AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the accepted manuscript version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Larsen, C. (2017). Inventory control with and without deliveries in several pieces. OR Spectrum - Quantitative Approaches in Management, 39(2), 607-622. <https://doi.org/10.1007/s00291-016-0466-7>

Publication metadata

Title:	Inventory control with and without deliveries in several pieces
Author(s):	Larsen, Christian
Journal:	OR Spectrum - Quantitative Approaches in Management
DOI/Link:	https://doi.org/10.1007/s00291-016-0466-7
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

Inventory control with and without deliveries in several pieces

Christian Larsen

Aarhus University, BSS, Department of Economics and Business Economics, CORAL,
Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark

Abstract

An inventory system is operated as a base stock system under a compound Poisson demand process. Besides having inventory and backorder costs, also a cost is incurred for each order that is delivered in several pieces to the customer (irrespective of whether some items in the order are delivered on time, or whether they are all delivered late but at different times). The possible split deliveries are assumed to happen automatically, meaning that there cannot at the same time be a positive on-hand inventory and a positive backlog. We develop a mathematical cost expression for this system to be optimized, denoted Model 1. We compare Model 1 to another model, denoted Model 2, where it is forbidden to make any of these split deliveries, thus at the same time, there can be a positive on-hand inventory and a positive backlog. We compare Model 2 to the standard textbook model, denoted Model 0, where split deliveries occur automatically and are costless. Thereby we address a claim made by Zipkin (2000). We also examine the threshold value for the delivery split cost, which makes Model 1 and 2 perform equally well, and we examine how this threshold value depends on the other common parameters of the models.

Keywords: Inventory control, base stock policy, compound Poisson process, split deliveries

1. Introduction

In most expositions on inventory control, it is tacitly assumed that if a customer request occurs and, say, the order is of size 6, and there is currently a positive amount on stock but less than the customer order, say, it is 4, then the customer is given the amount possible, in this case 4 units, and the remaining units, in this case 2, is sent to the customer later. What if, when this is not possible, that is, all deliveries from the inventory are only made when the customer order can be served in full? Or it is possible to make a split delivery, as described above, but it has an additional cost? Surely, whatever assumptions are made there are additional costs compared to the more generic textbook model assuming that split deliveries happen automatically and are costless. Furthermore, it might also be that the optimal control policies in any of these two variants will be different from what is prescribed in the generic textbook model. In the standard textbook literature (see for instance Zipkin (2000) and Axsäter (2006)), there is not much emphasis on this issue. Though it is mentioned in passing at p. 227 in Zipkin (2000) (without any references):

The second scenario (Model 2 in this paper) leads to a far more complex model, however, and only rarely to fundamentally different system behavior.

This paper will, as one of its research questions, address this quote and investigate what happens if we introduce these more complicating features into an inventory control model. We will do our investigation within the frame of base stock control policies (relevant, when there are no fixed order costs and thereby no order batching), and where the demand process is a compound Poisson process. Surely, if there is no compound element, the objective of this research project would be redundant. Also by investigating an inventory system with no cycle stocks due to order batching, it should be a scene for where to see the most substantial (if any) effects of these more detailed models. As in most expositions on mathematical inventory models, we assume that any unit on stock is incurred an inventory cost at rate h , and any unit backlogged is incurred a shortage cost at rate b . In the case where split deliveries are not possible, we assume that orders are served on a ‘First-Come-First-Served’ basis. This means, referring to the initial example, that, if immediately after the occurrence

of the order of size 6, another order of size 2 arrived, which could be immediately served, it is first served after the order of 6 has been served. In the case where split deliveries are possible, we assume a cost c per order that is split. The cost component c could be motivated by an administrative surcharge, since we have to inform the customer of the splitting of her order. This cost specification is consistent when measuring service performance as the ability to deliver complete customer orders opposed to focusing on the traditional fill rate. For more discussion and analysis of this alternative service measure, the order fill rate, see Larsen and Thorstenson (2008, 2014).

In Section 2, we present, in two separate subsections, our mathematical models. In the first subsection, we present Model 2, where it is not possible to make split deliveries. Here, we apply the unit tracking methodology of Axsäter (1990) to derive a mathematical expression for the expected costs incurred per time unit. In the second subsection, we develop a mathematical expression for the expected costs per time unit for the case, where split deliveries are possible, that is, Model 1. Here, we can start reusing standard cost expressions for the expected inventory and shortage costs per time unit, which is Model 0 of this paper. But in addition to that, we must develop a mathematical expression for the probability that an incoming order can be subject to a split delivery. Also, for this derivation, we rely on the unit tracking methodology of Axsäter (1990). Note that the split delivery can occur, as sketched in the example of the beginning of this section, when there is a positive amount on stock. But it could also occur in the case, where there is no on-hand inventory at the time of the customer request, and the first incoming replenishment cannot cover the customer order. We also show how this probability is related to the order fill rate service measure. In Section 3, we formulate two research questions and outline a numerical experiment. Then, in Section 4, we do our numerical investigation and comment on it with regard to the two research questions. Finally, Section 5 contains some concluding remarks and suggestions for future research.

2. Derivation of mathematical expressions

In all models, we assume a base stock policy with base stock level S . The demand process is a compound Poisson process with rate λ customer arrivals per time unit, and the size of a customer order is represented by a positive integer-valued random variable J . The base stock policy implies that a replenishment order, of the same size as the customer order, is issued immediately when receiving the customer request. It is assumed that all replenishment orders have a constant lead-time of L time units. We start deriving the results of Model 2, which can be considered a special case of Model 1, in the sense that the option of making split deliveries is not possible.

Model 2: No split deliveries

We will apply the unit tracking methodology of Axsäter (1990). This has been extended to the case of a compound Poisson process by Forsberg (1995) and Axsäter and Zhang (1996). So the exposition here relies on these sources, but it attempts to be made in a self-contained manner. In order to make the explanations easier to follow, a sample path of base stock system serving a compound demand process has been generated in Table 1.

<Table 1 about here>

The table shows the arrival of in all 5 orders with a total demand of 16 units, with $S = 4$ and $L = 2$. The initial inventory is equal to the base stock level $S = 4$, meaning, initially, there are no outstanding replenishment orders. The first 4 unit demands are served from the initial inventory. But all the other unit demands are, according to the logic of the base stock policy, served from the replenishment of

unit demand number S occurring before the considered unit demand. That is for example unit demand number 11, occurring at time 4.97, is served from the replenishment of unit demand number 7, which occurs at time 3.19. As we are concerned with the fulfillment of complete customer orders, we will rather focus on the tracking of replenishments to fulfill a complete customer order. As an example, let the order of size 4 arriving at time 3.19 play the role as the considered-order. It is served from the unit demands occurring as numbers 4, 3, 2, 1 before the arrival of the considered-order. As there, until the arrival of the considered-order, has been a total demand of 5, it means the considered-order is served from replenishments of unit demand numbers 2, 3, 4, 5. The first two of these unit demand numbers, unit demand numbers 2 and 3, belong to the order that arrived at time 1.01, that is, $3.19 - 1.01 = 2.18$ time units before the arrival of the considered-order. It means the replenishment of unit demand numbers 2 and 3 will arrive $2.18 - 2 = 0.18$ time units earlier than needed, which is why they are recorded with earliness value of 0.18 in the last column of Table 1. Unit demand numbers 4 and 5 belong to the order that arrived at time 2.08, that is, 1.11 time units before the arrival of the considered-order. It means the replenishment of unit demand numbers 4 and 5 will arrive at the inventory at time $2.08 + 2 = 4.08$ and thereby cause a delay of at least $4.08 - 3.19 = 0.89$ time units for the considered-order, and they are therefore given an earliness value of -0.89 in the last column of Table 1. The reason for the wording “at least” above, is due to the fact that the total delay of the considered-order is determined from the last incoming replenishment. Note that if the considered-order contains a unit demand number v with $v \geq S + 1$, then unit demand number v will be served from the replenishment of another unit demand in the considered-order, meaning it will first be served after L time units. This is illustrated in Table 1 for the order of size 5 arriving at time 5.23.

We now return to the general case and consider an order of size k , as above denoted the considered-order, arriving to the inventory system well ahead in time, so it is not served from the initial inventory, meaning we are in steady-state. It consists of unit demand number $v = 1, \dots, k$. As explained above, unit demand number v will be served from the replenishment of unit demand number $S + 1 - v$, occurring before the considered-order. Denote by the random variable $W(r)$ the time interval from when unit demand number r occurred before the considered-order. Per definition $W(r) \equiv 0$ when $r \leq 0$. The considered-order will first be delivered to the customer, in one piece, when the replenishment of unit demand number $S + 1 - k$, occurring before the considered-order, has arrived to the inventory. Therefore, the expected accumulated waiting time of the considered-order is

$$kE[L - W(S + 1 - k)]^+ \quad (1)$$

Here, x^+ is the usual notation for $\max\{0, x\}$. The items that are replenished in order to serve the considered-order incurs an expected accumulated sojourn time at the inventory of

$$kE[W(S + 1 - k) - L]^+ + \sum_{v=1}^{k-1} \{E[W(S + 1 - v)] - E[W(S + 1 - k)]\} \quad (2)$$

The first part in (2) accounts for the case, where the replenishment of unit demand number $S + 1 - k$ arrives to the inventory before the considered-order occurs, and the whole batch, of size k , of incoming replenishments is waiting to serve the considered-order. The second part of (2), which is positive in case $k \geq 2$, reflects the fact that the incoming replenishment do not arrive simultaneously, and $E[W(S + 1 - v)] - E[W(S + 1 - k)]$ is the expected time difference between when the replenishment of unit demand number $S + 1 - v$ arrives until the whole order, of size k , of incoming replenishments is collected.

As an order of size k arrives with rate $\lambda P(J=k)$ to the inventory system, we need to multiply this rate to the sum of (1) (with weight b) and (2) (with weight h) and then make a summation over k in order to specify the expected costs per time unit, which we denote $C_2(S)$. When using that $W(r) \equiv 0$ when $r \leq 0$ and that $E[L - W(r)]^+ = L - E[W(r)] + E[W(r) - L]^+$, we get

$$\begin{aligned} C_2(S) = & \lambda b L E[J] \\ & + (h + b) \lambda \sum_{k=1}^S k P(J = k) E[W(S + 1 - k) - L]^+ \\ & - b \lambda \sum_{k=1}^S k P(J = k) E[W(S + 1 - k)] \\ & + h \lambda \sum_{k=2}^S P(J = k) \sum_{v=1}^{k-1} \{E[W(S + 1 - v)] - E[W(S + 1 - k)]\} \\ & + h \lambda (1 - P(J \leq S)) \sum_{v=1}^S E[W(S + 1 - v)] \end{aligned} \quad (3)$$

Denote by the random variable $J(n)$ the sum of n independent random variables, all having the same probability distribution as J . We define $J(0) \equiv 0$ (because $P(J(0)=0) = 1$) and per definition $J(1) \equiv J$. Denote by the random variable $N(t)$ the number of customer arrivals during a time interval of length t . Per definition, $N(t)$ is Poisson distributed with mean λt . It follows from Forsberg (1995) and Axsäter and Zhang (1996) that (in order to make the paper appear self-contained proofs of this are given in Appendix)

$$E[W(r)] = \begin{cases} \frac{1}{\lambda} & r = 1 \\ E[W(r-1)] + \frac{1}{\lambda} \sum_{n=1}^{r-1} P(J(n) = r-1) & r \geq 2 \end{cases} \quad (4)$$

and

$$E[W(r) - L]^+ = \begin{cases} \frac{P(N(L) \leq 0)}{\lambda} & r = 1 \\ E[W(r-1) - L]^+ + \frac{1}{\lambda} \sum_{n=1}^{r-1} P(J(n) = r-1) P(N(L) \leq n) & r \geq 2 \end{cases} \quad (5)$$

Note as a point of validation that (5) converges to (4), when L converges to 0.

Model 1: Split deliveries with a cost per order that is split

If we ignore the costs of the split deliveries, then we have the standard textbook model, in this paper denoted Model 0, which has expected inventory and shortage costs per time unit (see for instance p. 130 in Axsäter (2006))

$$C_0(S) = bE[D(L) - S]^+ + hE[S - D(L)]^+ \quad (6)$$

Here, $D(L)$ is the aggregate demand recorded during a lead-time period L , and its probability distribution can be computed by the Adelson recursion, Adelson (1966). We will again resort to the unit tracking methodology when specifying the probability for that an in-incoming order of size k , (which we again will denote the considered-order) is delivered in full. We denote this probability $Q(k, S)$. There are two easy cases. First, if $k = 1$, then split delivery is not an issue, and accordingly $Q(k, S)$ is 1. Secondly, if $k \geq S + 1$, then the delivery of the considered-order will always be split. This is due to the fact that unit demand number 1 of the considered-order is replenished from a customer order that occurred prior to the considered-order, while unit demand number k of the considered-order is replenished from a unit demand belonging to the considered-order. Therefore, in this case $Q(k, S)$

$= 0$. So, it remains to investigate the case $2 \leq k \leq S$. With probability $P(D(L) \leq S - k)$, the considered-order will be served immediately and thus without any splitting. If the considered-order is delayed, but served without splitting, it can only happen if the on-hand inventory was zero upon arrival, and when unit demand number S and unit demand number $S + I - k$ that are occurring before the considered-order belong to the same customer order, which is either customer order number $m = 1, \dots, S + k - I$ occurring before the considered-order. Unit demand number S and unit demand number $S + I - k$ that are occurring before the considered-order both belong to customer order number m occurring before the considered-order with probability

$$\sum_{x=m-1}^{S-k} P(J(m-1) = x) P(J \geq S - x) \quad (7)$$

The expression above takes into account all combinations where the first $m - I$ orders occurring before the considered-order contains at most $S - k$ units in total while order number m occurring before the considered-order is so large that it brings the total amount of all the m orders at or above S . Due to the definition that $J(0) \equiv 0$, (7) also is valid for the case $k = S$ where the only possible m is I . We must then combine (7) with the probability that in all n ($n \geq m$) customers have arrived in the lead-time period. Therefore,

$$\sum_{n=1}^{\infty} P(N(L) = n) \sum_{m=1}^n \sum_{x=m-1}^{S-k} P(J(m-1) = x) P(J \geq S - x) \quad (8)$$

is the probability that the considered-order arrives to see an empty inventory but is served together because an incoming replenishment, destined to serve considered-order, is greater than or equal to the considered-order. When also adding the probability $P(D(L) \leq S - k)$ and keeping in mind that the counter m in (8) at most can be $S + I - k$, we get that $Q(k, S)$ can be specified as

$$\begin{aligned} Q(k, S) &= P(D(L) \leq S - k) \\ &+ \sum_{n=1}^{S-k} P(N(L) = n) \sum_{m=1}^n \sum_{x=m-1}^{S-k} P(J(m-1) = x) P(J \geq S - x) \\ &+ P(N(L) \geq S - k + 1) \sum_{m=1}^{S+1-k} \sum_{x=m-1}^{S-k} P(J(m-1) = x) P(J \geq S - x) \end{aligned} \quad (9)$$

After having specified $Q(k, S)$, we can now specify $SplFR(S)$, defined to be the expected fraction of orders that are delivered in several pieces. As the occurrence of an order of size k has probability $P(J=k)$, we get

$$SplFr(S) = \sum_{k=2}^S P(J = k)(1 - Q(k, S)) + P(J \geq S + 1) \quad (10)$$

The expected total costs per time period, which we denote $C_1(S)$, is (10) multiplied by λc and added to (6). That is,

$$C_1(S) = bE[D(L) - S]^+ + hE[S - D(L)]^+ + c\lambda SplFr(S) \quad (11)$$

From the expressions above, we can also get the measure for the order fill rate, $OFR(S)$, as it is defined in Larsen and Thorstenson (2008, 2014). Using the notation of this paper, it is

$$OFR(S) = \sum_{k=1}^S P(J = k) P(D(L) \leq S - k) \quad (12)$$

From (9), we can define another measure, the full order delivery rate, $FODR(S)$, as

$$FODR(S) = \sum_{k=1}^S P(J = k) Q(k, S) \quad (13)$$

Obviously, we have $FODR(S) \geq OFR(S)$ for any value of S . This $FODR(S)$ measure has some resemblance to the SCOR metric: *RL2.1 % of orders delivered in full*, see Supply Chain Council (2010).

3. Formulation of research questions and set-up of the numerical investigation

We will explore two research questions:

Research Question 1

Assume it is not possible to do a split delivery. How biased will the predictions of the textbook model be compared to the true model assuming no split? And how does this bias depend on the other parameters of the models?

Research Question 2

Assume that orders are automatically delivered in several pieces but at a cost per order for each order that is split. How large can this cost component be, before it is better not to allow split delivery at all? And how does this threshold value depend on the other parameters of the models?

By ‘the other parameters of the model’ in the research questions, we refer to the parameters λ , L , h , b as well as distributional assumptions of J . In order not vary all these parameters, which could make the analysis too overwhelming and possibly confusing, we keep the parameter L fixed at 1, we fix λ to be 2, we fix h to be 1, and let b be either 5 or 10. Finally, we let J be negative binomial distributed, that is, it has distribution:

$$P(J = k) = \begin{cases} (1-p)^r & k = 1 \\ P(J = k-1) \frac{p(r+k-2)}{k-1} & k = 2, 3, \dots \end{cases} \quad (14)$$

Here, we let the probability parameter p be either 0.25, 0.5 or 0.75 and let the shape parameter r attain the values 1, 2, 3, 4, 5 (when $r = 1$, it means J is geometrically distributed). The reason why we choose the negative binomial distribution, is that it is a two-parameter distribution (contrary to the logarithmic and the geometric distribution presented in Chapter 5 of Axsäter (2006)), and therefore a lot of different shapes of the probability distribution of J can be generated. When we vary r and p , we do not find it necessary to vary λ and L as well, because the variation of r and p secures various levels of the aggregate demand in the lead-time period, which is the most important random input to our models. The mean and standard-deviation of the aggregate demand in the lead-time period are, respectively,

$$E[D(L)] = \lambda L E[J] = \lambda L \frac{1+(r-1)p}{1-p} \quad (15)$$

and

$$\sigma[D(L)] = \sqrt{\lambda L E[J^2]} = \frac{1}{1-p} \sqrt{\lambda L (rp + (1 + (r-1)p)^2)} \quad (16)$$

In most base stock models, what matters is the ratio between the inventory costs and the shortage cost which determines (via a newsvendor equation) the optimal base stock level. Therefore by varying b and keeping h fixed we get two different ratios. All combinations of (b, r, p) are chosen such that $S = 0$ will never be an optimal solution in any models.

Research Question 1 is investigated by examining:

RQ1.1: C_2^* , the minimum value of (3) when minimizing (3) with respect to the base stock level S .

RQ1.2: The difference between the optimal base stock levels in respectively the no-split model and the textbook model, that is, $S_2^* - S_0^*$, where S_2^* is the optimal solution of the former (Model 2), while S_0^* is the optimal solution of the latter (Model 0).

RQ1.3: The percentage increase when evaluating the costs in the textbook model compared to the no-split model, that is, $\frac{C_2^* - C_0^*}{C_0^*}$, where C_0^* is the minimum value of (6) with respect to the optimal base-stock level.

RQ1.4: The percent decrease when switching from S_0^* to S_2^* and doing the cost evaluations in the true model of Model 2 by evaluating $\frac{C_2(S_0^*) - C_2^*}{C_2(S_0^*)}$

Research Question 2 concerns how to compute the threshold value. At first glance, it might seem as a very time consuming work (or requiring additional programming development) to figure out the threshold value of c , which furthermore must be done for any combination of (b, r, p) . The point is, though, that this threshold value can be deduced immediately just by running Model 2 and evaluating (6) and (10) for each value of S .

The threshold value of c , c_{thr} , is now the minimum value of c that secures

$$C_0(S) + \lambda c * SplFr(S) \geq C_2^* \quad \forall S \quad (17)$$

Or, more operationally,

$$c_{thr} = \max \left\{ \frac{C_2^* - C_0(S)}{\lambda * SplFr(S)} : S > 0 \right\} \quad (18)$$

4. Numerical results

Tables 2 - 6 contains an examination of both research questions. For each combination of $(b, r, p, E[D(L)], \sigma[D(L)])$ is stated five entries. The first four of these address the answer of Research Question 1, by answering the four sub-questions: RQ1.1, RQ1.2, RQ1.3 and RQ1.4 stated in the previous section, while the last entry computes the threshold value (18), thereby answering Research Question 2.

<Tables 2-6 about here>

Concerning the analysis of Research Question 1, we see a cost increase in the magnitude of 10% to 30% when comparing the textbook model, Model 0, to Model 2. This must be considered as a quite substantial cost increase. This cost increase seems to be independent of whether there is high frequent demand (in our case an expected lead-time demand about 30) or low frequent demand (in our case an expected lead-time demand about 2). Furthermore, unless p is low (0.25), the textbook model will not

predict the optimal base stock level either, but it will predict a base stock level that is too small compared to what will be the outcome when using Model 2. However, by examining the fourth entry, RQ1.4, in Tables 2-6, we see that the economic consequences of using a base stock level that is too low is considerably less compared to when doing the direct comparison between Models 0 and 1 (the third entry, RQ1.3), and it is less than 6% for all the cases. So this leads to a mixed conclusion whether we by this analysis can claim that we have raised question marks to the claim made by Zipkin (2000), which we quoted in Section 1. The base stock levels that we can compute from the textbook model are not completely different from the optimal ones. On the other hand, the cost predictions you get from the textbook model are not correct as they clearly under-estimates the true costs, and in order to make a correct cost assessment, you need to develop a more accurate model, as we have done with Model 2. Concerning the analysis of Research Question 2, it seems, for the same values of r and p , that the higher b is, the higher gets the threshold value. So, in surroundings with focus on avoiding shortages it becomes more favorable to use Model 1 than Model 2. Also, when the parameter p increases, meaning the individual customer orders possess more variance, the threshold value increases, meaning Model 1 is more competitive to Model 2.

<Table 7 about here>

A further generalization of the model could concern that there is a cost per additional delivery, and therefore the number of splits of an order should be measured. In order to investigate this issue, a simulation model, programmed in Arena, see Kelton et al (2007), has been developed. For the cases in Table 2 (with $r = 1$) and using the base stock level S_2^* from Table 2, we have collected the number of splits of those orders that are not delivered in one piece. The results can be seen in Table 7. For the case $p = 0.25$ we can interpret the results so that roughly 1 out of 5 orders that are delivered in several pieces are delivered in three pieces, while 4 out of those 5 orders are delivered in two pieces giving the recorded average value of about 1.2. Similarly, the case $p = 0.75$ can be interpreted so that roughly 1 out of in all 3 orders that are delivered in several pieces are delivered in three pieces, while 2 out of those 3 orders are delivered in two pieces giving the recorded average value of about 1.5. Of course it depends on what the real cost structure is. It may be reasonable to argue, as we did in Section 1, that there is a cost c per order split, and then Model 1 is accurate. But if it is more reasonable to assume the same cost c is per each additional delivery, then the real threshold value is lower than the threshold values reported in Table 2. However, the computed threshold value by (18), using an incorrect model, still has a purpose as it then provides an upper bound. We do believe it will be quite difficult to develop a mathematical expression for in how many pieces an order is delivered. So any analysis concerning a cost per additional delivery must probably rely on simulation analyses.

5. Concluding remarks

We have developed two models to complement the textbook model for a base stock system under a compound Poisson demand process. The focus has been on that the textbook model might be too crude about that it is possible and costless to make delivery splits. For the case of the same cost structure as in the textbook model, but where it is forbidden to make split deliveries, our numerical experiment indicates that there can be a significant difference in the prediction of the cost performance. With regard to the recommendations about what is the optimal base stock level, the textbook model will recommend a base stock level that is too low compared to what the true model will recommend. However, the cost consequences of this in-optimal behavior are less dramatic. In all fairness, it must be mentioned that using a base stock policy probably makes these differences most significant. If we also include order costs and thereby a cycle stock, then, most of the time there will

be an ample inventory and shortage situations therefore occur rather seldom, so these differences will tend to vanish. However, many inventories, like spare parts inventories, are operated as base stock control systems, so findings from our Research Question 1 can be of practical relevance for these settings. The second research question addresses the situation when there is a cost related to a split delivery. In our paper you can either forbid this or you can assume that you do a split whenever it happens. Based on these premises, we have developed a method to find a threshold value of the split cost that makes the models break even. Of course, one could argue that the choice of no split or split (at a cost) should be done in a more flexible way, such that you dynamically over time consider how to handle each specific customer order. This would lead to the formulation of a semi-Markov decision model which per definition will perform better than the two models in our paper. The development and analysis of such a semi-Markov decision model could be a subject for further research. Furthermore, in such a model, there are no restrictions on which control policy to apply, so the conclusions from the analysis could be that the optimal control policy is different from the base stock policy. A reason for this could be that doing some degree of batch ordering (although the replenishment order costs are zero) is optimal as it would eliminate the occurrences of split deliveries, as we commented above, and therefore be optimal. It would ultimately also more fundamentally challenge the quote of Zipkin (2000) about ‘only rarely to fundamentally different system behavior’.

References

- Adelson RM (1966) Compound Poisson distributions. *Op Res Quart* 17: 73-75
- Axsäter S (1990) Simple solution procedures for a class of two-echelon inventory problems. *Op Res* 38: 64-69
- Axsäter S (2006) *Inventory control* (2'nd ed.). Springer, New York, USA
- Axsäter S, Zhang W-F (1996) Recursive evaluations of order-up-to-S policies for two-echelon inventory systems with compound Poisson demand. *Nav Res Log* 43: 151-157
- Forsberg R (1995) Optimization of order-up-to-S policies for two-level inventory systems with compound Poisson demand. *Eur J Oper Res* 81: 143-153
- Kelton W, Sadowski R, Sturrock D (2007) *Simulation with Arena* (4th ed.). McGraw Hill, Boston, USA
- Larsen C, Thorstenson A (2008) A comparison between the order and the volume fill rate for a base-stock inventory control system under a compound renewal demand process. *J Oper Res Soc* 59: 798-804
- Larsen C, Thorstenson A (2014) The order and volume fill rates in inventory control systems. *Int J Prod Econ* 147: 13-19
- Supply Chain Council (2010) *SCOR Overview*, Version 10.
<http://cloud.ld.ttu.ee/idu0010/Portals/0/Harjutustunnid/SCOR10.pdf>. Accessed June 20, 2016
- Zipkin P (2000) *Foundations of inventory management*. McGraw-Hill, Boston, USA

Tables

Arrival time	Order size	Unit demand number of the order	Unit demand number in total	Served from the replenishment of unit demand number (or initial inventory)	W-observation	Earliness value
<i>1.01</i>	<i>3</i>	<i>1</i> <i>2</i> <i>3</i>	<i>1</i> <i>2</i> <i>3</i>	In. inv. In. inv. In. inv.		
<i>2.08</i>	<i>2</i>	<i>1</i> <i>2</i>	<i>4</i> <i>5</i>	In. inv. <i>1</i>	<i>1.07</i>	<i>-0.93</i>
<i>3.19</i>	<i>4</i>	<i>1</i> <i>2</i> <i>3</i> <i>4</i>	<i>6</i> <i>7</i> <i>8</i> <i>9</i>	<i>2</i> <i>3</i> <i>4</i> <i>5</i>	<i>2.18</i> <i>2.18</i> <i>1.11</i> <i>1.11</i>	<i>0.18</i> <i>0.18</i> <i>-0.89</i> <i>-0.89</i>
<i>4.97</i>	<i>2</i>	<i>1</i> <i>2</i>	<i>10</i> <i>11</i>	<i>6</i> <i>7</i>	<i>1.78</i> <i>1.78</i>	<i>-0.22</i> <i>-0.22</i>
<i>5.23</i>	<i>5</i>	<i>1</i> <i>2</i> <i>3</i> <i>4</i> <i>5</i>	<i>12</i> <i>13</i> <i>14</i> <i>15</i> <i>16</i>	<i>8</i> <i>9</i> <i>10</i> <i>11</i> <i>12</i>	<i>2.04</i> <i>2.04</i> <i>0.26</i> <i>0.26</i> <i>0.00</i>	<i>0.04</i> <i>0.04</i> <i>-1.74</i> <i>-1.74</i> <i>-2.00</i>

Table 1: Sample path of base stock system with base stock level $S = 4$ and lead-time $L = 2$. The first three columns contain the information of the respective considered-orders. The first column keeps track of which unit demand numbers whose replenishments are designated to serve the considered-order. The column with W observation contains actual realizations of the random variables $W()$ referred to in Section 2. In the last column, is computed the earliness value of each incoming replenishment designated to serve the considered-order. If positive, it means they arrive too early with regard to the arrival time of the considered-order, and if negative, they are causing a delay for serving the considered-order.

p	$E[D(L)]$	$\sigma[D(L)]$	$b = 5$		$b=10$	
0.25	2.07	2.11	4.04	(RQ1.1)	5.03	(RQ1.1)
			$5 - 5 = 0$	(RQ1.2)	$6 - 6 = 0$	(RQ1.2)
			12.16%	(RQ1.3)	10.47%	(RQ1.3)
			0.00%	(RQ1.4)	0.00%	(RQ1.4)
			3.02	(c_{thr})	5.21	(c_{thr})
0.5	4.00	3.46	7.41	(RQ1.1)	9.17	(RQ1.1)
			$8 - 7 = 1$	(RQ1.2)	$10 - 9 = 1$	(RQ1.2)
			23.29%	(RQ1.3)	19.06%	(RQ1.3)
			3.79%	(RQ1.4)	3.22%	(RQ1.4)
			6.06	(c_{thr})	10.67	(c_{thr})
0.75	8.00	7.48	17.17	(RQ1.1)	21.20	(RQ1.1)
			$18 - 15 = 3$	(RQ1.2)	$23 - 19 = 4$	(RQ1.2)
			30.34%	(RQ1.3)	24.73%	(RQ1.3)
			4.38%	(RQ1.4)	5.20%	(RQ1.4)
			14.92	(c_{thr})	26.73	(c_{thr})

Table 2: Results on the numerical examination of Research Questions 1 and 2 with shape parameter $r = 1$. In column 1 is stated the demand inputs p , $E[D(L)]$ and $\sigma[D(L)]$. For the inputs $b = 5$ and $b = 10$ in column 2 and 3, in each cell, we first list the four answers to the questions RQ1.1, RQ1.2, RQ1.3 and RQ1.4, stated in Section 3, which constitutes the answer of Research Question 1. Thereafter comes the computation of the threshold value of c_{thr} using (18), which constitutes the answer of Research Question 2.

p	$E[D(L)]$	$\sigma[D(L)]$	$b = 5$		$b=10$	
0.25	3.33	2.71	5.41	(RQ1.1)	6.64	(RQ1.1)
			$7 - 6 = 1$	(RQ1.2)	$8 - 7 = 1$	(RQ1.2)
			17.25%	(RQ1.3)	13.25%	(RQ1.3)
			0.48%	(RQ1.4)	3.79%	(RQ1.4)
			3.41	(c_{thr})	5.79	(c_{thr})
0.5	6.00	5.10	10.80	(RQ1.1)	13.24	(RQ1.1)
			$13 - 11 = 2$	(RQ1.2)	$15 - 13 = 2$	(RQ1.2)
			22.81%	(RQ1.3)	18.28%	(RQ1.3)
			2.68%	(RQ1.4)	5.26%	(RQ1.4)
			6.65	(c_{thr})	11.52	(c_{thr})
0.75	14.00	12.08	26.43	(RQ1.1)	32.23	(RQ1.1)
			$31 - 25 = 6$	(RQ1.2)	$37 - 31 = 6$	(RQ1.2)
			26.43%	(RQ1.3)	21.11%	(RQ1.3)
			5.47%	(RQ1.4)	6.00%	(RQ1.4)
			15.88	(c_{thr})	27.27	(c_{thr})

Table 3: Results on the numerical examination of Research Questions 1 and 2 with shape parameter $r = 2$. For further comments, see Table 2.

p	$E[D(L)]$	$\sigma[D(L)]$	$b = 5$		$b=10$	
0.25	4.00	3.27	6.56	(RQ1.1)	8.09	(RQ1.1)
			8 - 7 = 1	(RQ1.2)	10 - 9 = 1	(RQ1.2)
			17.81%	(RQ1.3)	14.34%	(RQ1.3)
			2.71%	(RQ1.4)	0.85%	(RQ1.4)
			3.74	(c_{thr})	6.26	(c_{thr})
0.5	8.00	6.63	13.91	(RQ1.1)	19.67	(RQ1.1)
			17 - 14 = 3	(RQ1.2)	20 - 18 = 2	(RQ1.2)
			22.37%	(RQ1.3)	17.74%	(RQ1.3)
			4.56%	(RQ1.4)	2.58%	(RQ1.4)
			7.35	(c_{thr})	12.51	(c_{thr})
0.75	20.00	16.49	35.19	(RQ1.1)	42.76	(RQ1.1)
			42 - 35 = 7	(RQ1.2)	51 - 44 = 7	(RQ1.2)
			24.53%	(RQ1.3)	19.64%	(RQ1.3)
			5.45%	(RQ1.4)	4.34%	(RQ1.4)
			17.55	(c_{thr})	29.94	(c_{thr})

Table 4: Results on the numerical examination of Research Questions 1 and 2 with shape parameter $r = 3$. For further comments, see Table 2.

p	$E[D(L)]$	$\sigma[D(L)]$	$b = 5$		$b=10$	
0.25	4.67	3.80	7.69	(RQ1.1)	9.42	(RQ1.1)
			9 - 8 = 1	(RQ1.2)	11 - 10 = 1	(RQ1.2)
			18.70%	(RQ1.3)	15.02%	(RQ1.3)
			3.66%	(RQ1.4)	3.12%	(RQ1.4)
			4.13	(c_{thr})	6.95	(c_{thr})
0.5	10.00	8.12	16.91	(RQ1.1)	20.58	(RQ1.1)
			21-18 = 3	(RQ1.2)	25-22 = 3	(RQ1.2)
			21.97%	(RQ1.3)	17.56%	(RQ1.3)
			2.98%	(RQ1.4)	3.04%	(RQ1.4)
			8.10	(c_{thr})	13.65	(c_{thr})
0.75	26.00	20.83	43.80	(RQ1.1)	53.11	(RQ1.1)
			54-46 = 8	(RQ1.2)	64-56 = 8	(RQ1.2)
			23.52%	(RQ1.3)	18.81%	(RQ1.3)
			4.16%	(RQ1.4)	4.36%	(RQ1.4)
			19.58	(c_{thr})	33.13	(c_{thr})

Table 5: Results on the numerical examination of Research Questions 1 and 2 with shape parameter $r = 4$. For further comments, see Table 2.

p	$E[D(L)]$	$\sigma[D(L)]$	$b = 5$		$b=10$	
0.25	5.33	4.32	8.75	(RQ1.1)	10.70	(RQ1.1)
			11 - 9 = 2	(RQ1.2)	13 - 12 = 1	(RQ1.2)
			18.71%	(RQ1.3)	14.79%	(RQ1.3)
			4.78%	(RQ1.4)	1.22%	(RQ1.4)
			4.31	(C _{thr})	7.33	(C _{thr})
0.5	12.00	9.59	19.86	(RQ1.1)	24.12	(RQ1.1)
			25 - 21 = 4	(RQ1.2)	29 - 26 = 3	(RQ1.2)
			21.79%	(RQ1.3)	17.37%	(RQ1.3)
			3.99%	(RQ1.4)	3.31%	(RQ1.4)
			8.87	(C _{thr})	14.94	(C _{thr})
0.75	32.00	25.14	52.35	(RQ1.1)	63.38	(RQ1.1)
			65 - 56 = 9	(RQ1.2)	78 - 68 = 10	(RQ1.2)
			22.88%	(RQ1.3)	18.27%	(RQ1.3)
			4.20%	(RQ1.4)	4.34%	(RQ1.4)
			21.73	(C _{thr})	36.62	(C _{thr})

Table 6: Results on the numerical examination of Research Questions 1 and 2 with shape parameter $r = 5$. For further comments, see Table 2.

r	p	$b = 5$	$b=10$
1	0.25	1.21	1.20
	0.5	1.36	1.33
	0.75	1.46	1.42

Table 7: Results for simulation model to collect the expected number of deliveries per order, given that a split occurs. Each simulation runs over 200000 time units with a starting condition that the inventory level equals the base stock level. Each simulation is replicated 10 times. In columns 3 and 4 is stated the estimated number of splits of the orders not delivered in one piece. These are estimated in Arena with a 95% confidence interval of 0.00.

Appendix

Denote by P_r^n the probability that demand unit number r occurring before the considered order is contained in customer order number n occurring before the considered order. It holds, see also Axsäter and Zhang (1996), that (the probability is zero for all other combinations of (r, n) than listed below)

$$P_r^n = \begin{cases} P(J \geq r) & n = 1; r = 1, 2, \dots \\ P(J(n) \geq r) - P(J(n-1) \geq r) & n = 2, 3, \dots; r = n, n+1, \dots \end{cases} \quad (\text{A1})$$

For any $n \geq I$ denote by the random variable $T(n)$ the time interval from the arrival of the considered-order until order number n , occurring before the considered-order, arrived. It is an n -phased Erlang distribution with mean n/λ . The cumulative distribution function of $W(r)$ is for any $r \geq I$ given as (see also Axsäter and Zhang (1996))

$$P(W(r) \leq t) = \sum_{n=1}^r P_r^n P(T(n) \leq t) \quad (\text{A2})$$

Therefore,

$$E[W(r)] = \frac{1}{\lambda} \sum_{n=1}^r P_r^n n \quad (\text{A3})$$

We have

$$E[W(1)] = \frac{1}{\lambda} \quad (\text{A4})$$

For $r > I$ we have

$$\begin{aligned} E[W(r)] &= \frac{1}{\lambda} [\sum_{n=1}^{r-1} \{P(J(n) \geq r-1) - P(J(n-1) \geq r-1) - P(J(n) = r) + P(J(n-1) = r)\}n + \\ &\quad rP(J(n-1) = r-1)] \end{aligned} \quad (\text{A5})$$

By collecting terms (A5) can be rewritten as the lower half of (4).

From (A2), we also get that

$$E[W(r) - L]^+ = \sum_{n=1}^r P_r^n E[T(n) - L]^+ \quad (\text{A6})$$

The loss function of the Erlang distribution is, see also Axsäter and Zhang (1996)

$$E[T(n) - L]^+ = \frac{e^{-\lambda L}}{\lambda} \sum_{x=0}^{n-1} \frac{n-x}{x!} (\lambda L)^x \quad (\text{A7})$$

In the case $n = I$, we have from (A7)

$$E[T(1) - L]^+ = \frac{e^{-\lambda L}}{\lambda} = \frac{1}{\lambda} P(N(L) \leq 0) \quad (\text{A8})$$

When $n > I$, we get from (A7)

$$E[T(n) - L]^+ = E[T(n-1) - L]^+ + \frac{1}{\lambda} P(N(L) \leq n-1) \quad (\text{A9})$$

It holds that

$$E[W(1) - L]^+ = E[T(1) - L]^+ = \frac{1}{\lambda} P(N(L) \leq 0) \quad (\text{A10})$$

When $r > 1$, we can similarly, as for derivation in (A5), rewrite $E[W(r) - L]^+$ to

$$E[W(r) - L]^+ = \frac{1}{\lambda} [\sum_{n=1}^{r-1} \{P(J(n) \geq r-1) - P(J(n-1) \geq r-1) - P(J(n) = r) + P(J(n-1) = r)\} E[T(n) - L]^+ + E[T(r) - L]^+ P(J(n-1) = r-1)] \quad (\text{A11})$$

By collecting terms and using (A10), we can rewrite (A11) to the lower half of (5).