this point.

   Luckily the consequences are not really severe in our opinion because most of our numerical work relied on the exponential leadtimes; the analysis of this case was (and is) correct. Thus, all managerial insights remain the same.

   The case with constant L and $r = 1$ was also correct, so this part is not affected. In fact, we expanded our analytical work for this case, because while trying to address your above comment, we realized that a few extra steps were easy to achieve when $r = 1$.

We decided to handle the cases with constant L and $r > 1$ with simulation. This enabled us to check all our other cases too. All models are now consistent.

Finally, we changed Figure 3 into Figures 3 and 4, also in view of your point (10) below.

All in all we hope that new text is much clearer now.

6. P14, 1.275. In summary the referee asks us to explain how we dealt with the numerical aspects of solving the stationary distribution p for a Markov chain with transition rate matrix Q.

   We now provide a reference to Tijms' book, section 3.4, where a few numerical methods are discussed how to solve $pQ = 0$. We also refer to the numerical package that we used to actually compute p.

   We are open to making our code publicitly available, either on our homepage or via github. If the reviewer thinks this will serve a general purpose (and Omega or the publisher does not object), then we would be glad to comment the code and share it.

7. P. 16, l. 295 (but also earlier in the paper): Maybe I missed it, but the authors never clearly define what Q is. The reader get a rough intuition from the model analysis, but maybe a list of notation would be helpful.

   Given the above mentioned changes, this point is no longer relevant as it refers to text that has been removed (which we did to meet the earlier comment 5)

8. MODEL ANALYSIS: I missed one important aspect. How did you optimize the base stock level S and the reservation level r? Section 4 'Analysis' is mainly about determining the steady-state probabilities for a given S and r. In Section 2 you note that the fixed backorder cost complicate an exact optimality analysis. Nevertheless, I think the paper needs some more analysis regarding the solvability of the problem and how one can find the joint optimal solution (r^*, S^*).

   We just used full enumeration over a large grid. In the new remark, Remark 5.1, we specifically address how we found suitable upper bounds on S and r to ensure that the optimal values would lie in the grid.

9. In the numerical section, the authors present several graphs where they either fix r or S and show the effect on expected inventory level, expected backorder, and fill rate. Is it possible to find the joint optimum? In Section 5.2, the authors further present the influence of the fixed backorder cost π on the percentage gain G from allowing stock reservation. Here it would be interesting: (i) what is the optimal solution (r^*, S^*) for the different levels of p and (ii) what $E(I)$ and $E(B)$ under the optimal policy? This relates to my previous point and requires beforehand a mathematical analysis how to find $r^*(S)$ and $S^*(r)$ as well as (r^*, S^*).

10. P. 18, Figure 2: I would use separate captions for the left and right figure, e.g., Figure 2a and Figure 2b. Additionally, the captions should be more 'to the point'.

    We decided to split most of the figures of the paper into separate figures. We also organized the former Figure 3 such that now S runs along the x-axis and r is fixed along a set of experiments. In our opinion the graphs are much nicer now.

    We hope to have improved the clarity our captions.

11. Managerial explanation of the graphical results: The authors primarily repeat the numbers that can be seen from the graphs. However, they should also give a brief managerial intuition why these results are like that. What are drivers or trade-offs?

    We now provide short summaries at the end of each numerical section. In hindsight, we completely overlooked this critical point. Hopefully the referee is satisfied with these more managerial summaries.

1. We consider single-item inventory systems and motivate that, for many practical purposes, the cost structure should include costs per backorder per unit time and costs per backorder.

2. Under such cost structures, it is no longer optimal to serve backordered demand in a FIFO sequence.

3. We propose another class of policies, so-called reservation policies, to handle replenishments and backorders. Such policies aim to keep sufficient items on-hand to meet new demand, i.e., prevent the occurrence of backorders. Only when the on-hand inventory is above a certain level, a reservation policy starts to use replenishment orders to meet backlogged demand. Thus, it does not satisfy backorders when the on-hand inventory level is too low.

4. We provide analytical and numerical models to show that reservation policies can significantly reduce average cost, while the average backlogging times do not increase considerably.

,

,

Basestock Policies with Reservations

Nicky D. Van Foreest, Ruud H. Teunter

Department of Economics and Business, University of Groningen, P.O.Box 800 9700 AV Groningen, The Netherlands

Aris A. Syntetos

Logistics and Operations Management, Cardiff Business School, Cardiff University, Cardiff CF10 3EU, UK

Abstract

All intensively studied and widely applied inventory control policies satisfy demand in accordance with the First-Come-First-Served (FCFS) rule, whether this demand is in backorder or not. Interestingly, this rule is sub-optimal when the fill-rate is constrained or when the backorder cost structure includes fixed costs per backorder *and* costs per backorder per unit time. In this paper we study the degree of sub-optimality of the FCFS rule for inventory systems controlled by the well-known base-stock policy. As an alternative to the FCFS rule, we propose and analyze a class of generalized base-stock policies that reserve some maximum number of items in stock for future demands, even if backorders exist. Our analytic results and numerical investigations show that such alternative stock reservation policies are indeed very simple and considerably improve either the fillrate or reduce the total cost, without having much effect on the backorder level.

Keywords: Inventory Theory and Control; base-stock models; stock reservation policies

*Corresponding author

Email addresses: `n.d.van.foreest@rug.nl` (Nicky D. Van Foreest), `r.h.teunter@rug.nl` (Ruud H. Teunter), `syntetosa@cardiff.ac.uk` (Aris A. Syntetos)

URL: `http://nicky.vanforeest.com` (Nicky D. Van Foreest)

1. Introduction

Nearly all single-item continuous-time inventory models with positive replenishment leadtimes and backlogging use one of two 'extreme' cost structures. In one extreme, backorder costs are proportional to the backorder duration and there is no fixed cost associated with backordering a customer. In the other extreme, there is just a fixed cost per backorder and the cost per unit time per backorder is ignored. See e.g. Axsäter (2006) or Zipkin (2000) for further discussion of both cost models.

It is well-known, see e.g. (Zipkin, 2000, Section 3.3.), that when the cost b per unit time per backorder is positive and the cost π per backorder is 0, meeting backorders in a First-come-First-Served (FCFS) sequence minimizes backlogging durations, and that a base-stock policy makes an optimal trade-off between backlogging and inventory cost. In the other case, i.e., $\pi > 0$ and $b = 0$, there is no need to satisfy backorders at all; in fact, it is best to drop the FCFS rule altogether, and consider an inventory model with lost sales. In this situation the incentive is to limit the rate at which backorders occur, thereby explaining the popularity of the fillrate service criterion for such cases.

It is actually quite remarkable that most literature on inventories with backlogging and positive leadtimes concentrates on either of the cases, i.e., $(b > 0, \pi = 0)$ or $(b = 0, \pi > 0)$. Hadley and Whitin (1963) discuss in their seminal work, already published over fifty years ago, that a realistic cost structure for backlogging should contain *both* cost components. They model the cost of notifying the backordered customer as a fixed cost and argue that loss of goodwill may be proportional to the backlog duration. Chen and Zheng (1993) also reason that backorder costs are partially fixed and partially duration-dependent, although they do not provide any real life examples. Empirical evidence for the relevance of both cost components is provided by Kline and Betke (2004). In a report prepared for FedEx Services, they distinguish six categories of backorder costs, c.f. Table 1 : 'customer notification', 'information processing', 'packaging', 'warehousing', 'freight', and 'other issues'. Note from the descriptions that

Table 1: Categories and estimates of costs per backorder, based on an empirical study involving 40 companies in a variety of industries (Kline and Betke (2004)).

Cost Category	Description	Cost per Backorder
Customer notification	Notify customer of backordering	\$0.74
Information processing	Extra calls for processing backorders	\$3.94
Packaging	Carton and void	\$0.49
Warehouse	Efficiency loss in picking and handling	\$2.68
Freight	Shipping backordered item to customer	\$5.43
Other	Loss of goodwill or orders, cancellations	???
Total		USD 13.28

five of these categories include fixed costs per backorder and are (mostly) independent of the backorder's duration. One reason is that these costs originate from the need to handle backorders separately from the regular items in the same order line. Only the sixth 'other issues' category concerns less tangible costs related to loss of goodwill and loss of (future) demand. Kline and Betke (2004) do not provide an estimate on the nature of these costs. This is not entirely surprising: as already remarked by Hadley and Whitin (1963), it is very difficult to quantify cost related to these 'other issues'. Notwithstanding this difficulty, it is reasonable to model the unwillingness to wait for long amounts of time as cost per unit time, hence this sixth component is best modeled by a *combination* of a fixed cost per backorder and a cost per unit time. Thus, taking the costs of all six categories together, we see that fixed backorder costs often make up a large part of the total backorder costs for companies.

Besides stock-out costs related to logistics, firms can also opt to compensate customers for out-of-stock experiences by offering monetary payments, store credit, rain checks or discount coupons, see Su and Zhang (2009). Verhoef and Sloot (2005) find through consumer surveys that such marketing instruments are appreciated by customers, and Bhargava et al. (2006) and Dong and Rudi (2004) discuss their use by firms. These marketing related costs are all per backorder rather than per backorder per unit time.

Based on these logistic and marketing arguments, we contend that practically relevant cost structures should comprise a fixed cost π per backorder *and* a cost b per backorder per unit time.

Given this situation, we argue in this paper that the sequence in which backorders are satisfied needs attention. The reason is neither the policy that satisfies backorders using the FCFS rule, which is optimal for $(b > 0, \pi = 0)$, nor the policy that never satisfies backorders, which is optimal for $(b = 0, \pi > 0)$, is close to optimal for cases in which the cost includes components per backorder and per backorder per unit time, i.e., when $(b > 0, \pi > 0)$. As an alternative, we recommend for situations with $(b > 0, \pi > 0)$ that we should still meet *all* backorders but *sometimes later* than under the FCFS rule. In other words, we propose to replace the common FCFS rule by another sequencing rule to satisfy backorders.

This new rule is a simple generalization of the base-stock policy. Observe that base-stock policies balance holding costs against the total cost of backordering by suitably setting the order-up-to level S. In our proposal, the main idea is to also address the trade-off between the fraction of backorders (related to π) and their average duration (related to b). Rather than delivering in FCFS sequence, the new policy does *not always* satisfy backorders at the very moment a replenishment arrives, but *reserves* replenishments to increase the on-hand inventory level when it is low. This on-hand stock then serves to directly meet new demand, hence prevent new backorders to occur. Only when the inventory level is above some fixed *reservation level*, backorders are satisfied with replenishments. A consequence of this rule is that average backlogging times are longer than under the FCFS rule, which might seem ‘unfair’. In our opinion, however, fairness should be captured by the relative value of b and π. When π increases, it becomes more important to prevent backlogging.

With this paper we make three contributions. First we start a discussion about the (implicit) use of FCFS to meet backorders. In view of the above, we claim that delivering backorders in a FCFS sequence is often misaligned with practically relevant cost models, i.e., cost models in which fixed backorder costs are not negligible. Second, we propose a class of simple (stock reservation) policies to exploit the possibilities offered by deviating from FCFS. Third, we develop models to analyze the benefits of these reservation policies. In these models we assume that demands arrive according to a Poisson process, and we consider both constant lead-times and exponential replenishment lead-times. For both cases, we provide examples in which we determine all relevant performance measures, including the holding cost, fillrate, number of backorders and backorder duration. By varying the reservation level, we study the trade-off between inventory and backordering levels on the one hand, and fillrate on the other hand. Finally, we compute the costs savings that can be obtained with reservation policies. The numerical investigations provide insightful and encouraging results: using reservation policies can lead to substantial cost savings, up to 30% if π is considerably larger than b, and still more than 5% for more balanced cases.

The remainder of this paper is organized as follows. In Section 2 we discuss relevant literature. Section 3 provides a detailed system description. Next, we consider in Sections 4.2 and 4.1 the consequences of modeling the replenishment lead-times as constant or as exponentially distributed random variables. Section 5 presents numerical insights obtained from these models. In Section 6 we conclude and discuss interesting directions for further research.

2. Related Literature

The cost structure that is most commonly used in the inventory control literature assumes that backorder costs are incurred per backorder per unit time, i.e., analogous to inventory holding costs. Obviously, for such a cost structure it can never be optimal to reserve stock for future demand while backorders exist, and so FCFS is optimal. It is also well-known that then the optimal policy is characterized by a re-order level and an order-up-to level, or just by an order-up-to (i.e., base-stock) level if the fixed ordering cost is negligible, c.f.

e.g. Axsäter (2006) or Zipkin (2000).

The case with $(b = 0, \pi > 0)$ is discussed in (Zipkin, 2000, Section 3.3.7) and it is concluded that, as the cost b per unit time per backorder "captures more of customers' actual experience than" the cost π per backorder, the case $(b > 0, \pi = 0)$ deserves the most focus. Notwithstanding this, in view of Table 1, there are significant costs related to backordering customers, even though these costs are not directly observable by the customer. Thus, customers' experience is not the only relevant aspect that needs to be incorporated in the cost structure, internal costs should not be overlooked. Moreover, just the fact that customers have to wait, independent of the estimated duration, may make them turn away. For this reason also, π should be positive. Therefore, in realistic backordering cost models both π and b need to be non-negligible.

The fixed cost component π, however, complicates an exact optimality analysis, due to the fact that the resulting cost function is in general no longer a convex function of the inventory position (the on-hand inventory minus backorders plus all items on order, if any). Rosling (2002) proves that, under some mild conditions on the demand distribution, the resulting cost function is a quasi-convex function of the inventory position, so that base-stock policies are still long-run average optimal for such inventory systems. However, it is important to remark that this optimality result only holds under the assumption that backorders are satisfied on a FCFS basis. As discussed in the Introduction, under the general backorder cost structure, it may be better to reserve some stock for possible future demands even if backorders exist.

Even though the concept of stock reservation is not new, to the best of our knowledge it has not been considered for basic inventory systems with a single stocking point and a single demand class. In multi-echelon models, as considered by many authors, see e.g. Clark and Scarf (1960); Federgruen and Zipkin (1984); Graves and Willems (2000); Rosling (1989), stock can be reserved (although this is not the term usually employed) at an upstream echelon to correct future imbalances at downstream echelons. Thus, these policies use stock reservation to reduce the effects of demand variability. In our setting, we reserve stock to reduce the occurrence of backorders. Inventory policies with (complete or partial) reservations have also been considered for systems with multiple demand classes, in which case those policies could be referred to as *rationing* or *discrimination* policies, depending on the context, see, e.g., de Véricourt et al. (2001); Gayon et al. (2009); Marklund (2006); Teunter and Klein Haneveld (2008). The latter article also provides a classification of the existing literature on systems with multiple demand classes. Such systems arise in practice if, for instance, customers can choose different service level agreements (platinum, gold, silver), if equipment criticality affects the criticality of spare part demand, or if customers receive preferential treatment based on Advanced Demand Information (ADI). We refer to Marklund (2006) for a discussion of other mechanisms that achieve some form of reservation. It might seem that our reservation policies also split the demand into two types, i.e., new demand and backordered demand, so that new demand is served directly from on-hand stock (if available) while backordered demand is only served when the on-hand inventory level is sufficiently high. This policy, however, differs considerably from rationing policies: there the class of the order depends on the customer type, which is known beforehand, not on the inventory level as observed by arrivals. In other words, rationing policies deal with situations in which there is a natural classification of customer demand, while we deal with only one demand class. In summary, in the multi-echelon or rationing settings, the policies do not address (the prevention of) backorders, nor the consequences of serving backorders from replenishments.

3. Model

We consider a single-item continuous-review inventory system with backordering and positive lead-times. Demand arrives in single units in accordance with a Poisson process with rate λ; let $N(t)$ denote the demand that arrived up to time t and, with this, let $N(s,t] = N(t) - N(s)$ denote the the demand during $(s,t]$. Each demand generates a replenishment order that will be delivered a replenishment lead-time later. We assume that the replenishment lead-times L_i, $i = 1, 2, \ldots$, associated with the i-th order are i.i.d. random variables with mean $\mathbb{E}L$.

Let $I(t)$ be the inventory level at time t, and $B(t)$ the number of backorders. The (right-continuous with left limits) inventory process $\{I(t); t \geq 0\}$ and backorder process $\{B(t); t \geq 0\}$ are controlled by a base-stock policy with order-up-to level S, and a *reservation* level $r \geq 0$. This policy places a replenishment order when the inventory position hits a reorder level of $S - 1$, so that the inventory position (after ordering) is always equal to the base-stock level S. Only when the inventory level is equal or higher than r, replenishments are used to meet the backorders. Otherwise, they are added to the inventory on hand. The policy also includes a rejection level R to limit the number of backorders in the system: demand that arrives when the backorder level is R and the on-hand inventory is zero will be rejected. The aim of including R is to avoid technical and numerical difficulties related to infinite state spaces. (For notational convenience, we suppress in the notation of the random variables $I(t)$ and $B(t)$ the dependency on the parameters of the policy.) The working of the policy is illustrated in Figure 1 for the case that $S = 4$, $r = 2$, and $R = 3$.

More formally, suppose that a demand arrives at time t, then the behavior is the same as under the regular base-stock policy, i.e.,

$$\big(B(t), I(t)\big) = \begin{cases} (B(t_-), I(t_-) - 1), & \text{if } I(t_-) > 0, \\ (B(t_-) + 1, 0) & \text{if } I(t_-) = 0, \\ (R, 0) & \text{if } B(t_-) = R, \text{ and } I(t_-) = 0, \end{cases} \tag{1}$$

where $B(t_-) = \lim_{s \uparrow t} B(s)$, and similarly for $I(t_-)$. Thus, demand that cannot be met from inventory is backordered. When a replenishment arrives at time t, the reservation level comes into play:

$$(B(t), I(t)) = \begin{cases} (0, I(t_-) + 1), & \text{if } B(t) = 0, \\ (B(t_-), I(t_-) + 1)), & \text{if } 0 \leq I(t_-) < r \text{ and } B(t) > 0, \\ (B(t_-) - 1, r), & \text{if } I(t_-) = r \text{ and } B(t) > 0. \end{cases} \tag{2}$$

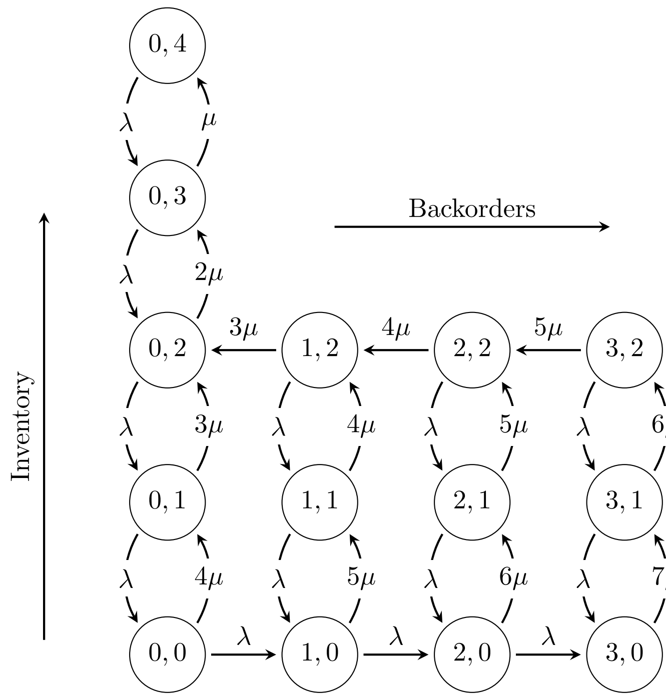

Figure 1: The transitions of an inventory system controlled by a reservation policy with order-up-to level $S = 4$, reservation level $r = 2$ and rejection level $R = 3$. A state (b, i) has b backorders and i items on hand. Thus, in state (b, i), the number of outstanding replenishments is $k = S + b - i$. The replenishment leadtimes are exponentially distributed with mean μ. The arrows labeled by λ correspond to demand arrivals; the arrows with label $k\mu$ correspond to replenishments when k orders are outstanding.

Thus, in case $B(t) > 0$ and $I(t) < r$, each replenishment is put on stock, rather than used to decrease the number of backorders. Only when $I(t) = r$, replenishments are used to fulfill backorders.

If we write $D(t)$ for the number of outstanding replenishments at time t, then the base-stock policy with reservations ensures that $I(t)$, $B(t)$, $D(t)$, and the base-stock level S are related such that

$$S = I(t) - B(t) + D(t), \text{ for all } t. \tag{3}$$

If the reservation level $r = 0$ it follows from (1) and (2) that

$$I(t) \cdot B(t) = 0, \qquad \text{for all } t. \tag{4}$$

That is, if there is on-hand stock, there cannot be backorders, and vice versa. Consequently, in case $r = 0$, the joint backorder-inventory process reduces to the (essentially) one-dimensional regular base-stock model without reservations. It is well-known that in this system the number of replenishments at an arbitrary moment in time is equal to the number of customers in an $M/G/\infty$ queue, hence the steady state probability p_n that the system contains n replenishments is given by

$$p_n = e^{-\lambda\,\mathbb{E}L}\frac{(\lambda\,\mathbb{E}L)^n}{n!}. \tag{5}$$

Observe also that p_n is insensitive to the distribution of the lead time so that this expression applies to L constant or exponentially distributed.

In case the reservation level r is not zero, the constraint in Eq. (4) no longer holds and the analysis of the inventory system becomes considerably more difficult. In particular, an insensitivity result similar to (5) does not exist, as we show numerically in Section 5. For this reason we restrict the analysis in the remainder to exponentially distributed and constant lead-times. In Section 4.1 we show that for both leadtime distributions, stationary distributions $p = (p_{b,i})$ exist that are equal to the limiting time-averages

$$p_{b,i} = \lim_{t\to\infty}\frac{1}{t}\int_0^t 1\{B(s) = b, I(s) = i\}\,ds, \tag{6}$$

where $1\{\cdot\}$ is the indicator variable[1]. Let (B, I) be the (pair of) limiting random variables of $\{B(t), I(t)\}$ as $t \to \infty$.

It is of interest to remark in passing that the sum of the probabilities 'along a diagonal' in Figure 1 must add up to the probabilities in (5). Specifically, if $n = S + b$, then

$$p_{b,0} + p_{b+1,1} + \cdots + p_{b+r,r} = p_n. \tag{7}$$

This follows since an external observer who just counts the number of 'jobs' in the system and does not distinguish between whether the job is an inventory item or a backlogged demand will attribute the right-hand probability to the lumped set of states along the diagonal.

[1] $1\{B\} = 1$ if B is true and $1\{B\} = 0$ otherwise.

Once the stationary distribution $\{p_{b,i}\}$ is known, we can compute, as functions of r and S, the expected inventory level

$$\mathbb{E}I = \sum_{b,i} i\, p_{b,i}, \tag{8}$$

the expected number of backorders

$$\mathbb{E}B = \sum_{b,i} b\, p_{b,i}, \tag{9}$$

the fillrate (as follows from the PASTA property, Wolff (1982))

$$FR := 1 - \mathbb{P}(I = 0) = 1 - \sum_{b} p_{b,0}, \tag{10}$$

and the rejection probability

$$PR := \mathbb{P}(B = R, I = 0) = p_{S+R}, \tag{11}$$

where p_{S+R} is given by (5) for $n = S + R$.

If $S + R$ is so large that the rejection probability can be neglected, we can take the limit of the time-average of Eq. (3) to see that $S = \mathbb{E}I - EB + \lambda\,\mathbb{E}L$, and where we use that the expected demand during the leadtime $\mathbb{E}D = \lambda\,\mathbb{E}L$. Thus, $\mathbb{E}I$ and $\mathbb{E}B$ are related through

$$\mathbb{E}I = S - \lambda\,\mathbb{E}L + \mathbb{E}B. \tag{12}$$

Finally, noting that $\lambda(1 - FR)$ is the rate at which demand is backlogged, it follows from Little's law (Little (1961)) that the average backorder duration satisfies

$$\mathbb{E}W_B = \frac{\mathbb{E}B}{\lambda(1 - FR)} \tag{13}$$

Note that these relations already provide some insight into the consequences of stock reservation policies. We expect, by reserving items, that the average on-hand inventory $\mathbb{E}I$ increases. If S, λ and $\mathbb{E}L$ remain the same, an immediate consequence of (12) is then that $\mathbb{E}B$ then also increases, which is perhaps somewhat counter intuitive. From (13) we conclude that it is undesirable to aim for a fillrate of nearly 1 for any inventory system that allows for backlogging: the average backlog time will become very large.

The cost structure includes a holding cost h per item on-hand per unit time, a cost b per backorder per unit time, and a cost π for each backorder. We do not include ordering costs—in a sense this is implied by the fact that replenishments occur in single units. With respect to rejection costs, recall that we use the rejection level R to limit the state space. In our (numerical) investigation, we search for an R that is so large that the rejection probability PR is negligible. Thus, we do not consider rejection costs. These considerations lead us to define the long-run expected cost of a reservation policy with base-stock level S and reservation level r as

$$\begin{aligned} C_{S,r} &= h\,\mathbb{E}I + b\,\mathbb{E}B + \pi\lambda(1 - FR) \\ &= h(S - \lambda\,\mathbb{E}L) + (h + b)\,\mathbb{E}B + \pi\lambda(1 - FR), \end{aligned} \tag{14}$$

where we use (12). Note that the performance measures $\mathbb{E}I$, and so on, implicitly depend on S and r. The minimal cost for reservation level r becomes then

$$C_r^* = \min_S C_{S,r}, \tag{15}$$

and the minimal cost without reservations, i.e., $r = 0$, as

$$C_0^* = \min_S C_{S,0}. \tag{16}$$

With these concepts, let the *relative efficiency gain* be given by

$$G = 100\% \frac{C_0^* - C_r^*}{C_r^*}. \tag{17}$$

4. Analysis

First we analyze the inventory model under a policy with reservations when the lead-times are exponential, then we consider the model with constant lead-times.

4.1. Exponential lead-times

If replenishment lead-times are exponentially distributed then the process $\{B(t), I(t), t \geq 0\}$ is a continuous-time Markov chain. Demand arrives at rate λ and, in view of (3), replenishment orders arrive at rate

$$\mu_{b,i} = (S + b - i)/\,\mathbb{E}L$$

when $B = b$ and $I = i$. From the relations (1) and (2) it is then straightforward to obtain the state space and the associated transition rate matrix Q. Once we have Q, we can obtain (numerically) the stationary probabilities $p = (p_{b,i})$ as the unique normed solution of the system $p\,Q = 0$, c.f., Tijms (2003, Theorem 4.3.1).

With respect to the numerical analysis, Tijms (2003, Section 3.4) discusses a number of general numerical procedures to solve $p\,Q = 0$. For our case we use a numerical toolbox[2] to compute the left-eigenvector v, say, associated with the eigenvalue 0 of the matrix Q. Then $p = v/||v||$, where $||v|| = \sum_i |v_i|$, is the stationary probability vector.

4.2. Constant lead-times and Reservation level $r = 1$

In this section we derive a closed-form solution for the stationary distribution of the inventory system with constant lead-time L and a reservation level $r = 1$. Moreover, the procedure allows us to handle an arbitrary number of backorders, hence in the present case the rejection level R can be set to ∞. In Remark 4.3 below we discuss why the analysis below does not carry over to cases with $r > 1$. To analyze such more general cases we therefore resort to simulation.

Before we analyze the case with $r = 1$ in detail, we remark that the limits (6) exist by Tijms (2003, Theorem 2.2.1) for any reservation level r. The condition to check is that the process $\{B(t), I(t), t \geq 0\}$ is regenerative, but this follows right away from the fact that whenever $N(t - L, t] = 0$, i.e., no arrivals occur

[2] The `linalg` module of `scipy`, a library with numerical tools developed for `python`

during $[(t-L,t]$, we have that $(B(t),I(t))=(0,S)$, hence the inventory process restarts at such moments.

Let us start for the case $r=1$ with the computation of the probabilities

$$p_{b,0}(t) = \mathbb{P}(B(t)=b, I(t)=0)$$

for $t \geq 2L$. Clearly, the process $\{B(t),I(t),t\geq 0\}$ changes at demand epochs, i.e., moments in time at which a demand arrives, and replenishment epochs, i.e., moments in time at which a replenishment arrives. Now observe, e.g., from Figure 2, that when $I(t)=0$ the last epoch before t was a demand epoch rather than a replenishment epoch. Moreover, when $B(t)=b$ and $I(t)=0$, it is necessary that the demand $N(t-L,t]$ during $(t-L,t]$ is equal to $S+b$. Hence,

$$p_{b,0}(t) = \mathbb{P}(D \,\&\, N(t-L,t] = S+b)$$

where we write D for the event

$$D = \{\text{Last epoch before } t \text{ was a demand}\}.$$

Then, by conditioning on the event $N(t-L,t]=S+b$, we see that

$$\begin{aligned} p_{b,0}(t) &= \mathbb{P}(D \,|\, N(t-L,t]=S+b)\, \mathbb{P}(N(t-L,t]=S+b) \\ &= \mathbb{P}(D \,|\, N(t-L,t]=S+b)\, e^{-\lambda L}\frac{(\lambda L)^{S+b}}{(S+b)!}, \end{aligned} \tag{18}$$

where we use that $N(t-L,t]$ is Poisson distributed. Thus, we need to find an expression for $\mathbb{P}(D \,|\, N(t-L,t]=S+b)$.

For this purpose we condition next on the number of outstanding replenishments at time $t-L$, which is equal to the demand $N(t-2L,t-L]$ that occurred during $(t-2L,t-L]$. Now note that, as the demand process is Poisson, the random variables $N(t-2L,t-L]$ and $N(t-L,t]$ are independent. Moreover, by Tijms (2003, Theorem 1.1.5), if it is given that $N(t-2L,t-L]=k$ and $N(t-L,t]=S+b$, the k replenishment and $S+b$ demands occur uniformly distributed over the interval $(t-L,t]$. But in that case, the probability that the last epoch before time t was a demand, must then be equal to $(S+b)/(S+b+k)$,

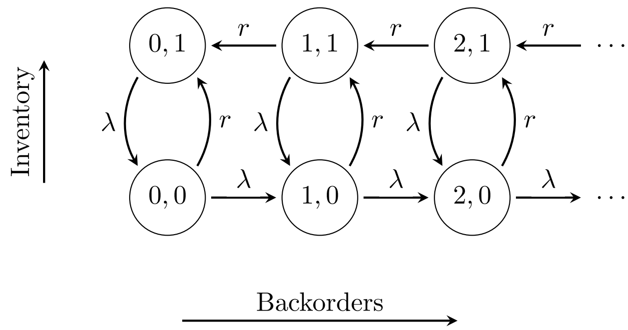

Figure 2: The up- and down-steps corresponding to replenishments, marked as r, and demands, marked as λ, on the process graph with $S = r = 1$, compare Figure 1. Observe that states (b, i) with $i = 1$ can only be entered when a replenishment arrives (whether the distribution of L is exponential or constant). Similarly, the process $\{B(t), I(t)\}$ only enters states (b, i) with $i = 0$ when a demand occurs.

i.e.,

$$\mathbb{P}(D \,|\, N(t-L, t] = S + b, N(t-2L, t-L] = k) = \frac{S+b}{S+B+k}. \tag{19}$$

Thus, with this, we obtain

$$\mathbb{P}(D \,|\, N(t-L, t] = S + b) = e^{-\lambda L} \sum_{k=0}^{\infty} \frac{S+b}{k+S+b} \frac{(\lambda L)^k}{k!}.$$

We can simplify this to

$$\mathbb{P}(D \,|\, N(t-L, t] = S + b) = e^{-\lambda L} \frac{S+b}{(\lambda L)^{S+b}} \int_0^{\lambda L} y^{S+b-1} e^y dy.$$

by using that, for $\alpha > 0$,

$$\sum_{k=0}^{\infty} \frac{\alpha^{k+n}}{(k+n)k!} = \sum_{k=0}^{n} \int_0^{\alpha} \frac{y^{k+n-1}}{k!} \, dy = \int_0^{\alpha} y^{n-1} \sum_{k=0}^{\infty} \frac{y^k}{k!} \, dy = \int_0^{\alpha} y^{n-1} e^y dy,$$

and taking $\alpha = \lambda L$ and $n = S + b$.

Finally, combining the above with (18) and noting that the result does not depend on t, we obtain the stationary distribution

$$p_{b,0} = \frac{e^{-2\lambda L}}{(S+b-1)!} \int_0^{\lambda L} y^{S+b-1} e^y dy. \tag{20}$$

From this and (7) it follows that for $b > 0$, $p_{b,1}$ can be obtained from

$$p_{b,0} + p_{b+1,1} = p_{S+b}, \tag{21}$$

where p_{S+b} is given by (5) with $n = S + b$. Of course, for $b = 0$ and $i \geq 1$, $p_{0,i}(t) = p_{S-i}$, i.e., equal to the basestock probabilities.

Remark 4.1. The fraction of time that the inventory level is zero is

$$\mathbb{P}(I = 0) = 1 - \sum_{b=0}^{\infty} p_{b,0}(t). \tag{22}$$

A particularly interesting formula for $\mathbb{P}(I = 0)$ results in case the basestock level $S = 1$. Combining (20) with (22) (and using the positivity of the summands to reverse the integration and the sum) we have that

$$\begin{aligned}\mathbb{P}(I = 0) &= e^{-2\lambda L} \int_0^{\lambda L} \sum_{b=0}^{\infty} \frac{y^b}{b!} e^y dy = e^{-2\lambda L} \int_0^{\lambda L} e^{2y} dy \\ &= \frac{1 - e^{-2\lambda L}}{2}.\end{aligned}$$

Hence,

$$FR = 1 - \mathbb{P}(I = 0) = \frac{1 + e^{-2\lambda L}}{2}. \tag{23}$$

Thus, in this case the probability of having stock on hand is at least $1/2$. Interestingly, this holds even for situations with very large lead-times, whereas the traditional base-stock policies without reservations would lead to fillrates close to 0%. This shows that even a reservation level of only one unit can increase the fillrate very significantly. The numerical results of the next section will show that, also more generally, small reservation levels can have large effects.

Remark 4.2. From (21), $p_{b,0} \leq p_{S+b}$. Hence, for sufficiently large S, the backlog probabilities can be neglected.

Remark 4.3. As a final remark, the above analysis does not generalize to cases with $r > 1$. The reason is that, for $r > 1$, the occurrence of a replenishment just before t does not guarantee that $I(t) = 1$. To see this, consider first Figure 2. To enter state $(1, 1)$, the last epoch must have been a replenishment epoch. In Figure 1, state $(1, 1)$ can be entered from states $(1, 0)$ by means of a replenishment, or from $(1, 2)$ by means of a demand. Note further that the process $\{B(t), I(t), t \geq 0\}$ is *not* a Markov process. For this, it is necessary to also keep track of the demand and replenishments epochs during $(t - L, t]$.

5. Numerical Results

We now provide several numerical scenarios to analyze the effect of the reservation level r on the main performance measures. We start in Section 5.1 by taking a fillrate perspective, i.e. minimizing the holding cost whilst attaining a certain minimum fillrate. As discussed in Section 1, this corresponds to a backorder cost structure with a fixed cost per backorder only, where maximum benefits can be expected from stock reservations. We then continue in Section 5.2 by also including costs per backorder per unit time.

For the numerical analysis we need to choose the rejection level R. Since we only use a finite value of R to enforce a finite state space, we set R to 30 for all our examples, as this turns out to be so large that the rejection probability is negligible. Note that the search for an appropriate R is trivial as the rejection probability is a decreasing function of R.

5.1. fillrate perspective (cost per backorder only)

In our first scenario, c.f., Figures 3 and 4, we plot the fillrate FR and the (time-)average number of backlogged jobs $\mathbb{E}B$ as functions of S for reservation levels $r = 0$ i.e., the basestock model, $r = 1, 2$ and 3, respectively. Since the average on-hand inventory $\mathbb{E}I$ and $\mathbb{E}B$ are related through (12) we leave out $\mathbb{E}I$. For exponential L the procedure of Section 4.1 yields the computation of the steady-state distribution p, from which FR and $\mathbb{E}B$ follow by the definitions in Section 3. Note that for $r = 0$, FR does not depend on the distribution of L by (5). For constant L and $r = 1$ we use the closed form solutions (20) and (21). Finally, for constant L and $r = 2, 3$ we use simulation. The demand arrival rate is $\lambda = 2$ per period and the average replenishment leadtime is $\mathbb{E}L = 4$ periods.

Based on the graphs in Figure 3 we can make a number of interesting observations. First, it is apparent that, as expected, the fillrate FR increases in S. Second, the results are very similar for constant and exponentially distributed lead-times. Third, for small order-up to levels S, policies with stock reservation, i.e., $r > 0$, dramatically outperform the regular basestock policy with respect to the fillrate criterion. As mentioned earlier, c.f. Eq. (23), when $r = 1$ and $S = 1$, the fillrate FR is already at least $1/2$, while it is nearly 0 for the basestock model when there is high load during the leadtime, which is $\lambda \mathbb{E}L = 2 \cdot 4 = 8$ in this case. Finally, for large values of S, e.g. $S = 12$, setting the reservation level to 1 rather than 0 increases the fillrate from 0.89 to 0.93. Thus, even for $S = 12$, roughly a 4% increase in fillrate can be achieved simply by reserving just one item on stock.

In Figure 4 we see that when S is small, the number of backorders $\mathbb{E}B$ is large (due to the large leadtime demand) and it increases by about half a demand for each increment in r. When S is large, the expected backorder level is anyway small, hence just marginally affected by r.

Furthermore, as constant and exponentially distributed lead-times apparently lead to very similar results, we expect that the obtained insights carry over to practical situations, as the constant and exponential distribution are at either extreme of nearly any reasonable model for practical leadtime distributions. Moreover, given the similar results, we henceforth only consider exponentially distributed lead-times in our explorations, as the latter yields to a numerical analysis rather than simulation.

Another interesting way to use the reservation level r is to increase r and decrease S such $FR \geq 0.9$ with the aim to decrease the average inventory level $\mathbb{E}I$. We investigate this effect in Figure 5; the parameter values are the same as in Figure 3. Reading the graphs from right to left, we observe that by reducing the order-up-to level S from 13 to 12 and increasing r from 0 to 1, the fillrate can be kept above 90%, whilst $\mathbb{E}I$ decreases from about 5 to 4. As observed earlier, and as follows from (12), the expected backorder level remains nearly zero. When S becomes quite a bit smaller, and r larger, $\mathbb{E}I$ remains more or less constant, while $\mathbb{E}B$ becomes quite large. However, comparing this to Figure 4, $\mathbb{E}B$ is large when S is small, even for the basestock model. Thus, in situations in which high fillrates are required, e.g. 90% or higher, S needs to be large. But then the resulting large on-hand inventory level can be reduced simply by reserving one item, while $\mathbb{E}B$ is hardly affected.

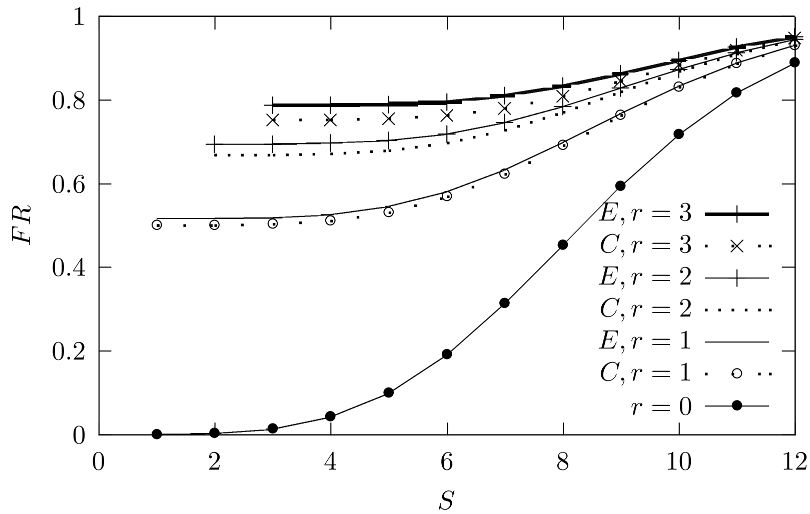

Figure 3: The fillrate FR as a function of the order-up-to level S and reservation level $r = 0$, i.e., the basestock model, and $r = 1, 2$ and 3. The leadtimes are exponential distributed ('E') or constant ('C'). Note that $S \geq r$, hence when $r = 2$, S starts at 2, and likewise for $r = 3$.

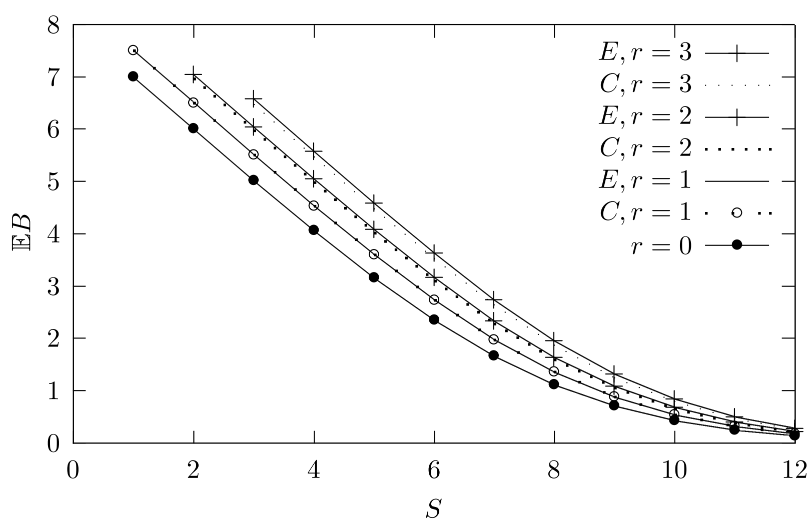

Figure 4: The expected number of backorders $\mathbb{E}B$ as a function of the order-up-to level S and reservation level $r = 0, 1, 2$ and 3, c.f. Figure 3.

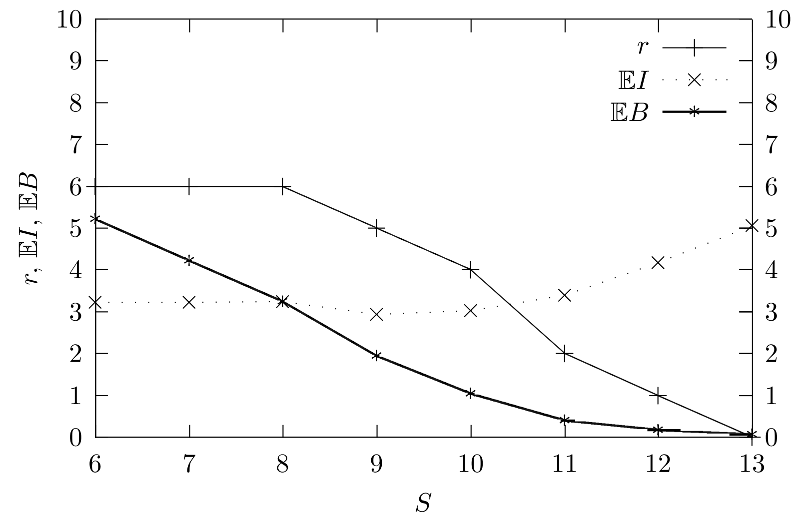

Figure 5: The reservation level r, $\mathbb{E}I$ and $\mathbb{E}B$, as functions of S such that the fillrate is at least 90%.

Finally, in Figure 6, we vary FR from 0.85 to 0.99 and compute the minimal S required by the basestock policy to achieve the given FR and by the $r = 1$ policy. (Finding the minimal S is easy as FR is increasing in S.) Again we observe that the reservation policy often allows to reduce S by one and, as a result, $\mathbb{E}I$ also by about one, thereby implying a saving of almost h in the expected inventory cost. Also, as S is large, the average number of backorders is hardly affected.

The above numerical analysis result in a set of simple heuristics. If the aim is to increase the fillrate (considerably), keep the order-up-to level S fixed and change the reservation level from $r = 0$ to 1 or 2; the backorder level $\mathbb{E}B$ is typically only marginally affected. If the aim is to reduce on-hand inventory, set r to 1 or 2, and decrease S by 1 or 2. Perhaps reservation policies are particularly interesting in the presence of constraints on the inventory space, such as at shops in city centers. As an example, suppose S can be at most 4. For our present parameter setting, we see in Figure 3 that FR increases from about 0.05 to about 0.5, i.e,. a ten-fold increase, by simply reserving one item,

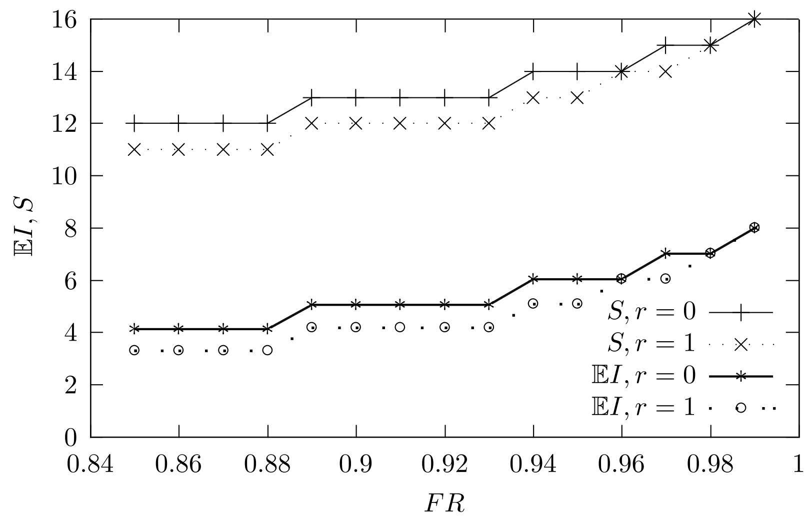

Figure 6: The minimal for S and the associated inventory level $\mathbb{E}I$ for the basestock policy and the $r = 1$ policy such that the fillrate exceeds a given minimal value.

while the relative change in $\mathbb{E}B$ is just $(4.5 - 4)/4 \approx 13\%$.

5.2. General backorder cost structure

In most practical situations both the number of backorders and their (average) duration will matter. To understand this relation we consider a number of numerical examples where we vary the backorder cost per unit time b and cost per backorder π. We normalize the holding cost h to 1, and take $b \in \{1, 3, 10\}$ to study the effect of the relative value of backordering versus holding cost, and $\pi = 0, 1, \ldots, 25$. Finally, we let $\lambda \in \{10, 20\}$ and $\mathbb{E}L \in \{1, 2\}$ to study the effect of the demand rate and the lead-time. Figure 7 shows the relative cost savings, as defined in (17), of the best policy with reservations compared to the best regular base-stock policy. To identify the optimal policy parameters we use a full grid search, c.f., Remark 5.1.

It is evident from Figure 7 that the best reservation policy nearly always outperforms the best regular base-stock policy quite significantly. Moreover, the steep increase at the left for all combinations of λ, $\mathbb{E}L$ and b implies that even for relatively small values of π considerable cost savings can be achieved.

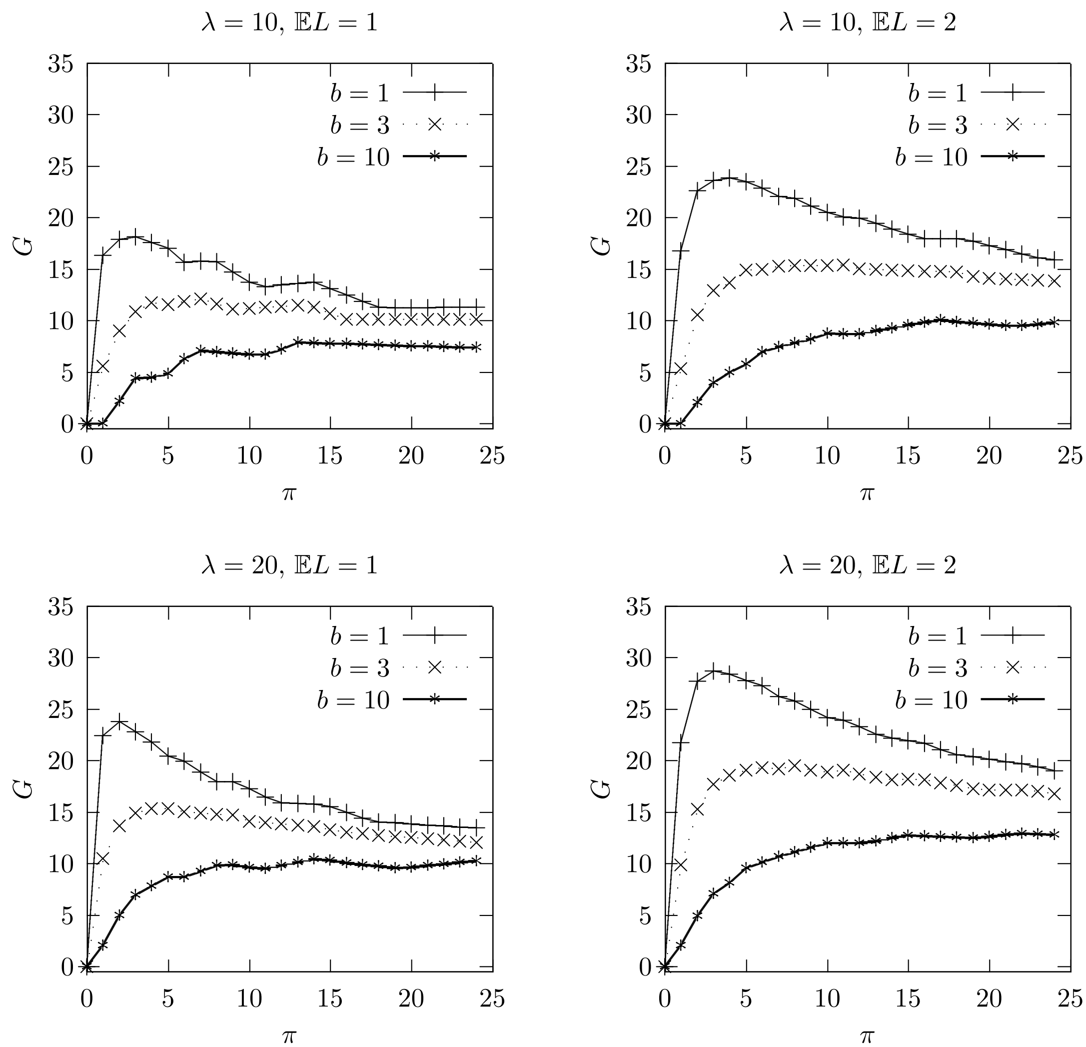

Figure 7: The influence of the fixed backorder cost π on the percentage gain from allowing stock reservations for all combinations of demand rate $\lambda = 10, 20$, lead-time $\mathbb{E}L = 1, 2$ and backorder cost rate $b = 1, 3, 10$, and holding cost rate h normalized to 1.

Of course, the cost savings do decrease as the backorder cost b increases, as this implies that the cost of the number of backorders becomes relatively less important than the cost of backlog duration. However, even for $b = 10$, a moderate value of $\pi = 5$ is sufficient to obtain savings of at least five per cent for all considered combinations of the demand rate and lead-time. We remark that the absolute savings always increase in π as expected, but that they increase at a slower rate than the cost of the best policy without reservations at the high end of the considered π range, which explains why the relative gain starts to decrease at some point.

Viewing Figure 7 from left to right and from top to bottom, it is apparent that both a larger lead-time and a larger demand rate lead to larger savings from reserving stock. This can be explained by the increased safety stock without reservations, creating a higher savings potential from reserving stock. Especially the result that savings from stock reservations increase with the demand rate is very encouraging, because demand rates may be (much) larger in many practical settings. So, real savings may be even higher than the 5 to 30 per cent for the relatively small inventory systems we considered here.

In summary, in realistic cost settings in which backorder costs split into costs per backorder π, whether this is hidden for the customer or not, and cost per backorder per unit time b, reservation policies can reduce overall costs. This is of course reasonable: reservation policies increase the fillrate hence reduce the cost associated with backordering demand. As this cost component is, practically speaking, often not negligible, it actually should have been included in the total cost for any basestock model, even though this leads a (much) harder optimization problem.

Remark 5.1. We approach the search for the optimal S and r in Eqs. (15) and (16) by means of a full grid computation. To obtain suitable boundaries so that the grid contains the optimal S and r, note that necessarily r satisfies $0 \leq r \leq S$. Thus, finding an upper bound on S suffices. For the basestock policy, i.e., $r = 0$ this is easy. From the second expression (14), i.e.,

$$C_{S,r} = hS - h\lambda\,\mathbb{E}L + (h+b)\,\mathbb{E}B + \pi\lambda(1 - FR),$$

we see that $C(S,0)$ is a quasi-convex function of S, as hS is linear in S and $\mathbb{E}B$ and $1 - FR$ are monotone decreasing functions of S. For $r \neq 0$, observe that (5) and (7) imply that $\mathbb{E}B \to 0$ and $1 - FR \to 0$ as $S \to 0$. Thus, for sufficiently large S, $C_{S,r}$ becomes (nearly) linearly increasing in S, for any $0 \leq r \leq S$. In practice, a suitable upper bound S need not be large because of the supra-exponential decay of p_n in (5).

6. Conclusions and Suggestions for Further Research

In this paper, we study single-item inventory systems with backordering under continuous-review, positive lead-times, and controlled by a simple modification of the regular base-stock policy. This modification is based on the introduction of a reservation level r, which is used as follows. When a replenishment order arrives and the inventory level is less than r, the replenishment is put on stock rather than used to satisfy backordered demand (if any). Only when the inventory level is equal to or higher than the reservation level r, backordered demand is met with replenishments. Thus, contrary to regular base-stock policies, a policy with reservations does not satisfy customer demand in a FCFS order. The idea behind reserving replenishments, and deviating from the FCFS delivery rule, is to be able to decrease the average inventory level, hence holding cost, and/or increase the fillrate so that fixed backorder costs decrease. We derive models to study the effects of reservations for the cases with constant lead-times and exponentially distributed lead-times.

With these models we show that, as a result of reserving stock, customers in backorder have to wait somewhat longer on average, but the decrease in fixed backorder costs significantly outweighs the time-related backordering cost, typically leading to a total cost reduction of 10% to 30%. Even if the fixed cost per backorder is relatively small compared to the cost per backorder per unit time, significant savings of 5% or more can be achieved for considered cases. Moreover, our results indicate that even larger savings may be achieved in realistic settings with demand rates and lead times that are larger than the relatively small ones that we considered. Also in situations in which on-hand inventory is (physically or financially) constrained, reservation policies can achieve a considerable increase in fillrate.

Of course, there can be situations where backorder durations are so important that FCFS is preferred, but then our results are also valuable as they provide motivation to reconsider the use of the fill rate as a, or the sole, relevant performance measure. More generally, in situations where both the number and durations of backorders matter, our results offer a starting point for comparing different inventory and replenishment policies and positioning the various inventory performance measures.

Based on these results, we suggest that stock reservations are an interesting alternative to FCFS deliveries in both theory and practice, hence research on adapting the FCFS rule is certainly merited.

In this initial exploration, we concentrate on fixed reservation levels, as such policies are easy to implement in practice and also allow us to embed the reservation structure in a base-stock policy which, as discussed above, is optimal under the FCFS rule. We expect, however, that fixing reservation levels is not (always) optimal, nor that base-stock policies are optimal if stock reservations are allowed. The optimal policy may be quite complex as it can depend on momentary base-stock and inventory levels. We also point out that other mechanisms exist that achieve a form of reservation. For example, sale personnel often quotes longer leadtimes to backordered customers than the average leadtime. This extra time serves as a hedge, but also enables to serve new customers from incoming replenishments. How these mechanisms compare is interesting to pursue further.

Although our exploration is restricted to backorder inventory systems, reservation policies are also interesting to deploy in inventory systems with partial backordering, c.f. Porteus (1990). In this case a shortage becomes a lost sale with probability β and is backlogged with probability $1 - \beta$. Each lost sale results in a penalty cost. As reserving stock reduces the rate at which shortages occur, the cost savings due to reservations may be considerable.

Besides analyzing these partial backordering systems, there are many other opportunities for further research. A first important direction is to extend our model, for example, by including a fixed order cost and considering more general policies such as (Q, r) or (s, S) policies. Another interesting direction is to develop efficient algorithms to compute the optimal policy parameters, both for models where the objective is to minimize the holding cost under a fillrate constraint as well as for models where the objective is to minimize the total cost. These could be based on the assumption of exponential lead times to speed up calculations, as our results indicate robustness with respect to the lead time distribution. Since companies typically stock thousands of different items, developing fast heuristics is also of interest.

References

References

Axsäter, S., 2006. Inventory Control. 2nd ed., Springer.

Bhargava, H., Sun, D., Xu, S.H., 2006. Stockout compensation: Joint inventory and price optimization in electronic retailing. J. Comput. 18, 255–266.

Chen, F., Zheng, Y.S., 1993. Inventory models with general backorder costs. European Journal of Operational Research 65, 175–186.

Clark, A., Scarf, H., 1960. Optimal policies for a multi-echelon inventory problem. Management Science 6, 475–490.

Dong, L., Rudi, N., 2004. Who benefits from transshipment? exogenous vs. endogenous wholesale prices. Management Science 50, 645–657.

Federgruen, A., Zipkin, P., 1984. Approximations of dynamic multilocation production and inventory problems. Management Science 30, 69–84.

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

Graves, S., Willems, S.P., 2000. Optimizing strategic safety stock placement in supply chains. Manufacturing and Service Operations Management 2, 68–83.

Hadley, G., Whitin, T., 1963. Analysis of Inventory Systems. Prentice-Hall, Englewood Cliffs, N.J.

Kline, J., Betke, B., 2004. A study of backorder costs for non-store retailers. White Paper. URL: `http://images.fedex.com/us/retail/backorders.pdf`.

Little, J., 1961. A proof of the queuing formula: $L = \lambda W$. Operations Research 9, 383–387.

Marklund, J., 2006. Controlling inventories in divergent supply chains with advance-order information. Operations Research 54, 988–1010.

Porteus, E., 1990. Stochastic inventory theory, in: Heyman, D., Sobel, M. (Eds.), Stochastic Models. Elsevier. volume 2 of *Handbooks in Operations Research and Management Science*. chapter 12, pp. 605–652.

Rosling, K., 1989. Optimal inventory policies for assembly systems under random demands. Operations Research 37, 565–579.

Rosling, K., 2002. Inventory cost rate functions with nonlinear shortage costs. Operations Research 50, 1007–1017.

Su, X., Zhang, F., 2009. On the value of commitment and availability guarantees when selling to strategic consumers. Management Science 55, 713–726.

Teunter, R., Klein Haneveld, W., 2008. Dynamic inventory rationing strategies for inventory systems with two demand classes, Poisson demand and backordering. European Journal of Operational Research 190, 156–178.

Tijms, H., 2003. A First Course in Stochastic Models. John Wiley & Sons, Chichester.

Verhoef, P., Sloot, L.M., 2005. Out-of-stock: Reactions, antecedents, management solutions, and a future perspective, in: Krafft, M., Mantrala, M.K. (Eds.), Retailing in the 21st Century: Current and Future Trends, Springer Berlin. pp. 239–254.

de Véricourt, F., Karaesmen, F., Dallery, Y., 2001. Assessing the benefits of different stock-allocation policies for a make-to-order production system. Manufacturing and Service Operations Management 3, 105–121.

Wolff, R., 1982. Poisson arrivals see time averages. Operations Research 30, 223–231.

Zipkin, P., 2000. Foundations of Inventory Management. McGraw-Hill/Irwin.