ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Econometrie & Operationele Research

Cost approximation for an (S-1, S) inventory model with rationing and demand lead times

Name student: Derk Barth
Student ID number: 476341

Supervisor: Dr. Oguzhan Vicil
Second Assessor: Professor Rommert Dekker

Date final version: July 3rd, 2022

Cost approximation for an (S-1, S) inventory model with rationing and demand lead times

We study an inventory policy regarding spare parts. In this policy, we need to satisfy the orders of two demand classes. The orders that belong to the first class need to be fulfilled straight away, while the orders of the second class should be satisfied in given demand lead time. Also, we have two different types of criticality. For the critical class, a higher service level, for which we will use fill rate in this case, needs to be maintained than for the non-critical class. To manage the inventory, we investigate a (S-1, S) policy with backordering. Assuming Poisson demand arrivals for both classes, we will first replicate the approximation and simulation for the service level proposed by Koçağa and Şen (2007). We will then create expressions that enable us to calculate holding and penalty costs. We compare these approximations with simulated values and introduce a model to minimize the stock levels using this cost function. Finally, we compare the optimal cost level indicated by our derived approximation with the simulated optimal cost to determine the accuracy of our cost approximation.

1. Introduction

For our paper, we want to replicate and extend on the paper of Koçağa and Şen (2007). In their paper, a new inventory management policy is suggested. Considering demand lead times, they propose a rationing strategy to reduce the stock level to be maintained. They assume two demand classes. The demand of the first class must be satisfied immediately, while the demand of the second class needs to be fulfilled in a given demand lead time. Besides that, they also introduce two different types of criticality. Critical orders should be satisfied whenever there is stock available, while non-critical orders must be rationed, meaning that they will only be satisfied when the amount of stock is above a critical stock level. For the demand, they assume Poisson demand arrivals. In their paper, they follow a one-for-one replenishment policy (S-1, S). With these given assumptions, and a set service level for each criticality class, which is defined as the fill rate, they examine what savings can be made in the stock levels. We want to extend by considering penalty and holding cost. In Deshpande et al. (2003) spare parts inventory management with rationing for different demand classes is considered. They involve several types of costs regarding the holding, setup and backlogging of orders. We will examine a combination of inventory management with demand lead time and rationing for an (S-1, S) ordering policy with a cost function, consisting of a combined penalty cost, and holding cost. First, we will replicate the comparison of the simulation described by Koçağa and Şen (2007) with the approximation of the service level they provide. After that, we will approximate a total annual cost function and compare this with the simulation results. Finally, we find new stock levels while optimizing cost determined by the approximation and find the cost difference between the optimal stock levels determined by the approximation and the actual optimal stock levels found by the simulation.

In this paper, we will first continue with explaining the relevance and the motivation of this research in the second section. In the third section, we perform a brief literature review, where we discuss existing papers regarding our subject. In section 4 we will discuss the theoretical background and the research methodology. We briefly discuss the service level calculations proposed by Koçağa and Şen (2007). After that, we derive an expression for calculating the penalty costs for the critical and non-critical customers. In section 5, we will discuss the results of our replication of the study by Koçaǵa and Şen (2007). After that, we will test our penalty cost calculations against a simulation to validate the accuracy of our calculations. We conclude our paper in section 6.

2. Relevance

Regarding the relevance of this matter, this research is focused on a combination of two different aspects: the rationing of inventory using advanced demand information and setting up an annual cost function. Already separate research has been done into both aspects. By combining these two aspects, the result of our research could be even more applicable in the corporate field. The currently existing research is either only focused on rationing with costs with no demand lead time or focused on rationing with demand lead time without considering a separate cost function. Companies that have to deal with the distribution of spare parts of production machines, big transportation vehicles or any other objects, in reality do not only consider minimizing the stock level, or only deal with customers that need their spare part straight away. Also, situations where spare parts are needed in the future occur, and costs come into play in pretty much every event. Common costs in inventory management often consist of order costs, storage costs, handling costs, stockout costs, etc. Also, getting orders with a demand lead time would be a realistic possibility, since companies that need spare parts often do routine maintenance. In this case, the demand for such a certain spare part would most likely be ordered before it is actually needed. If we consider both of these occasions, we suspect that combining annual cost with advance demand information, could very well complement on situations that occur in reality. Also, for scientific relevance, it could widen the current field of research. Our research can again be extended on, for example by incorporating multiple demand classes. This would contribute to keep on extending the research of this general matter.

3. Literature review

For our literature review, we will first review general literature on inventory systems with demand lead time. We will then append rationing inventory based on criticality. Finally, we will review literature on minimizing cost for a rationing policy.
One of the first research papers about demand lead time was the Simpson (1958) paper. In this research, service time for inventory in multi-stage production systems was examined. This was basically the first concept of a DLT. Hariharan & Zipkin (1995) then introduced the idea of customers that provide an advance warning of their demands. This introduced the idea of demand lead times bigger than zero. What actually both of these papers conclude, is that DLT reduces the inventory needed while still achieving the same service level, and has the same effect as the opposite of supply lead time. This is one of the factors that we also account for in our model. However, we have a complicating factor. This is the existence of the two different service level classes in our research. Kleijn & Decker (1999) explored an inventory system with several demand classes. Their first example implies a single-echelon system. In their model, they do consider more than two different demand classes. However, they are one of the first to introduce multiple demand classes in rationing inventory.

Another interesting piece of literature is Wang *et al.* (2002). In their work, they also studied a single-location system. Expressions for the inventory level distribution and the random customer delay were derived. They found that, if the probability that a replenishment order corresponding to a positive demand lead time arrives before the demand due date, the service level with demand lead times is higher than the service level for zero demand lead times. They extend their model to a two-echelon system after deriving the steady-state performance metrics for a single-location system. After analyzing this two-echelon results, it turned out that a system with two classes with different service levels results in quite significant inventory cost savings.

The first implementation of rationing was presented in Veinott (1965). This paper considered several demand classes in inventory systems. Analyzed was an inventory model with zero lead time and n demand classes, as well as backlogging of unfilled orders. He also introduced the critical-level policy, which will be used in our paper as well. Another segment incorporates in his report are holding and penalty costs. His goal was to find an ordering policy to minimize these costs over an infinite time horizon. Ha (1997) considers an inventory rationing system with Poisson demand processes and lost sales. He examines optimal critical level policies, and with that comes up with time-independent stock and critical stock levels. De Véricourt et al. (2002) studies a similar model, but assume backordering is possible.

To continue with costs, we consider Deshpande *et al.* (2003), which examined a rationing policy for two different demand classes and shortage penalty costs. They assumed Poisson demand arrivals in a continuous-review (Q, r) environment. In their model, they assume that there can be more than one order outstanding, which complicates the allocation of arriving replenishment orders. To overcome this issue, they compose a threshold clearing mechanism and created an accompanying algorithm to compute the optimal policy parameters (Q, r, K), with K as the threshold level. As a part of their model, they define an expected annual cost function for a given (Q, r, K) policy. This function consists of expected annual holding, setup and penalty costs. Our cost function will be very similar, apart from that we don't include the holding costs, since we assume a fixed (S-1, S) policy.

Tan et al. (2009) investigates the consequences of advance demand information that is not complete. They do this in an inventory system with rationing and two priority classes. Their main goal is to analyze this system with the objective of minimizing the expected total costs, with the assumption that all unsatisfied demand is lost. Kranenburg & Houtum (2005) consider cost optimization for an (S-1, S) inventory model with lost sales. Also, they assume more than two demand classes. Vicil (2022) also looks into cost optimization specifically for a (S-1, S) ordering policy, but with backorders. He considers two demand classes and rationing. He considers supply lead times that can be either deterministic or stochastic.

4. Methodology

To discuss our methodology, we first have to briefly discuss the theoretical background. Here, we will state the framework, definitions and theories we will use in our research. We will first start by setting our theoretical assumptions. Since we are extending the Koçaǵa and Şen (2007) paper, our assumptions will mostly match theirs. However, we will need additional

assumptions.
In the basis of our research, we will investigate fill rates for a spare parts inventory management, where we will also incorporate annual cost. We will consider two different demand classes. The orders of the first demand class have to be satisfied immediately. For the second demand class orders, a demand lead time is given in which the orders should be fulfilled. Also, we have two different classes of criticality. We have a critical class and a non-critical class. For the critical class, the service level requirement, which we set as the fill rate, is higher than for the non-critical class. We will look at two different cases:

i. the first demand class consists of critical orders, the second demand class of non-critical orders
ii. The first demand class consists of non-critical orders, the second demand class of critical orders

We will model a single-echelon system, with Poisson demand for the two defined demand classes. We note these rates as λ_1 and λ_2. For this inventory management, we will study a (S-1, S) ordering policy, with S being the order-up-to level. For each spare part demand event, we order a new part with a deterministic replenishment lead time L. To calculate the service level, we will use the fill rate for each demand class. This is defined as the number of orders that can be satisfied at the needed demand lead time corresponding to the class, divided by the total placed orders by that class. Also, we assume that every order gets accepted. If not sufficient stock is available at the time of the demand event, the order will be placed in a queue, operating according to the FIFO strategy. This means that demand of the non-critical class will be backlogged when the stock level falls below the critical stock level, while demand of the critical class keeps being filled until no inventory is left.
Furthermore, we also want to incorporate a cost function into our model. To do this, we will consider the following assumptions, also used in Deshpande et al. (2003). We assume a stockout cost per unit that has to be backordered π_c and π_n. We also assume a delay cost per period of delay per unit, $\hat{\pi}_c$ and $\hat{\pi}_n$. Proceeding this, we assume that $\pi_c > \pi_n$ and $\hat{\pi}_c > \hat{\pi}_n$, since the critical class is more important, which is also reflected in the service levels. The last cost we consider, is a holding cost per unit of stock per time period, h. We create a total expected cost function C(S-1, S), which consists of three parts in our case: $\Pi(S-1, S)$, $\widehat{\Pi}(S-1, S)$ and $H(S-1, S)$, which are the total stockout cost, delay cost and holding cost respectively. These assumptions are similar to the assumptions made by *Deshpande et al. (2003).* In our research, we won't consider setup cost, since we assume a (S-1, S) policy, and do not compare different ordering policies. We want to minimize these costs, while still meeting fixed service levels. First, we will discuss the theorem needed for the replication study of Koçaǵa and Şen (2007) in subsection 4.1. In section 4.2 we will derive our own cost function approximation, using expressions from section 4.1 and Deshpande et al. (2003). We will finish with section 4.3, in which we state and discuss the cost optimization problem.

4.1 replication study expressions

In this section, we will discuss the service level expressions derived by Koçaǵa and Şen (2007). For both the critical and the non-critical class, a service level function is derived for either the critical class or the non-critical class having DLT = T. Furthermore, the same assumptions we described above in the methodology are used, without considering the cost function assumptions. We will briefly explain the logic behind these expressions, and we will later use these functions in our replication study, as well as incorporate them into our annual stockout cost.

We will start with the service level expression for the non-critical class. To get to this expression, we consider an interval (t, t + L]. If we assume that no demand occurs in the interval, the inventory would be S at time t + L, since the delivery of orders at time t would arrive at time t + L. Next, for an order to be satisfied at time t + L, the inventory level at that time has to be at least S_c+1. This would only occur if the sum of the class 1 demand during time (t, t + L] and the class 2 demand that is due in (t + t, t + L] is less than S - S_c. Note that for DLT=T, we do not need to take into account demands that are due in (t, t + T], since the replenishments for these demands would be received before t + L (assuming that T < L). Therefore, these demands do not affect the inventory level at t + L. Combining both DLT possibilities, we get the following service level for the non-critical demand class, as derived by Koçaǵa and Şen (2007):

$$\beta_j^n(S, S_c) = P\{D_1(t, t+L] + D_2(t+T, t+L] \leq S - S_c - 1\} \quad (1)$$

To be able to calculate this, the probability function for the Poisson distribution is used. We note this as $p(i; \lambda) = e^{-\lambda}\lambda^i/i!$. If we combine expression (1) with the probability function, and using the sums of independent Poisson-distributed random variables, Koçaǵa and Şen (2007) come to the following service level expression for the non-critical demand class:

$$\beta_j^n(S, S_c) = \sum_{i=0}^{S-S_c-1} p(i; \lambda_1 L + \lambda_2(L-T)) \quad (2)$$

For the critical demand class, we again consider the interval (t, t+L]. If we assume that no demand occurs in the interval, the inventory would be S at time t + L, since the delivery of orders at time t would arrive at time t + L. Next, for an order to be satisfied at time t + L, the inventory level at that time has to be at least one unit. If we consider demand class 2, note that if the due date is between (t, t+T], the corresponding replenishment will arrive in time interval (t+L-T, t+L]. To calculate the probability that the inventory level at time t+L is at least one, we use hitting time H, which is the arrival of S-S_c units of total demand that has a negative influence on the inventory. We condition on whether H is in one of two intervals, which are (t, t+L-T] and (t+L-T, t+L], or after time t+L. This gives us the following expression:

$$\beta_j^c(S, S_c) = P\{D_j(t+H, t+L] \leq S_c - 1, H \leq L-T\}$$
$$+P\{D_j(t+H, t+L] \leq S_c - 1, L-T \leq H \leq L\} + P\{H \geq L\} \quad (3)$$

The next step in computing the service level would be to derive the density functions of the hitting time H for all three scenarios. We will not reproduce these derivations, but these derivations can be found in Koçaǵa and Şen (2007). We jump straight to the combination of these density functions, which together form the service level approximation for the critical demand class. This leads to the following expression:

$$\beta_j^c(S,S_c) = \int_0^{L-T} (\lambda_1+\lambda_2)^{S-S_c} e^{-(\lambda_1+\lambda_2)y} \frac{y^{S-S_c-1}}{(S-S_c-1)!} \times \left(\sum_{i=0}^{S_c-1} p\left(i:\lambda_j(L-y)\right)\right) dy + \int_{L-T}^{T} \lambda_1 e^{-(\lambda_1 y+\lambda_2(L-T))} \frac{[\lambda_1 y+\lambda_2(L-T)]^{S-S_c-1}}{(S-S_c-1)!} \times \left(\sum_{i=0}^{S_c-1} p\left(i:\lambda_j(L-y)\right)\right) dy + \sum_{i=0}^{S-S_c-1} p(i:\lambda_1 L+\lambda_2(L-T)) \quad (4)$$

Note that the expression for the critical service level approximates the actual service level, since it is not known how incoming replenishment orders are handled after the hitting time.

The last part that we need to fully perform our replication study, is the optimization problem for the stock level. Koçaǵa and Şen (2007) optimize the stock according to the following minimization problem:

$$\min_{S,S_c} S, \quad (5)$$

$$\text{subject to } \beta_j^c(S,S_c) \geq \delta_j \bar{\beta}_j, \qquad j = 1,2, \quad (6)$$

$$\beta_j^n(S,S_c) \geq (1-\delta_j)\bar{\beta}_j, \; j = 1,2, \quad (7)$$

$$S, S_c \geq 0, \quad (8)$$

Where

$$\delta_j = \begin{cases} 1, & \text{if } \bar{\beta}_j = \max_k \bar{\beta}_k \\ 0, & \text{otherwise.} \end{cases}$$

In this model, we minimize the stock level, while maintaining a minimum fixed service level. With δ_j, the constraint for the appropriate situation is set active. With now all the expressions from Koçaǵa and Şen (2007) stated, we can perform the replication study in section 5.1. We will now continue with the cost function approximation.

4.2 Cost function approximation

To get a better insight into the annual costs, we will derive an annual cost function approximation. This approximation should approach the actual cost, just like the service level approximation approaches the service level. We will compare our approximation with a simulation study in order to check how accurate our approximation is in different cases. In this part of the methodology, we will guide you through the structure of the approximation.

In order for us to create an approximation for the annual costs, there are several things we need. As discussed before, we already made assumptions for the stockout and delay costs. We will later vary these parameters to test the approximation. The second thing that is of big importance, is the development of the demand events. We have to approximate how often a demand event cannot be satisfied for each class, and how long that demand is backlogged before it is finally satisfied.

First, we will create an expression for estimating the stockout costs. These penalty costs are calculated by multiplying the amount of backlogged demand events with the stockout cost π. This stockout cost π differs for the critical and non-critical class, but can be chosen to be any value. Therefore, we will start with finding an expression that approximates the amount of backlogged demand events as close as possible. One last thing we have to note, is that the service level expressions proposed by Koçaǵa and Şen (2007), can be used to compute the stockout costs. We will start with these same expressions and clarify why we come to these expressions.

We already have the first part we need for the stockout cost, namely the service level expressions from 4.1. These expressions display what percentage of the demands from either the critical or non-critical class can be fulfilled at their due date. If we subtract this percentage from 1, we get the percentage of non-critical demands that could not be fulfilled at their due date. Multiplying this with the according stockout cost π and the amount of expected demand events of the corresponding class λ, we get the total stockout cost. If λ, as well as L and T are expressed in years, this function automatically becomes the annual cost for stockout. Combining these factors, we get the following expressions for the total stockout cost of the critical and non-critical demand class

$$\Pi_j^n = \left(1 - \beta_j^n(S, S_c)\right) * \pi_n * \lambda_n \tag{9}$$

and

$$\Pi_j^c = \left(1 - \beta_j^c(S, S_c)\right) * \pi_c * \lambda_c \tag{10}$$

The next part of cost we want to approximate, are the delay costs. We will first consider the non-critical demand class. When an order cannot be satisfied at its due date, it will be backlogged. In this case, that would be because the inventory level is below the critical stock level. With backlogging, we basically mean that it gets placed in a FIFO-queue with all other non-critical demands that are currently backlogged. When a replenishment order comes in, one of the following three events occurs. Either a critical class backorder is cleared, the replenishment order is added to the on-hand stock, or a non-critical class backorder is cleared if the on-hand stock is equal to or above the critical stock level. We will take as basis the expressions for the long-run average number of backorders, derived by Deshpande et al (2003). They start off with deriving the number of backorders at a given time t, denoted as $BO_i(t)$. For both the demand classes, this is noted as

$$BO_1(t + \tau) = [D_1(t_K, t + \tau) - K]^+ \tag{11}$$

$$BO_2(t + \tau) = D_2(t_K, t + \tau) \tag{12}$$

with $D_i(t_j, t_j + \tau)$ the number of demands of class I that arrive between the placement and the receipt of replenishment order j. Also, t_K is defined as the time of $y - Kth$ demand arrival in the interval $(t, t + \tau)$. If $y < K$, t_K is defined as the last time the inventory was at the critical stock level K, before reaching y at time t. With these expressions, Deshpande et al (2003) compute the limiting backorder distributions for both demand classes. With α_i = λ_i/ λ being the probability of an arrival being class i, i = 1, 2, and $\tau = L$, the limiting backorder distributions are given by

$$Prob[BO_i(\infty) = j] = \frac{1}{Q} \Sigma_{y=r+1}^{r+Q} Prob\,[BO_1(\infty) = j | IP(\infty) = y] \tag{13}$$

where

$$Prob[BO_1(\infty) = j|IP(\infty) = y] = \begin{cases} \sum_{x=y+j}^{\infty} b(\alpha_1; x - y + K; K + j)p(x: \lambda\tau) & if\ j > 0 \\ 1 - \left(\sum_{h=1}^{\infty} \sum_{x=y+h}^{\infty} b(\alpha_1; x - y + K; K + h)p(x: \lambda\tau) \right) & if\ j = 0 \end{cases}$$

and

$$Prob[BO_2(\infty) = j|IP(\infty) = y] = \sum_{x=(j+y-K)^+}^{\infty} b(\alpha_1; x - y + K; K + j)p(x: \lambda\tau)$$

These expressions are used to calculate the average number of backorders in a system with rationing, but without DLT. The expressions below denote the limiting backorder distribution for both the critical and the non-critical demand class, in the case where DLT = 0 for both the critical and the non-critical class. With α_i = λ_i / λ being the probability of an arrival being class i, i = 1, 2, $\tau = L$, $B_i(Q, r, K)$ and denoting the average number of backorders for class I, we get

$$B_i(Q, r, K) = \frac{1}{Q} \sum_{y=r+1}^{r+Q} b_i(y, K), \tag{14}$$

where

$$b_1(y, K) = \sum_{x=y}^{\infty} \sum_{j=0}^{x-y} jb(\alpha_1; x - y + K; K + j)p(x: \lambda\tau)$$

and

$$b_2(y, K) = \begin{cases} \sum_{x=y-k}^{\infty} \alpha_2(x - y + K)p(x: \lambda\tau) & if\ K \leq y \\ \lambda_2\tau + \alpha_2(K - y) & if\ K > y \end{cases}$$

Note that up to this point, these expressions were derived by Deshpande et al (2003). Now that we have these expressions, a few adaptations have to be made to fit our assumption of DLT. First of all, since we consider an (S-1, S) ordering policy, the summation of $b_i(y, K)$ becomes unnecessary. This is because the value of y that is summed over can only attain the stock level S. Therefore, we get $b_i(S, K)$ as an expression for the average number of backorders of class i.

Since we only test the (S-1, S) policy, we can fix y simply to the stock level. We adjust this function to work for j = 1, 2, with j indicating the demand class that is the non-critical class. Now we want to add demand lead time into the average backorder expression. For the first demand class, the DLT = 0. Therefore, we still multiply λ_1 with the replenishment lead time T. However, we now have the second demand class with DLT = T. Since from its due date, it takes (L – T) until the replenishment order arrives, we multiply λ_2 with (L – T). Now that we have incorporated the DLT, we get the expression for the average backorders of the non-critical class as

$$b_j^n(S,K) = \sum_{x=S-k}^{\infty} \alpha_n (x - S + K) p(x: \lambda_1 L + \lambda_2 (L - T)) \tag{15}$$

Now, we can add the delay cost $\hat{\pi}_n$ to get the following expression for the total delay cost of the non-critical delay cost:

$$\widehat{\Pi}_j^n = \hat{\pi}_n b_j^n(S,K) \tag{16}$$

Next, we want to derive an expression for the average delay cost of the critical demand group. We again consider an expression derived by Deshpande et al (2003) from the previous expression (14), which is very similar to the expression for the non-critical class. However, in this expression, a binomial distribution is needed. Again, we incorporate the DLT by changing λ to λ_1L + λ_2(L-T), set y to the stock level and make the function suitable for j = 1, 2, with j in this case indicating the demand class that is the critical class. Implementing these adjustments gives us

$$b_j^c(S,K) = \sum_{x=S}^{\infty} \sum_{j=0}^{x-S} j b(\alpha_c; x - S + K; K + j) p(x: \lambda_1 L + \lambda_2 (L - T)) \tag{17}$$

Now, we can add the delay cost $\hat{\pi}_n$ to get the following expression for the total delay cost of the critical delay cost:

$$\widehat{\Pi}_j^c = \hat{\pi}_c b_j^c(S,K) \tag{18}$$

Finally, we consider the holding cost of the inventory. This cost is the cost of holding one unit of stock per time unit times the average inventory. To calculate this, we use another expression derived by Deshpande et al (2003) and modify this expression to work for our case with DLT. This expression is

$$Prob[OH(\infty) = j | IP(\infty) = y] = \begin{cases} p(y - j; \lambda\tau) \; if \; y \geq j \geq K, \;\; j > 0 \\ \sum_{x=y-j}^{\infty} b(\alpha_1; x - y + K; K - j) p(x: \lambda_1 L + \lambda_2 (L - T)) \;\; if \; 0 < j < K \\ \sum_{x=y}^{\infty} \sum_{z=K}^{x-y+K} (\alpha_c; x - y + K; z) p(x: \lambda_1 L + \lambda_2 (L - T)) \;\; if \; j = 0 \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad otherwise. \end{cases}$$

In this function, OH(t) is referred to as the on-hand inventory, and IP(t) is referred to as the inventory position process. With this expression, we can calculate the chance that the inventory level will be j. If we do this for all possible stock levels j ≤ S, we get the average inventory level. We add demand lead time in all Poisson probability functions in the same manner we did before with the delay cost. Furthermore, we make sure that the average inventory is usable for both critical and non-critical customers having either DLT=0 or DLT=T. Also, we can again set y to the stock level, because of our fixed ordering policy. Applying all these modifications, we get the following functions for the stock levels:

$$Prob[OH(\infty) = j | IP(\infty) = S]$$
$$= \begin{cases} p(S-j; \lambda_1 L + \lambda_2 (L-T)) \; if \; S \geq j \geq K, \;\; j > 0 \\ \sum_{x=S-j}^{\infty} b(\alpha_1; x-S+K; K-j) p(x: \lambda_1 L + \lambda_2(L-T)) & if \; 0 < j < K \\ \sum_{x=S}^{\infty} \sum_{z=K}^{x-y+K} b(\alpha_1; x-S+K; z) p(x: \lambda_1 L + \lambda_2(L-T)) & if \; j = 0 \\ 0 & otherwise. \end{cases}$$

With the functions now determined for the inventory levels, we can obtain the average inventory level, and multiply this with the average cost of holding one unit of stock. This gives us the total holding cost

$$H^T = h * \sum_{j=0}^{S} (j * Prob[OH(\infty) = j | IP(\infty) = S]) \qquad (19)$$

With all of the separate cost functions now defined, we can now setup our total cost function, which will be tested in our study. The total annual cost function is defined as

$$C(S-1, S) = \mathrm{S}_\mathrm{j}^\mathrm{n} + \mathrm{S}_\mathrm{j}^\mathrm{c} + \mathrm{D}_\mathrm{j}^\mathrm{n} + \mathrm{D}_\mathrm{j}^\mathrm{c} + H^T$$

We can now fill in all parts of this cost function. This gives us the cost function that we will use to calculate the total annual cost later in our study:

$$C(S-1, S) = \left(1 - \beta_j^n(S, S_n)\right) * \pi_n + \left(1 - \beta_j^c(S, S_c)\right) * \pi_c + \hat{\pi}_n b_j^n(S, K) + \hat{\pi}_c b_j^c(S, K) + H^T \qquad (20)$$

4.3 Cost optimization

Now that we have derived all components of the cost function, we can setup a minimization model to optimize the annual cost. In this model, we also want to take service level requirements into account:

$$Min \; C(S-1, S) \qquad (21)$$

$$subject \; to \;\; \beta_j^c(S, S_c) \geq \delta_j \bar{\beta}_j, \qquad j = 1, 2, \qquad (22)$$

$$\beta_j^n(S, S_c) \geq (1 - \delta_j) \bar{\beta}_j, \;\; j = 1, 2, \qquad (23)$$

$$S, S_c \geq 0, \qquad (24)$$

Where

$$\delta_j = \begin{cases} 1, & \text{if } \bar{\beta}_j = \max_k \bar{\beta}_k \\ 0, & \text{otherwise.} \end{cases}$$

This optimization is very similar to the optimization problem stated by Koçaǵa and Şen (2007), but in our case we of course want to minimize total cost instead of stock level. The first constraint makes sure that if the critical class is of demand class j, the service level is at least the minimum required service level by setting $\delta_j = 1$. If $\delta_j = 0$, class j is the non-critical class, and the second constraint ensures that the non-critical service level is at least the minimum required service level. The last constraint makes sure that the stock and critical stock levels are non-negative. In section 4, we will apply this model to multiple instances by using brute force. Since the chosen instances are quite small, it is not very time consuming. For instances with larger parameters, obviously time could be saved by implement an optimizing algorithm.

5. Study

Our study consists of three parts. In section 5.1, we will replicate a part of the study performed by Koçağa and Şen (2007). We calculate their proposed service level expressions, and setup a simulation[1] to replicate the accuracy testing of their critical class service level approximation. In section 5.2, we test our composed cost function. We use the same simulation program to test the accuracy of our cost function for different instances of parameters. In section 5.3, we will look at the performance of the total cost composed by the approximation against the simulated cost.

5.1 replication study

In this section, we replicate the study performed by Koçağa and Şen (2007). We take the same steps of computing their simulation results, to see whether we come to the same results. All tables show the difference between the exact and the simulated non-critical service level. This could give an indication on how accurate our simulation is. Furthermore, we find the simulated critical service level, the approximated critical service levels and the percentage difference between these two service levels, which is calculated as ($\beta_c$sim-$\beta_c$app)*100. Also, the values between brackets next to the simulated critical service level shows the difference between our simulated values and the values simulated by Koçağa and Şen (2007).

				c = 1, n = 2				c = 2, n = 1			
λ_c	λ_n	S	S_c	β_n diff (%)	β_c (sim)	β_c (approx.)	Perc. diff (%)	β_n diff (%)	β_c (sim)	β_c (approx.)	Perc. diff (%)
1	4	5	3	0.00	0.9995 (0.0000)	0.9976	0.19	0.28	0.9990 (0.0003)	0.9976	0.14
2	4	6	3	0.07	0.9979 (0.0002)	0.9927	0.52	0.18	0.9975 (0.0002)	0.9927	0.48
3	4	7	3	0.13	0.9970 (0.0002)	0.9892	0.78	0.02	0.9965 (0.0001)	0.9891	0.74
4	4	8	3	0.00	0.9964 (0.0002)	0.9877	0.87	0.03	0.9960 (0.0002)	0.9877	0.83
5	4	9	3	0.08	0.9957 (0.0001)	0.9876	0.81	0.02	0.9961 (0.0003)	0.9877	0.84
6	4	10	3	0.04	0.9957 (0.0001)	0.9884	0.73	0.02	0.9966 (0.0003)	0.9885	0.81
7	4	11	3	0.05	0.9960 (0.0000)	0.9896	0.64	0.08	0.9973 (0.0000)	0.9898	0.75
8	4	12	3	0.05	0.9961 (0.0003)	0.9909	0.52	0.03	0.9977 (0.0002)	0.9913	0.64
9	4	13	3	0.14	0.9966 (0.0001)	0.9922	0.44	0.05	0.9983 (0.0000)	0.9927	0.56
10	4	14	3	0.01	0.9971 (0.0000)	0.9934	0.37	0.00	0.9984 (0.0003)	0.9940	0.44
11	4	15	3	0.06	0.9974 (0.0001)	0.9945	0.29	0.02	0.9988 (0.0002)	0.9951	0.37
12	4	16	3	0.02	0.9979 (0.0001)	0.9954	0.25	0.02	0.9991 (0.0001)	0.9961	0.3
2	4	8	1	0.00	0.9982 (0.0001)	0.9963	0.19	0.02	0.9973 (0.0000)	0.9957	0.16
3	4	8	2	0.05	0.9973 (0.0001)	0.9928	0.45	0.04	0.9968 (0.0001)	0.9927	0.41
4	4	8	3	0.00	0.9964 (0.0002)	0.9877	0.87	0.03	0.9960 (0.0003)	0.9877	0.83
5	4	8	4	0.04	0.9941 (0.0002)	0.9802	1.39	0.08	0.9950 (0.0004)	0.9802	1.48
6	4	8	5	0.14	0.9920 (0.0003)	0.9697	2.23	0.05	0.9946 (0.0002)	0.9697	2.49
7	4	8	6	0.02	0.9909 (0.0001)	0.9554	3.55	0.05	0.9946 (0.0001)	0.9554	3.92
8	4	8	7	0.00	0.9918 (0.0003)	0.9368	5.50	0.01	0.9953 (0.0003)	0.9367	5.86

Table 1: Replication of the Koçağa and Şen (2007) approximation performance for a fixed service level of 99% (L = 0.5 and T = 0.1)

[1] The simulation code is adapted from the Monte Carlo and discrete event simulation Java code provided by Nemanja Milovanovic, Erasmus University Rotterdam

In table 1, we find the first results of our simulation, with the value in between the brackets being the absolute difference in simulation outcome values between our service level, and the service level obtained by Koçağa and Şen (2007).

We will also replicate table 2 of Koçağa and Şen (2007). For this table, we also assume the same supply lead time, as well as the same DLT, which will again be L = 0.5, and T = 0.1. Below, we can find our replication of this second table.

				c = 1, n = 2				c = 2, n = 1			
λ_c	λ_n	S	S_c	β_n diff (%)	β_c (sim)	β_c (approx.)	Perc. diff (%)	β_n diff (%)	β_c (sim)	β_c (approx.)	Perc. diff (%)
4	1	5	2	0.29	0.9395 (0.0015)	0.9190	2.05	0.15	0.9602 (0.0007)	0.9208	3.94
5	1	6	2	0.05	0.9486 (0.0005)	0.9339	1.47	0.16	0.9712 (0.0000)	0.9368	3.44
6	1	7	2	0.12	0.9573 (0.0000)	0.9467	1.06	0.02	0.9788 (0.0002)	0.9505	2.83
7	1	8	2	0.20	0.9656 (0.0004)	0.9573	0.83	0.08	0.9847 (0.0003)	0.9617	2.30
8	1	9	2	0.09	0.9720 (0.0002)	0.9658	0.62	0.15	0.9892 (0.0000)	0.9706	1.86
9	1	10	2	0.07	0.9768 (0.0004)	0.9726	0.42	0.06	0.9917 (0.0006)	0.9776	1.41
5	1	7	1	0.05	0.9762 (0.0001)	0.9722	0.40	0.01	0.9881 (0.0002)	0.9785	0.96
6	1	7	2	0.12	0.9573 (0.0000)	0.9467	1.06	0.02	0.9788 (0.0002)	0.9505	2.83
7	1	7	3	0.05	0.9328 (0.0007)	0.9118	2.10	0.14	0.9666 (0.0000)	0.9130	5.36
8	1	7	4	0.05	0.9039 (0.0001)	0.8671	3.68	0.07	0.9512 (0.0005)	0.8673	8.39

Table 2: Replication of the Koçağa and Şen (2007) approximation performance for a fixed service level of 95% (L = 0.5 and T = 0.1)

For this instance, they tried to create ten instances, where the service level would be around 95%. We see that the differences between our simulated service levels, and the service levels simulated by Koçağa and Şen (2007), are again considerably small. This means that we observe the same as Koçağa and Şen (2007). The approximation works well for the 95% service level, however not as well as for the 99% service level cases. The average differences between the approximation and our simulation are 1.37% and 3.33% for the two different cases.

If we compare the results from Koçağa and Şen (2007) with our own simulation, we can assume that the results they obtained are very plausible. Also, we see that their calculation for the non-critical, and their approximation for the critical service levels, match with our calculated service levels. Therefore, we presume that all these results are correct, and that we can use the approximation as a foundation for our subsequent research.

There is however a minor error in their definitions. They choose the percentage difference to be 100 times simulation approximation / simulation, however the values they placed in the table is calculated as simulated service level **minus** the approximated service level, and then multiplied by 100 to get a percentage value. We will use this exact method to calculate the percentage differences.

Finally, we reproduce the third table presented by Koçağa and Şen (2007), where they test the performance of their approximation by varying a single parameter. They test for only varying stock, arrival rate of the critical class, arrival rate of the non-critical class and the Demand Lead Time.

						c = 1, n = 2				c = 2, n = 1			
						β_n diff (%)	β_c (sim)	β_c (approx.)	Perc. diff (%)	β_n diff (%)	β_c (sim)	β_c (approx.)	Perc. diff (%)
S	S_c	λ_c	λ_n	L	T								
7	2	6	2	0.5	0.1	0.06	0.9490 (0.0004)	0.9225	2.79	0.02	0.9672 (0.0006)	0.9258	4.28
8	2	6	2	0.5	0.1	0.00	0.9773 (0.0000)	0.9655	1.21	0.04	0.9869 (0.0002)	0.9678	1.94
9	2	6	2	0.5	0.1	0.03	0.9907 (0.0002)	0.9861	0.46	0.08	0.9954 (0.0000)	0.9874	0.80
10	2	6	2	0.5	0.1	0.09	0.9966 (0.0001)	0.9949	0.17	0.02	0.9985 (0.0000)	0.9955	0.30
11	2	6	2	0.5	0.1	0.04	0.9988 (0.0001)	0.9983	0.05	0.03	0.9995 (0.0000)	0.9986	0.09
5	2	1	1	1	0.5	0.10	0.9947 (0.0003)	0.9860	0.87	0.14	0.9959 (0.0035)	0.9989	-0.30
5	2	2	1	1	0.5	0.14	0.9474 (0.0007)	0.9008	4.92	0.11	0.9780 (0.0173)	0.9906	-1.29
5	2	3	1	1	0.5	0.19	0.8390 (0.0013)	0.7378	12.06	0.06	0.9410 (0.0425)	0.9662	-2.68
5	2	4	1	1	0.5	0.15	0.6971 (0.0010)	0.5438	21.99	0.14	0.8876 (0.0733)	0.9208	-3.74
5	2	5	1	1	0.5	0.23	0.5597 (0.0017)	0.3668	34.46	0.27	0.8202 (0.1055)	0.8543	-4.16
5	2	1	1	1	0.5	0.07	0.9947 (0.0003)	0.9860	0.87	0.14	0.9959 (0.0035)	0.9989	-0.30
5	2	1	2	1	0.5	0.09	0.9938 0.0002)	0.9686	2.54	0.05	0.9893 (0.0092)	0.9972	-0.80
5	2	1	3	1	0.5	0.09	0.9929 (0.0001)	0.9484	4.48	0.11	0.9848 (0.0125)	0.9946	-1.00
5	2	1	4	1	0.5	0.03	0.9925 (0.0002)	0.9274	6.56	0.15	0.9835 (0.0127)	0.9914	-0.80
5	2	1	5	1	0.5	0.03	0.9922 (0.0001)	0.9072	8.57	0.03	0.9845 (0.0109)	0.9980	-1.37
14	3	10	4	0.5	0.10	0.00	0.9971 (0.0000)	0.9934	0.37	0.00	0.9986 (0.0001)	0.9940	0.46
14	3	10	4	0.5	0.20	0.13	0.9982 (0.0001)	0.9953	0.29	0.04	0.9997 (0.0001)	0.9975	0.22
14	3	10	4	0.5	0.30	0.10	0.9990 (0.0000)	0.9973	0.17	0.01	1.0000 (0.0000)	0.9996	0.04
14	3	10	4	0.5	0.40	0.02	0.9994 (0.0000)	0.9986	0.08	0.00	1.0000 (0.0000)	1.0000	0.00
14	3	10	4	0.5	0.50	0.38	0.9995 (0.0000)	0.9993	0.02	0.00	1.0000 (0.0000)	1.0000	0.00

Table 3: Replication of the Koçağa and Şen (2007) performance of the approximation with varying system parameters

In the top part of the table, we can see that all service levels become higher if we increase the stock level. Also, the approximation becomes better as the stock level rises. In the second and third part of the table, we look at the impact of the arrival rates. In the second part, we change the critical arrival rate, in the third part we change the non-critical arrival rate. We can see in the second part of the table that the critical arrival rate has quite a big impact on the performance of the approximation. The approximation performs considerably worse as the critical arrival rate starts becomes higher. This impact on the performance of the approximation is much smaller if we test for different levels of non-critical arrival rates. If we look at the fourth part of the table, we study the impact of the Demand Lead Time T. If T increases, the service levels also increase for both criticality classes. However, the approximation performs quite well for all different Demand Lead Times.

	c = 1, n = 2							c = 2, n = 1						
$\bar{\beta}_c$	S_T^*	S	S_c	Percentage Saving (%)	S^*	S_c^*	Percentage Saving (%)	S_T^*	S	S_c	Percentage Saving (%)	S^*	S_c^*	Percentage Saving (%)
0.900	33	32	2	3.03	31	1	6.06	35	35	0	0,00	34	1	2.86
0.925	33	33	0	0.00	31	1	6.06	36	35	2	2.78	34	1	5.56
0.950	34	34	0	0.00	31	1	8.82	37	36	3	2.70	34	1	8.11
0.970	36	35	5	2.78	32	2	11.11	39	36	3	7.69	35	2	10.26
0.980	37	35	5	5.41	32	2	13.51	40	37	4	7.50	35	2	12.50
0.985	37	36	6	2.70	32	2	13.51	40	37	4	7.50	35	2	12.50
0.990	38	36	6	5.26	33	3	13.16	41	38	5	7.32	36	3	12.20
0.995	40	37	7	7.50	33	3	17.50	43	39	6	9.30	36	3	16.28

Table 4: Replication of the Koçağa and Şen (2007) optimal parameters: approximation vs simulation ($\lambda_c = 5, \lambda_n = 10, L = 2, T = 0.5$ *and* $\bar{\beta}_n = 0.80$)

Finally, we replicate the comparison of the approximation savings with the simulated savings. In table 4, we find this for fixed parameters, and only varying the required service level of the critical class. If we compare the approximated savings against the simulated savings, we see that in none of the 16 instances, the approximation estimates the same savings as the simulated savings displays. It seems though as our simulation obtains lower stock levels in some situations than the stock level obtained in the simulation made by Koçağa and Şen (2007). In 13 of the 16 instances, the approximation does however provide a lower stock level than would be needed if no rationing was applied. Also, we see that the savings become higher as the service level of the critical class has to be higher. This seems logical, since when the difference between the two service levels becomes bigger, a round-up policy is not convenient, since the non-critical class gets a much higher service level than is required.

If we compare the simulation of Koçağa and Şen (2007) with our simulation results, we can see that the service levels are in most situations very comparable. However, if we look in the second and third part of the table, for the situation where the second demand class is critical, our simulation seems to get very different results from the results of Koçağa and Şen (2007). The first and fourth part of the table are however almost identical to the results of Koçağa and Şen (2007). This could be due to a mistake in the simulation, since our simulated service levels are lower than the service level that follow from the approximation. We have marked these simulation results with red crossed out lettering. However, it is also possible that the results obtained by Koçağa and Şen (2007) are wrong for these instances, since all other simulated service levels seem to correspond to their simulated values.

If we summarize the comparison between our simulation results and the results obtained by Koçağa and Şen (2007), we see that our simulated service levels are mostly corresponding. In both cases, the approximation Koçağa and Şen (2007) proposed seems to be quite accurate. We will therefore use their approximation as basis for our extension, where we will include the costs of delay and stockout.

5.2 Cost approximation study

Now that we have established that the service level approximation created by Koçağa and Şen (2007) looks correct, we can start to test our cost function approximation. We will compare the approximation results with a simulated cost function. For this simulation, we use a modified version of our simulation used for the replication study. In table4, we choose the same parameter values as in table 1 used for the replication study. In column 5-9, we find the difference in simulation and approximation for all 5 cost function components for the critical customers being the first demand class, and the non-critical customers being the second demand class. In column 10-14, we find the same, but now for the case where the critical customers are of the second demand class, and the non-critical customers are of the first demand class. As for the different colors, they roughly indicate how big the simulated cost component is. Red differences correspond with an absolute simulated value bigger than 1, orange differences correspond with absolute simulated values between 0.1 and 1, and green differences correspond with absolute simulated values between 0.001 and 0.1. If the difference has no color, the simulated value is smaller than 0.0001. We also apply this grouping later in table 6 and 7.

				c = 1, n = 2					c = 2, n = 1				
λ_c	λ_n	S	S_c	Π_c Diff.	Π_n Diff.	$\hat{\Pi}_c$ Diff.	$\hat{\Pi}_n$ Diff.	H Diff.	Π_c Diff.	Π_n Diff.	$\hat{\Pi}_c$ Diff.	$\hat{\Pi}_n$ Diff.	H Diff.
1	4	5	3	-0,002	0,004	0,000	0,043	0,774	-0,001	-0,004	0,000	0,065	0,941
2	4	6	3	-0,011	0,001	-0,001	0,049	0,466	-0,010	-0,001	-0,001	0,066	0,542
3	4	7	3	-0,024	-0,002	-0,001	0,044	0,301	-0,023	0,001	-0,001	0,053	0,326
4	4	8	3	-0,033	0,002	-0,001	0,037	0,199	-0,034	0,003	-0,001	0,038	0,200
5	4	9	3	-0,040	-0,002	-0,002	-0,096	0,140	-0,043	0,000	-0,001	0,027	0,127
6	4	10	3	-0,044	0,000	-0,001	0,024	0,099	-0,050	-0,002	-0,001	0,019	0,082
7	4	11	3	-0,045	-0,004	-0,001	0,019	0,071	-0,053	-0,001	0,000	0,013	0,049
8	4	12	3	-0,043	0,002	-0,001	0,015	0,051	-0,053	-0,001	0,000	0,009	0,037
9	4	13	3	-0,040	0,002	-0,001	0,012	0,037	-0,051	0,001	0,000	0,006	0,022
10	4	14	3	-0,037	0,001	-0,001	0,009	0,028	-0,046	0,001	0,000	0,004	0,011
11	4	15	3	-0,033	0,000	-0,001	0,007	0,019	-0,043	0,000	0,000	0,002	0,010
12	4	16	3	-0,028	0,000	-0,001	0,006	0,018	-0,038	0,001	0,000	0,002	0,005
2	4	8	1	-0,004	0,014	0,000	0,002	0,502	-0,004	0,001	0,000	0,001	0,006
3	4	8	2	-0,008	0,114	0,000	0,015	0,513	-0,013	0,001	0,000	0,007	0,047
4	4	8	3	-0,015	0,323	-0,001	0,071	0,573	-0,034	0,003	-0,001	0,038	0,200
5	4	8	4	-0,025	0,454	-0,002	0,192	0,688	-0,076	0,001	-0,003	0,140	0,524
6	4	8	5	-0,038	0,304	-0,002	0,346	0,843	-0,150	0,000	-0,004	0,385	0,951
7	4	8	6	-0,050	0,115	-0,003	0,507	1,003	-0,275	0,006	-0,009	0,842	1,376
8	4	8	7	-0,441	0,001	-0,024	1,690	1,946	-0,471	0,004	-0,016	1,534	1,860

Table 5: Performance of the separate cost component approximations for a fixed critical service level of 99% (L = 0.5 and T = 0.1) with all cost π_c, π_n, $\hat{\pi}_c$, $\hat{\pi}_n$ and h to 1 for comparison purposes

If we look at the results of our first performance results, a few things become obvious straight away. The stockout cost for the non-critical class should be the same for the simulation as for our expression, since the expression uses an exact calculation for the backorders of the non-critical class. The small differences in column 6 are therefore probably due to a slight error in the simulation. Also, we can see that our approximation for the delay cost of the critical class gives almost the same output as the simulation suggests, but this seems reasonable since the simulated values are very small. What we see is that all other approximations work quite well, but start to differ more from the simulated values for a critical stock level close to the total stock level. Mainly the approximation for the delay cost for the non-critical class and the holding cost then become considerably worse. This corresponds to what we see in our replication study for the service level approximation, which also performs worse in the same cases as we see here.

In the table 6, we repeat the previous performance, but we now choose parameters that result in a critical service level of around 95% instead of 99%. Again, in this instance we have supply lead time L = 0.5, and DLT T=0.1. Again, the stockout cost for the non-critical class calculated by the approximation is equal to the outcome of the simulation. Also, the delay cost for the critical class calculated by the approximation is again practically equal to the simulation results. We see that for the rest of the terms, the difference becomes bigger when we get towards a worse service level approximation, which we can see in table 2 is much worse for the last case for example. The differences we see here are however smaller than for the fixed critical service level of 99%.

				c = 1, n = 2					c = 2, n = 1				
				Π_c	Π_n	$\hat{\Pi}_c$	$\hat{\Pi}_n$	H	Π_c	Π_n	$\hat{\Pi}_c$	$\hat{\Pi}_n$	H
λ_c	λ_n	S	S_c	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*
1	4	5	3	-0,076	0,000	-0,005	0,031	0,082	-0,159	0,000	-0,003	0,021	0,060
2	4	6	3	-0,073	0,001	-0,005	0,021	0,051	-0,171	0,001	-0,002	0,013	0,034
3	4	7	3	-0,065	0,001	-0,003	0,016	0,037	-0,170	0,001	-0,002	0,007	0,020
4	4	8	3	-0,055	0,001	-0,003	0,012	0,024	-0,163	0,001	-0,001	0,005	0,012
5	4	9	3	-0,049	0,001	-0,002	0,009	0,019	-0,148	-0,001	0,000	0,003	0,007
6	4	10	3	-0,039	0,000	-0,002	0,006	0,010	-0,132	0,000	0,000	0,002	0,002
7	4	11	3	-0,020	0,000	-0,001	0,003	0,004	-0,049	0,000	0,000	0,001	0,002
8	4	12	3	-0,065	0,001	-0,003	0,016	0,037	-0,170	0,000	-0,002	0,007	0,020
9	4	13	3	-0,143	-0,002	-0,009	0,064	0,118	-0,375	0,001	-0,004	0,036	0,083
10	4	14	3	-0,299	0,001	-0,023	0,202	0,276	-0,672	-0,001	-0,010	0,123	0,208

Table 6: Performance of the separate cost component approximations for a fixed critical service level of 95% (L = 0.5 and T = 0.1) with all cost π_c, π_n, $\hat{\pi}_c$, $\hat{\pi}_n$ and h to 1 for comparison purposes

In table 7, we don't attain parameters to reach a certain service level. We now choose the parameters freely and vary also in supply lead time and DLT. We only vary one parameter at a time, to see for which parameter change the approximations work well, and for which they approach the simulation results a bit worse. We see that varying the DLT basically has no impact on any of the cost component approximations. The biggest differences appear in the case where we vary the parameter of the critical class demand arrivals. We can see that mainly the critical stockout cost component is quite different from the simulated critical

						c = 1, n = 2					c = 2, n = 1				
						Π_c	Π_n	$\hat{\Pi}_c$	$\hat{\Pi}_n$	H	Π_c	Π_n	$\hat{\Pi}_c$	$\hat{\Pi}_n$	H
S	S_c	λ_c	λ_n	L	T	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*	*Diff.*
5	2	1	1	1	0.5	-0,009	0,000	0,000	0,006	0,060	-0,010	0,000	0,000	0,008	0,063
5	2	2	1	1	0.5	-0,094	0,002	-0,004	0,039	0,141	-0,127	-0,001	-0,003	0,025	0,091
5	2	3	1	1	0.5	-0,297	0,001	-0,028	0,123	0,208	-0,470	0,001	-0,010	0,053	0,119
5	2	4	1	1	0.5	-0,611	-0,001	-0,109	0,280	0,265	-1,044	0,000	-0,025	0,095	0,144
5	2	5	1	1	0.5	-0,970	-0,001	-0,279	0,521	0,307	-1,730	0,002	-0,053	0,151	0,169
5	2	1	1	1	0.5	-0,009	0,000	0,000	0,460	0,060	-0,010	0,000	0,000	0,008	0,063
5	2	1	2	1	0.5	-0,025	0,000	0,000	0,017	0,178	-0,025	0,001	-0,003	0,049	0,304
5	2	1	3	1	0.5	-0,045	0,002	-0,001	0,035	0,341	-0,045	0,002	-0,006	0,120	0,645
5	2	1	4	1	0.5	-0,065	0,003	-0,001	0,055	0,528	-0,066	-0,001	-0,011	0,202	0,963
5	2	1	5	1	0.5	-0,085	-0,001	-0,002	0,078	0,720	-0,088	0,001	-0,017	0,289	1,202
14	3	10	4	0.5	0.10	-0,037	0,001	-0,001	0,009	0,028	-0,046	0,001	0,000	0,004	0,011
14	3	10	4	0.5	0.20	-0,028	0,000	0,000	0,005	0,015	-0,022	0,001	0,000	0,001	0,002
14	3	10	4	0.5	0.30	-0,017	0,001	0,000	0,003	0,090	-0,003	0,000	0,000	0,000	0,001
14	3	10	4	0.5	0.40	-0,008	0,000	0,000	0,002	0,007	0,000	0,000	0,000	0,000	0,000
14	3	10	4	0.5	0.50	-0,002	-0,012	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000

Table 7: Performance of the separate cost component approximations for varying parameters (L = 0.5 and T = 0.1) with all cost π_c, π_n, $\hat{\pi}_c$, $\hat{\pi}_n$ and h to 1 for comparison purposes

stockout cost in the case where the critical customers are of demand class 2. This however coheres with a probable error of the critical service level calculation in our simulation. We already saw in table 3 that these values are incorrect. While we directly use these service level approximation values to calculate our stockout cost, suspect that to be the cause of the big differences in this case, especially since the other cost components differ less from their simulated values. This also applies to the rows where we vary the non-critical demand events.

After analyzing the performance of our separate cost function expressions, we see that in most situations, the approximations can be used to estimate actual cost quite accurately. For the stockout cost for the non-critical class we have an exact expression. We see that for the stockout cost and the delay cost for the critical class, we actually have an upper bound. Note that the delay cost for the critical class appears to be very similar to the simulation, but this is due to the delay cost being very small in most instances. There are however a few instances where the delay cost is slightly larger, and there you can see the approximation starting to differ from the simulated delay cost. For the delay cost of the non-critical class and the holding cost, we have an upper bound. We keep in mind that for the situations, for which in the first three tables we see that the critical service level approximation is close to the simulated service level, we have a good accurate cost function approximation.

5.3 Cost optimization study

After testing the expressions for the components of our composed function, we will now optimize the cost function as described in section 4.3. We find the minimum cost using the approximation of the cost function and compare this with the minimum cost found by the simulation. In table 8, we compare cost for fixed stock and critical stock level, lead times and critical demand events, for different values of non-critical demand events. In table 9, we switch the critical and non-critical demand events, so we fix the non-critical demand events and test different values of critical demand events. In the first column, we can find the chosen value for the corresponding demand event parameter. The table is split up in two parts. We again consider both cases. For columns 2-8, demand class 1 consists of critical customers, and for columns 9-15, demand class 2 consists of critical customers. For each case, the first three columns show the optimal stock levels and total cost found by the simulation. The next three columns after that, show the optimal stock levels and total cost for minimizing the approximated total cost level. In the last column, we can find the difference in total cost, which can be used as a measure of the performance of the approximation. If we look in table 8, we see that for the critical customers being of the first demand class, the cost difference

	c = 1, n = 2							c = 2, n = 1						
	Simulation cost			*Approximation cost*			*Difference*	*Simulation cost*			*Approximation cost*			*Difference*
λ_n	*S*	S_c	*C*	*S*	S_c	*C*	*in cost (%)*	*S*	S_c	*C*	*S*	S_c	*C*	*in cost (%)*
2	4	0	0.17	4	0	0.17	0.00	4	0	0.17	4	0	0.17	0.00
4	5	0	0.20	6	0	0.21	5.00	6	0	0.21	6	0	0.21	0.00
6	7	0	0.24	8	0	0.27	12.50	7	0	0.25	7	0	0.25	0.00
8	8	0	0.26	9	0	0.28	7.69	9	0	0.28	9	0	0.28	0.00
10	9	0	0.29	11	0	0.34	17.24	10	0	0.31	11	0	0.32	3.22

Table 8: Optimal stock with cost for $\bar{\beta}_c$ = 0.99, $\bar{\beta}_n$ = 0.80, with λ_c=1, L = 0.5, T=0.1, and cost parameters set to π_c = 0.5, π_n = 0.1, $\hat{\pi}_c$ = 0.1, $\hat{\pi}_n$ = 0.02 and h = 0.05

between the simulation and the approximation is not very large. The difference does however become larger as the parameter λ_n becomes larger. In the case where the non-critical customers are of the first demand class, we even see that the approximated cost are almost exactly the same as our simulated cost.

If we look at table 9, we see that the differences are again reasonably small. When the critical customers are of the first demand class, note that the difference in cost appears to get smaller as the parameter for the critical demand events becomes larger. When the non-critical customers are of the first demand class, this also seems to be the case. The difference in cost is however bigger in this table than in table 8, where we varied the non-critical demand event parameter.

	c = 1, n = 2							c = 2, n = 1						
	Simulation cost			*Approximation cost*			*Difference*	*Simulation cost*			*Approximation cost*			*Difference*
λ_c	*S*	S_c	*C*	*S*	S_c	*C*	*in cost (%)*	*S*	S_c	*C*	*S*	S_c	*C*	*in cost (%)*
2	5	1	0.20	6	0	0.23	15.00	5	1	0.20	6	0	0.24	20.00
4	7	1	0.25	8	0	0.29	16.00	7	1	0.26	7	0	0.26	0.00
6	9	1	0.30	9	0	0.31	3.33	8	1	0.28	8	0	0.29	3.57
8	11	2	0.35	11	0	0.35	0.00	9	3	0.32	10	0	0.34	6.25
10	12	2	0.37	12	0	0.38	2.70	11	2	0.36	11	0	0.36	0.00

Table 9: Optimal stock with cost for $\bar{\beta}_c = 0.99$, $\bar{\beta}_n = 0.80$, with λ_n=1, L = 0.5, T=0.1, and cost parameters set to π_c = 0.5, π_n = 0.1, $\hat{\pi}_c = 0.1$, $\hat{\pi}_n =$ 0.02 and h = 0.05

Finally, we create table 10. This table is a replication of table 4 of Vicil (2022), but now with DLT incorporated. Note that we only created this table for the case where the non-critical class has a DLT. In columns 1-3, model parameters are stated. In columns 4, 7, 10 and 13 the optimal stock levels suggested by the simulation are shown. In columns 5, 8, 11 and 14 we find the optimal stock levels that are suggested by the cost approximation. In column 6, 9, 12 and 15 we compare the simulated cost when using the simulation optimal stock levels, with

			c = 1, n = 2											
			$\lambda L = 2.5$			$\lambda L = 5$			$\lambda L = 10$			$\lambda L = 20$		
π_n	λ_c	λ_n	(S, S_c)	(S^*, S_c^*)	*% Gap*	(S, S_c)	(S^*, S_c^*)	*% Gap*	(S, S_c)	(S^*, S_c^*)	*% Gap*	(S, S_c)	(S^*, S_c^*)	*% Gap*
0.1	0.75	0.25	(7, 3)	(5, 0)	146,35%	(10, 5)	(8, 0)	192,82%	(15, 7)	(14, 1)	105,03%	(26, 8)	(20, 0)	1305,27%
0.5	0.75	0.25	(7, 3)	(5, 0)	144,38%	(10, 4)	(8, 0)	185,85%	(16, 5)	(14, 1)	99,21%	(26, 7)	(20, 0)	1226,56%
1	0.75	0.25	(7, 2)	(5, 0)	142,73%	(10, 3)	(8, 0)	177,43%	(16, 4)	(14, 1)	91,76%	(27, 5)	(20, 0)	1165,52%
2	0.75	0.25	(7, 1)	(5, 0)	140,76%	(11, 2)	(8, 0)	170,40%	(17, 3)	(14, 0)	127,92%	(28, 3)	(20, 0)	1102,96%
0.1	0.5	0.5	(5, 3)	(4, 0)	212,51%	(8, 3)	(7, 0)	153,33%	(12, 4)	(12, 0)	175,02%	(18, 6)	(19, 0)	488,63%
0.5	0.5	0.5	(5, 2)	(4, 0)	194,14%	(8, 3)	(7, 0)	142,14%	(12, 4)	(12, 0)	148,42%	(21, 4)	(19, 0)	401,76%
1	0.5	0.5	(6, 2)	(4, 0)	182,76%	(8, 2)	(7, 0)	124,19%	(13, 3)	(12, 0)	125,91%	(22, 3)	(19, 0)	344,02%
2	0.5	0.5	(6, 1)	(5, 0)	35,34%	(9, 1)	(7, 0)	116,81%	(14, 2)	(12, 0)	112,18%	(24, 2)	(19, 0)	321,70%
0.1	0.25	0.75	(4, 2)	(4, 0)	67,89%	(5, 2)	(6, 0)	129,54%	(8, 3)	(10, 0)	174,95%	(13, 3)	(17, 0)	240,11%
0.5	0.25	0.75	(4, 2)	(4, 0)	50,27%	(6, 2)	(6, 0)	98,78%	(9, 2)	(10, 0)	116,46%	(16, 3)	(17, 0)	146,10%
1	0.25	0.75	(4, 1)	(4, 0)	39,47%	(7, 1)	(6, 0)	71,07%	(11, 2)	(10, 0)	81,66%	(18, 2)	(17, 0)	105,97%
2	0.25	0.75	(5, 1)	(4, 0)	32,97%	(7, 1)	(6, 0)	62,16%	(12, 1)	(10, 0)	73,76%	(20, 1)	(17, 0)	93,64%

Table 10: Optimization table with parameters chosen as in Vicil (2022), $\mathrm{T} = 0.5 * \mathrm{L}$ and cost parameters set to $\pi_c = 10$, $\hat{\pi}_c = 20$, $\hat{\pi}_n = 2 * \pi_n$ and h = 1

the simulated cost when using the approximated optimal stock levels, calculated as ((app. cost – sim. cost)/sim. cost)*100%. Looking at the table, we see that the gap between the cost varies from reasonably small to inordinately large. Especially for cases where the replenishment lead time is very large, the cost approximation becomes wildly ineffective. After comparing specific cost components, it appears that especially the delay cost for the critical class is not well approached for cases where the critical service level is low. This shows a weak point in our cost approximation.

Concluding from these tables, it appears that our cost function operates reasonably well for instances where a high service level is present. However, a considerable drawback is that when a minimum service level is not required, the approximation starts to deviate very much from the actual cost, which can lead to very inefficient stock levels.

6. Conclusion

In this paper, we examined a cost function for an (S-1, S) inventory management system with two demand classes. We considered two different types of criticality. Customers can either be of the critical, or the non-critical class. Also, we have two different demand classes. Orders of the first demand class should be satisfied at the time of the incoming demand, while orders of the second demand class should be satisfied after a given DLT. The model we use is referred to as a single-echelon inventory model. We incorporate a rationing policy that sets a critical stock level, such that when the stock drops below this level, we only help the critical customers and therefore ration the non-critical class. For this system, we replicate the approximation proposed by Koçaǵa and Şen (2007) and replicate their study to test their approximation against a simulation. After conducting this replication, we compose an approximation for a cost function. This cost function consists of 2 types of penalty cost and holding cost. Next, we also test this total cost function approximation against a simulation. From this study, we find that our cost function approximation is quite accurate for most instances. With this cost function, we finally optimize stock levels, and compare the cost for several instances with the cost that follows from our simulation. We show with this comparison that considering cost, minimizing the stock level may often not be the most cost-efficient policy.

In our results, we see that mainly for only varying the non-critical demand event parameter, considerable cuts in cost can be made by actually attaining a higher inventory level. This is especially the case when the critical customers are of the first demand class, but also for the case when the critical customers are of the second demand class to a lesser extent. Of course, these results are dependent on the cost parameters we have chosen for our instances. However, this is a very good indication that choosing the stock based on minimizing the annual cost function, can lead to different stock levels which are more cost efficient.

Regarding the accuracy and robustness of the cost approximation, there are still some improvements to make. We observe that, especially when no fixed service level is required, the cost function can perform quite bad. After looking at the separate cost components, we conclude that this is due to the service level approximation that is used in the stockout cost for the critical class. This approximation performs much worse when the service level gets

lower, which leads to suboptimal stock levels. In future research, incorporating a more accurate service level approximation could very well lead to a better cost approximation. As an extension, considering stochastic DLT might be very interesting and useful. Also, expanding the criticality classes could be very enlightening.
As far as we are aware of, a cost function has not yet been incorporated into a system with rationing as well as DLT. Considering a form of a cost function could even further reduce the cost of an inventory system.

References

De Véricourt, F., Karaesmen, F., & Dallery, Y. (2002). Optimal stock allocation for a capacitated supply system. *Management Science*, *48*(11), 1486-1501.

Deshpande, V., Cohen, M. A., & Donohue, K. (2003). A threshold inventory rationing policy for service-differentiated demand classes. *Management science*, *49*(6), 683-703.

Hariharan, R., & Zipkin, P. (1995). Customer-order information, leadtimes, and inventories. *Management Science*, *41*(10), 1599-1607.

Kleijn, M. J., & Dekker, R. (1999). An overview of inventory systems with several demand classes. *New trends in distribution logistics*, 253-265.

Koçağa, Y. L., & Şen, A. (2007). Spare parts inventory management with demand lead times and rationing. *IIE Transactions*, *39*(9), 879-898.

Kranenburg, A. A., & Van Houtum, G. J. (2005). *Cost optimization in the (S-1, S) lost sales inventory model with multiple demand classes*. Beta, Research School for Operations Management and Logistics.

Simpson Jr, K. F. (1958). In-process inventories. *Operations Research*, *6*(6), 863-873.

Tan, T., Güllü, R., & Erkip, N. (2009). Using imperfect advance demand information in ordering and rationing decisions. *International Journal of Production Economics*, *121*(2), 665-677.

Veinott Jr, A. F. (1965). Optimal policy in a dynamic, single product, nonstationary inventory model with several demand classes. *Operations Research*, *13*(5), 761-778.

Vicil, O. (2022). Cost optimization in the $$(S-1, S) $$(S-1, S) backorder inventory model with two demand classes and rationing. *Flexible Services and Manufacturing Journal*, *34*(1), 101-124.

Wang, Y., Cohen, M. A., & Zheng, Y. S. (2002). Differentiating customer service on the basis of delivery lead-times. *Iie Transactions*, *34*(11), 979-989.

Appendices

Appendix 1

In table 5*, 6* and 7*, we can see table 5, 6 and 7, but now the differences are given in percentages. Note that percentages may appear high, while the cost values may be of multiple digits after the decimal point. Therefore, differences of well over a 1000% can be present in these tables.

				c = 1, n = 2					c = 2, n = 1				
				sc	*sn*	*dc*	*dn*	*h*	*sc*	*sn*	*dc*	*dn*	*h*
λc	*λn*	*S*	*Sc*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*
1	4	5	3	300%	0%	-70%	-8%	-23%	172%	0%	213%	-9%	-28%
2	4	6	3	275%	0%	0%	-14%	-12%	200%	0%	164%	-15%	-15%
3	4	7	3	267%	0%	1696%	-18%	-7%	230%	0%	211%	-19%	-8%
4	4	8	3	206%	0%	120%	-21%	-4%	227%	0%	133%	-21%	-4%
5	4	9	3	182%	0%	0%	9600%	-3%	239%	0%	112%	-24%	-2%
6	4	10	3	176%	0%	100%	-26%	-2%	263%	0%	117%	-26%	-1%
7	4	11	3	161%	1%	100%	-28%	-1%	279%	0%	21%	-27%	-1%
8	4	12	3	143%	0%	100%	-29%	-1%	312%	0%	44%	-28%	-1%
9	4	13	3	133%	-1%	100%	-32%	-1%	340%	0%	61%	-29%	0%
10	4	14	3	128%	0%	100%	-31%	0%	329%	-1%	51%	-29%	0%
11	4	15	3	122%	0%	100%	-32%	0%	391%	0%	51%	-22%	0%
12	4	16	3	108%	0%	0%	-35%	0%	422%	-1%	47%	-33%	0%
2	4	8	1	6352%	-17%	7520%	-29%	-8%	80%	-1%	49%	-11%	0%
3	4	8	2	1973%	-23%	1821%	-27%	-9%	144%	0%	79%	-15%	-1%
4	4	8	3	1500%	-22%	1187%	-29%	-11%	227%	0%	133%	-21%	-4%
5	4	8	4	625%	-16%	644%	-27%	-13%	330%	0%	300%	-29%	-12%
6	4	8	5	422%	-8%	425%	-23%	-16%	469%	0%	200%	-37%	-20%
7	4	8	6	357%	-3%	341%	-20%	-17%	743%	0%	450%	-45%	-27%
8	4	8	7	678%	0%	600%	-52%	-35%	1346%	0%	800%	-52%	-32%

Table 5*: Performance of the separate cost component approximations in percentages for a fixed critical service level of 99% (L = 0.5 and T = 0.1) with all cost π_c, π_n, $\hat{\pi}_c$, $\hat{\pi}_n$ and h to 1 for comparison purposes

				c = 1, n = 2					c = 2, n = 1				
				sc	*sn*	*dc*	*dn*	*h*	*sc*	*sn*	*dc*	*dn*	*h*
λc	*λn*	*S*	*Sc*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*
4	1	5	2	31%	0%	20%	-30%	-3%	101%	0%	23%	-29%	-2%
5	1	6	2	28%	0%	22%	-30%	-2%	118%	0%	18%	-32%	-1%
6	1	7	2	25%	0%	14%	-33%	-1%	134%	-1%	25%	-29%	0%
7	1	8	2	23%	-1%	17%	-35%	-1%	155%	-1%	17%	-33%	0%
8	1	9	2	22%	-1%	13%	-38%	0%	170%	1%	0%	-33%	0%
9	1	10	2	19%	0%	15%	-35%	0%	189%	0%	0%	-33%	0%
5	1	7	1	17%	0%	11%	-30%	0%	83%	0%	0%	-25%	0%
6	1	7	2	25%	0%	14%	-33%	-1%	134%	0%	25%	-29%	0%
7	1	7	3	30%	0%	22%	-41%	-4%	160%	0%	25%	-40%	-2%
8	1	7	4	39%	0%	33%	-52%	-9%	173%	0%	37%	-50%	-6%

Table 6*: Performance of the separate cost component approximations in percentages for a fixed critical service level of 95% (L = 0.5 and T = 0.1) with all cost π_c, π_n, $\hat{\pi}_c$, $\hat{\pi}_n$ and h to 1 for comparison purposes

						c = 1, n = 2					c = 2, n = 1				
						sc	*sn*	*dc*	*dn*	*h*	*sc*	*sn*	*dc*	*dn*	*h*
S	*S_c*	*λ_c*	*λ_n*	*L*	*T*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*	*diff.*
5	2	1	1	1	0.5	180%	0%	14%	-12%	-2%	250%	0%	56%	-15%	-2%
5	2	2	1	1	0.5	90%	0%	21%	-22%	-5%	289%	0%	60%	-26%	-3%
5	2	3	1	1	0.5	61%	0%	27%	-33%	-11%	267%	0%	48%	-34%	-4%
5	2	4	1	1	0.5	50%	0%	38%	-45%	-19%	233%	0%	45%	-41%	-6%
5	2	5	1	1	0.5	44%	0%	50%	-54%	-30%	192%	0%	46%	-48%	-9%
5	2	1	1	1	0.5	180%	0%	14%	-91%	-2%	250%	0%	56%	-15%	-2%
5	2	1	2	1	0.5	417%	0%	20%	-10%	-6%	227%	0%	300%	-15%	-11%
5	2	1	3	1	0.5	643%	0%	145%	-10%	-12%	300%	0%	300%	-14%	-27%
5	2	1	4	1	0.5	813%	0%	139%	-9%	-20%	388%	0%	550%	-13%	-46%
5	2	1	5	1	0.5	1063%	0%	289%	-9%	-30%	550%	0%	850%	-12%	-61%
14	3	10	4	0.5	0.10	128%	0%	100%	-31%	0%	329%	-1%	51%	-29%	0%
14	3	10	4	0.5	0.20	156%	0%	45%	-28%	0%	733%	-2%	38%	-33%	0%
14	3	10	4	0.5	0.30	170%	-1%	43%	-27%	-1%	862%	0%	0%	-22%	0%
14	3	10	4	0.5	0.40	133%	0%	1%	-33%	0%	0%	0%	0%	0%	0%
14	3	10	4	0.5	0.50	40%	28%	-45%	0%	0%	0%	0%	0%	0%	0%

Table 7*: Performance of the separate cost component approximations in percentages for varying parameters (L = 0.5 and T = 0.1) with all cost π_c, π_n, $\hat{\pi}_c$, $\hat{\pi}_n$ and h to 1 for comparison purposes