WORKING PAPER L-2006-03

WP

Christian Larsen

Computation of order and volume fill rates for a base stock inventory control system with heterogeneous demand to investigate which customer class gets the best service

Logistics/SCM
Research Group

Computation of order and volume fill rates for a base stock inventory control system with heterogeneous demand to investigate which customer class gets the best service

Christian Larsen
Department of Business Studies
Logistics/SCM Research Group
Aarhus School of Business
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Abstract We consider a base stock inventory control system serving two customer classes whose demands are generated by two independent compound renewal processes. We show how to derive order and volume fill rates of each class. Based on assumptions about first order stochastic dominance we prove when one customer class will get the best service. That theoretical result is validated through a series of numerical experiments which also reveal that it is quite robust.
Keywords Base stock policy, service measures, two customer classes, compound renewal processes.

1. Introduction

Most literature on inventory control in presence of stochastic demand assumes the demand process is homogeneous. In well-established textbooks, like Silver et al. (1998) and Zipkin (2000) no mention is made on how to handle inventory control in presence of heterogeneous demand, that is when the demand process is the aggregation of demands of several different customer classes. The earliest reference on inventory control in presence of several customer classes seems to be Veinott (1965) who developed a (time-dependent) base stock policy in a periodic review framework. All other references on the subject, that the author has been able to collect, all assume that it is possible to discriminate among the customer classes by introducing a rationing policy, to be applied in cases where the physical inventory is critically low. See Topkis (1968), Evans (1968), Kaplan (1969) and Frank et al. (2003) for periodic review models and Nahmias and Demmy (1981), Dekker et al. (1998 and 2002), Melchiors et al. (2000), Melchiors (2003) and Despande et al. (2003) for continuous review models. Our analysis rests on the premises there is no discrimination among the customer classes. We will give two reasons for our deviation from the main stream direction in this subject. The first address the issue about reputation and sensible business behavior. Of course if the item of concern is a very vital one, like a spare part and the potential customers are nuclear power plants, oil refineries or similar where an unfilled demand might imply a total stop of production with severe societal consequences, it would be very sensible to apply a rationing policy. But if the item of concern is a more everyday commodity, and (maybe more importantly) if the difference among customer classes is maybe distinct but not that profound, we find it less obvious to apply a rationing policy. For instance, the author has been involved in an intensive work of improving the inventory control policies for a large Danish manufacturer of equipment for refrigeration and air conditioning. In the author's discussions with senior staff members of logistics it was never an issue to invoke rationing policies, even if there were some differences among the different customers. These are industrial companies or wholesalers, and the difference can sometimes be seen in the demand patterns of the respective customer classes. In general we find it hard to believe that any sensible business minded person would pretend to a potential customer that he has no items in stock, in case he actually has. If he really applied such a policy he might very well lose customers in the long run, in case the low-prioritized customers found out or rumours spread around (likely to happen unless the organization is very tight) that their wishes had not been fulfilled. The second reason addresses the concept of postponement, which is a very popular phrase in the world of logistics. It may be *geographical postponement*. That is, instead of having an inventory to serve each geographical dispersed market, one has a central inventory from which all markets are served. For a good illustration see the case Risk Pooling in Simchi-Levi et al. (2003; pp 64-66). It may also be *product postponement*. That is, one applies a strategy of delayed product differentiation which means that instead of having several inventories of end-products, all with small variations of the same basic product, one only holds an inventory of the basic product, and then upon the receipt of a customer order does the final customization. If the demands of several markets/end products are pooled, and each market/end product has distinct characteristics with respect to their demand patterns, then the centralized inventory serving all markets/end products will be facing heterogeneous demands. There are many good reasons for implementing a postponement strategy; the most common argument is reduction in inventory investments. However, we find it less obvious that when taking such an initiative it is also on the premise that some markets/end-products should be judged less important. We would imagine that all markets/end-products would have the same priority from scratch. However, it might call for the incorporation of a rationing policy into the inventory control policy, if it turned out that there was a large discrepancy among the service levels offered the different markets/end-products - for instance if a particular market/end-product has a too low service level, So in that sense the analysis offered in this paper could be valuable tool for the management before contemplating the (maybe painful) decision to discriminate among its customers.

In our paper we focus on a situation where there are two customer classes both served from the same inventory. The inventory is operated using a base stock policy where all unfilled demands are being backlogged. We also assume that all replenishment orders have the same fixed lead-time L. For customer class j ($j=1,2$) the inter-arrival times between demand instances are independent and identical distributed as a positive continuous random variable T_j with mean $1/\Gamma_j$. The order sizes of the customers of class j

are independent and identical distributed as a positive integer valued random variable X_j. We assume independence between the two compound renewal processes. In order to complete the description of the base stock inventory policy, it is assumed that if a customer upon arrival can not get his full order, then he gets as much as possible and the remaining part is backlogged. Furthermore, it is assumed that for all customers irrespective of their class, an attempt is made to serve them immediately upon arrival. In our analysis we will focus on the two service measures, o*rder fill rate* and v*olume fill rate*, in the following when convenient abbreviated to *OFR* and *VFR*, respectively. The latter is often just denoted fill rate, in Silver et al. (1998; p. 245) it is also called the P_2 service measure. It measures the fraction of the total demand (aggregated over all customer orders) that is filled immediately. The former has received less attention in the textbook literature on inventory control. It measures the fraction of customer orders, where the whole order is filled immediately upon receipt of the order. The reason for our interest in that service measure is due to the fact that exactly that service measure is applied at the Danish company mentioned above. In the papers by Song (1998) and Hausman et al (1998) they analyse this service measure though in a multi-item setting, where we focus on a single-item setting only.

We assume that the base stock level S is set in order to achieve the same pre-specified level β for each customer class of the chosen service measure. This means that S should be set as a solution to

$$\min\{S : XFR_j(S) \geq \beta, j = 1,2\}, \qquad (1)$$

where $XFR_j(S)$ is the fill rate of customer class j, and where the first letter X is either O or V depending on which service measure is applied. This implies that the organization in charge of the inventory should be able to monitor the service levels offered to the different customer classes. This represents also a prerequisite if the organization is considering an introduction of a rationing policy. Alternatively, one would set the base stock level S in order to achieve an average service level. In this case of course the base stock level will be smaller than in (1). The important thing to note is that irrespective of which of the methods is applied, the two customer classes will not in general get the same service levels. Therefore it would be of interest to explore which customer class benefits most. It is interesting to know, because it gives the organization in charge of the inventory an idea from which customer class it gets its most satisfied/dissatisfied customers. It also gives an indication to which class (which might be a single large customer) it should invest in demand planning activities to improve the service to the customer class (which we believe is a more fruitful and proactive way than to introduce a rationing policy). We are able to formally prove that if the demand process of class *2* is a compound Poisson process while the demand process of class *1* is a compound Erlang process then if X_2 displays first degree stochastic dominance over X_1, class *1* will get the best service irrespective of which of our service measures are applied and irrespective of the value of β. It should be noted that we are able to prove this result under a more general condition; however the condition stated here emerges as a very operational and easy to understand sub-case of our general result. So the main contributions of our paper are first to develop mathematical expressions for the volume and order fill rates and to prove our theoretical result. Furthermore, by conducting a series of numerical experiments, assuming that inter-arrival times are Erlang distributed and customer order sizes are (delayed) negative binomially distributed, we also find our theoretical result is fairly robust. That is, the two assumptions mentioned above can be relaxed rather much and the result that class *1* gets the best service still holds.

There is much literature examining the impact of pooling strategies on inventory operations, see Garg and Lee (1999) and Aviv and Federgruen (1999) and the many references therein. But none of them seem to focus on heterogeneity neither in demand processes nor in delivered service. In the queuing theory literature the subject service management in the presence of heterogeneity among customer request has been analyzed very thoroughly; see among others Mendelson and Whang (1992) and Van Miegham (2000). But these analyses are often conducted on single server models (which corresponds to base stock systems with a base stock level of one unit). Moreover, here the main issue is most often how to price requests from different customer classes and explore whether the pricing scheme is incentive-compatible, which is outside the scope of our analysis.

In Section 2 we start making a mathematical derivation of our service measures in case of a single customer class. This is a brief recapitalization of the results in Larsen and Thorstenson (2006). We then show how to extend the result in the case of two customer classes. We finish Section 2 by proving our general theoretical result. In Section 3 we consider the case where T_j, $j=1,2$ are both Erlang distributed and X_j, $j=1,2$ are both (delayed) negative binomial distributed, for which we in Section 4 do numerical analyses. These concern the validity of our model, the robustness of our theoretical result and suggest some managerial insights that can be drawn from our model. Due to the fact that the Erlang and the (delayed) negative binomial distributions can represent a lot of different shapes, we do not find it limiting for our numerical contribution that we only work with these particular distributions. Finally in Section 5 we give some concluding remarks.

2. Derivations of the fill rate service measures

First we will derive the two fill rate service measures for the case where there is only customer class. We put a superscript *A* (*A* for alone) on the fill rate service measures in order to distinguish from the case of two customer classes. Let in the following τ_j be an arrival epoch of a customer of class *j*. Let the random variable $\overline{D}_j^L$ represent the aggregate demand of customer class *j* in the time interval $[\tau_j - L, \tau_j)$. Then a customer at an arrival epoch will see a net inventory *S-x* with probability $P(\overline{D}_j^L = x)$. Because he will get immediate delivery of his whole order if $X_j \le S - x$, the order fill rate service measure of customer class *j* in the stand-alone setting is

$$OFR_j^A(S) = \sum_{x=0}^{S-1} P(\overline{D}_j^L = x) P(X_j \le S - x) \tag{2}$$

It could also have been rewritten as

$$OFR_j^A(S) = \sum_{x=1}^{S} P(\overline{D}_j^L \le S - x) P(X_j = x) \tag{3}$$

It appears in this form in Song (1998; Equation 4), where it is denoted an item fill rate (note however that in her definition a demand is a customer order) and in Rosling (2002; Equation 5). It can also be stated as

$$OFR_j^A(S) = P(\overline{D}_j^L + X_j \le S) \tag{4}$$

If we instead focus on the volume fill rate, the customer upon arrival seeing a net inventory $S - x$, will get an expected fill $E[min\{S\text{-}x,X_j\}]$ if $x \le S$ and 0 otherwise. Therefore the volume fill rate is

$$VFR_j^A(S) = \frac{\sum_{x=0}^{S-1} P(\overline{D}_j^L = x) E[\min\{S - x, X_j\}]}{E[X_j]} \tag{5}$$

It can also be stated as

$$VFR_j^A(S) = \frac{E[\min\{\max\{S - \overline{D}_j^L, 0\}, X_j\}]}{E[X_j]} \tag{6}$$

We will now derive the fill rates of customer class j when the inventory serves both classes. Again we consider an arrival epoch τ_j of a customer of class j. Note that when considering customer class j then $3\text{-}j$ is the index of the other customer class. Let $\widetilde{D}_{3-j}^L$ be the aggregate demand of this other customer class in the time interval $[\tau_j - L, \tau_j)$. Note that concerning the compound renewal demand process of class $3 - j$, time point τ_j can be considered as a randomly chosen time point. Applying the same logic as in the case of a single customer class we therefore conclude that

$$OFR_j(S) = \sum_{x=0}^{S-1} P(\overline{D}_j^L + \widetilde{D}_{3-j}^L = x) P(X_j \le S - x) \tag{7}$$

and

$$VFR_j(S) = \frac{\sum_{x=0}^{S-1} P(\overline{D}_j^L + \widetilde{D}_{3-j}^L = x) E[\min\{S-x, X_j\}]}{E[X_j]}. \qquad (8)$$

Exploiting that random variables $\overline{D}_j^L$ and $\widetilde{D}_{3-j}^L$ are independent and changing the order of summation we get

$$\begin{aligned} OFR_j(S) &= \sum_{x=0}^{S-1}\left[\sum_{y=0}^{x} P(\overline{D}_j^L = x-y) P(\widetilde{D}_{3-j}^L = y)\right] P(X_j \le S-x) \\ &= \sum_{y=0}^{S-1}\left[\sum_{x=0}^{S-y-1} P(\overline{D}_j^L = x) P(X_j \le S-y-x)\right] P(\widetilde{D}_{3-j}^L = y) \\ &= \sum_{y=0}^{S-1} OFR_j^A(S-y) P(\widetilde{D}_{3-j}^L = y). \qquad (9) \end{aligned}$$

Similarly

$$VFR_j(S) = \sum_{y=0}^{S-1} VFR_j^A(S-y) P(\widetilde{D}_{3-j}^L = y) \qquad (10)$$

Thus for any of the two fill rates, the fill rate of class *j* in pooled set-up appears as the fill rate function of the stand-alone set-up convoluted to the lead-time demand probability distribution of the other class. As a point of validation we also note that if $P(T_{3-j} > L) = 1$ then $P(\widetilde{D}_{3-j}^L = 0) = 1$ and therefore $OFR_j(S) = OFR_j^A(S)$ and $VFR_j(S) = VFR_j^A(S)$ which also should be obvious, because if the demand of the other customer class is rather stable and low-frequent then the service offered to customer class *j* will be unchanged whether it is from a stand-alone inventory or from a pooled inventory. Again the service measures can also be written as

$$OFR_j(S) = P(\overline{D}_j^L + \widetilde{D}_{3-j}^L + X_j \le S) \qquad (11)$$

and

$$VFR_j(S) = \frac{E[\min\{\max\{S - \overline{D}_j^L - \widetilde{D}_{3-j}^L, 0\}, X_j\}]}{E[X_j]} \qquad (12)$$

Before we prove our theoretical result, we first define stochastic dominance of first degree. For a definition of various concepts of stochastic dominance, see a standard text book in Finance like Copeland and Weston (1979; pp 78-82). For a more research oriented exposition on stochastic dominance, see Levy (1992). We deliberately keep our exposition within the class of integer valued random variables.

Definition: Let W and V be two integer valued random variables. W displays first degree stochastic dominance over V if for any integer q it holds that $P(V \le q) \ge P(W \le q)$.

Note also that when W displays first degree stochastic dominance over V it holds that $E[W] \ge E[V]$. We are now able to prove

Theorem 1: Assume that X_2 displays first-degree stochastic dominance over X_1 and $\widetilde{D}_1^L + \overline{D}_2^L$ displays first-degree stochastic dominance over $\overline{D}_1^L + \widetilde{D}_2^L$. Then for any of the fill rate service measures and any levels of the service requirement β it holds that customer class *1* gets the best service.

Proof: The result for the case order fill rate follows immediately from (11). The inequality

$$\frac{E[\min\{\max\{S - \overline{D}_1^L - \widetilde{D}_2^L, 0\}, X_1\}]}{E[X_1]} \ge \frac{E[\min\{\max\{S - \overline{D}_2^L - \widetilde{D}_1^L, 0\}, X_2\}]}{E[X_2]}$$

can be rewritten to

$$E[\min\{X_2 \max\{S - \overline{D}_1^L - \widetilde{D}_2^L, 0\}, X_1 X_2\}] \ge E[\min\{X_1 \max\{S - \overline{D}_2^L - \widetilde{D}_1^L, 0\}, X_1 X_2\}]$$

where we have exploited that X_{3-j} and $\min\{\max\{S - \overline{D}_j^L - \widetilde{D}_{3-j}^L, 0\}, X_j\}$ are independent, $j=1,2$. The conclusion now follows because $X_2 \max\{S - \overline{D}_1^L - \widetilde{D}_2^L, 0\}$ displays first-degree stochastic dominance over $X_1 \max\{S - \overline{D}_2^L - \widetilde{D}_1^L, 0\}$.

It should be easy to detect when X_2 displays first-degree stochastic dominance over X_1. Later we explain how to do it for the class of (delayed) negative binomial distributions. It is less obvious to detect when $\widetilde{D}_1^L + \overline{D}_2^L$ displays first-degree stochastic dominance over $\overline{D}_1^L + \widetilde{D}_2^L$. From now on we will focus on the situation where random variables T_j, $j=1,2$ are Erlang distributed with k_j phases, where k_j is a positive integer. It is then obvious that $\widetilde{D}_L^j$ displays first-degree stochastic dominance over $\overline{D}_j^L$, and also that they are identical if demands of customer class j occur after a compound Poisson process, that is when $k_j = 1$. Therefore the following corollary emerges from Theorem 1

Corollary 1: Assume that X_2 displays first-degree stochastic dominance over X_1 and the demand process of customer class *2* is a compound Poisson process while the demand process of customer class *1* is a compound Erlang process. Then for any of the fill rate service measures and any levels of the service requirement β it holds that customer class *1* gets the best service.

The managerial insight that can be thought from Corollary 1 is that in order to get the highest service level compared to other customer classes one must aim at **small orders** and **regularity**. Later in Section 4 we try to illustrate that numerically.

Theorem 1 and Corollary 1 are our theoretical results and their robustness will among others be examined in Section 4. But before that we will further specify our service measure formulas when inter-arrival times are Erlang distributed and order sizes are (delayed) negative binomial distributed.

3. The set-up for the numerical investigation

3.a. Inter-arrival times

We will expand further on the derivation of the order fill rate expressions when random variables T_j, $j=1,2$ are Erlang distributed with k_j phases, where k_j is a positive integer. Let $\lambda_j = k_j\Gamma_j$, that is the intensity of the underlying Poisson process. Let $X_j(m)$ denote the aggregate demand of in all m customers of class j. Note that $X_j(1)$ is identical to X_j and $P(X(0) = 0) = 1$. It then follows, see also Rosling (2002, p. 1010), that

$$P(\overline{D}_j^L = x) = \sum_{m=0}^{\infty} P(\overline{N}_j^L = m) P(X_j(m) = x) \tag{13}$$

where the random variable $\overline{N}_j^L$ denotes the total number of customer arrivals of class j in the time interval $[\tau_j - L,\ \tau_j)$. Its probability distribution is given in Cox (1962; p. 37). Therefore (note that we explicitly have $P(X_j \geq 1) = 1$ and thus $P(X_j(m) \geq m) = 1$ for $m \geq 1$)

$$P(\overline{D}_j^L = x) = \begin{cases} e^{-\lambda_j L} \sum_{i=0}^{k_j - 1} \frac{(\lambda_j L)^i}{i!} & x = 0 \\ e^{-\lambda_j L} \sum_{m=1}^{x} P(X_j(m) = x) \sum_{i=0}^{k_j - 1} \frac{(\lambda_j L)^{i+mk}}{(i+mk)!} & x = 1, 2, \ldots \end{cases} \tag{14}$$

In a similar fashion the probability distribution of $\widetilde{D}_{3-j}^L$ can be determined, see also Rosling (2000, p. 1010) as

$$P(\widetilde{D}_{3-j}^L = x) = \sum_{m=0}^{\infty} P(\widetilde{N}_{3-j}^L = m) P(X_{3-j}(m) = x) \tag{15}$$

where the random variable $\widetilde{N}_L^{3-j}$ denotes the total number of customers of class 3-j in the time interval $[\tau_j - L, \tau_j)$. Its probability distribution is given in Cox (1962; p. 39). Therefore

$$P(\widetilde{D}_j^L = x) = \begin{cases} e^{-\lambda_j L} \sum_{i=0}^{k_j-1} \frac{k_j - i}{k_j} \frac{(\lambda_j L)^i}{i!} & x = 0 \\ e^{-\lambda_j L} \sum_{m=1}^{x} P(X_j(m) = x) \sum_{i=1-k_j}^{k_j-1} \frac{k_j - |i|}{k_j} \frac{(\lambda_j L)^{i+mk}}{(i+mk)!} & x = 1, 2, \ldots \end{cases} \tag{16}$$

3.b Customer order-size distributions

Our choice of distribution (temporarily ignoring class index j) to describe order-sizes is a (delayed) negative binomial distribution with probability parameter ρ, with $0 < \rho < 1$, and shape parameter s, with $s > 0$. Its probability distribution is

$$P(X = x) = \frac{(s + x - 2)!}{(x-1)!(s-1)!} (1-\rho)^s \rho^{x-1} \qquad x = 1, 2, \ldots \tag{17}$$

It is just a standard negative binomial distribution, as seen in most textbooks on probability theory, shifted to the right by 1. We find it most convenient, and it makes more sense in a real-life sense when it comes to data collection to have the restriction that any customer order is positive. Note that s needs not be integer valued since the factorials in (17) can then be interpreted as gamma functions, see Zipkin (2000; p 452). The (delayed) negative binomial distribution can attain a lot of shapes, therefore we do not find it that limiting for the findings we do in Section 4, that we exclusively work with that distribution. Because Theorem 1 concerns stochastic dominance of first degree, we will establish when one (delayed) negative binomial distribution displays first-degree stochastic dominance over another (delayed) negative binomial distribution. This is stated in the following lemma

Lemma 1: Let X_j, $j = 1,2$ be (delayed) negative binomial random variables with probability parameter ρ_j and shape parameter s_j, with $\rho_1 \leq \rho_2$ and $s_1 \leq s_2$. Then X_2 displays first-degree stochastic dominance over X_1.

Proof: First consider the case where $s_1 = s_2 = 1$, that is we consider two (delayed) geometric distributions. The result then follows from $P(X_j \geq q) = \rho_j^{q-1}$. The result, of course, also holds if we instead considered two standard geometric distributions. When $s > 1$, then X_j can be written as 1 plus the sum of s_j independent standard geometric distributions, each with probability parameter ρ_j. The result then follows by combining the result of $s_1 = s_2 = 1$ and the summation of positive random variables preserves stochastic dominance.

4. Numerical results

This section is organized around the following three topics: validity, robustness and managerial insights. First we consider validity of our model which concerns the following question:

Question 1: Can an arrival point of one demand process be considered as a random time point with respect to the other?

It may seem odd to pose this question, given that the positive answer to it is already accounted for via the mathematical derivations done in the previous sections. We want to remark that most contributions on inventory control, also the cited references on continuous-review rationing policies, assume that a demand process is a (compound) Poisson process. In that respect our exposition is more advanced, as we model a demand process as a compound renewal process. So also from a technical point of view our contribution is outside mainstream. Due to this we feel it might be challenged whether an arrival point in one renewal process can be considered a randomly chosen time point with respect to the other renewal process. This addresses the validity of the probability distributions $P(\widetilde{D}_L^j = x)$, $j=1,2$, in particular if the phase parameters k_1 and k_2 are large. Therefore we would like numerically to verify that the answer to Question 1 is positive. For various values of k_1 and k_2, we found the optimal base stock level for each service measure by optimization model (1). Then we simulated the base stock system and collected data on the service measure in order to compare with the computed values. In the following two tables the results for each service measure are summarized.

<Tables 1 and 2 about here>

As can be seen most often there is no significant difference between computed and simulated values. Several other experiments not reported in this paper give the same conclusion. Therefore we find that we have given an affirmative answer to Question 1. Accordingly, we do not in the following report any simulation results.

We now address the issue of robustness of Corrolary1. Since Corollary 1 consists of two assumptions we split our analysis into two. First we explore the assumption that the demand process of customer class *2* shall be compound Poisson.

Question 2: Given stochastic dominance of first degree of X_2 over X_1, how much can parameter k_2 be increased so that the result of Corollary 1 still holds?

We first look at the dataset also examined in Question 1, where X_2 dominates X_1 quite strongly, that is with $(s_1, \rho_1) = (1, 0.6)$ and $(s_2, \rho_2) = (2, 0.8)$. Here we let $L = 10$ and $\Gamma_1 = \Gamma_2 = 1.25$. We made two analyses where we had $k_1 = 1$ (which makes $\overline{D}_1^L + \widetilde{D}_2^L$ display stochastic dominance of first degree over $\widetilde{D}_1^L + \overline{D}_2^L$ which is the opposite of the assumption of Corollary 1) and $k_1 = 3$. For each of the analyses we then varied k_2 from *2* to *10*. The results are given in Tables *3* and *4*. Because it is not feasible to compute service levels up to a prohibitive large value of *S*, we have in the following stopped the computation of service levels when these become well above 0.995.

<Tables 3 and 4 about here>

For both $k_1 = 1$ and $k_1 = 3$ it turned out that for all considered values of *S* it holds that for the order fill rate service measure customer class 1 gets the best service independent of the value of k_2. It is almost also the case for the service measure volume fill rate, only for very small base stock levels *S*, which give rise to a service level of almost zero, customer class 2 will get the best service. In particular we find it interesting that the result of Corollary 1 holds in a case where $k_2 > 1$ and $k_1 = 1$, because here the second assumption of Theorem 1 is completely violated. We believe the reason is that X_2 exhibits very strong dominance over X_1. We therefore examined another case for the order distributions, namely with $(s_1, \rho_1) = (2, 0.7)$ and $(s_2, \rho_2) = (2, 0.8)$. The values of L, Γ_1 and Γ_2 are as before and we first let $k_1 = 1$ and then $k_1 = 3$.

<Tables 5 and 6 about here>

For $k_1 = 3$, the conclusion is as above, namely that for all realistic values of the base stock level, customer class 1 will get the best order and volume fill rate service levels irrespective of the value of k_2. For $k_1 = 1$ wee see that when $k_2 \geq 3$, then customer class *2* gets the best volume fill rate service. In particular the result of the case $k_2 = 3$ in Table *6* is interesting. Here we see that it is crucial which service measure is applied in order to state which customer class benefits most. In general it is the result of Corollary 1 with respect to the order fill rate that is least sensitive to the assumption $k_2 = 1$.

We now assume the second assumption of Corollary 1 to be fulfilled, that is $k_2 = 1$, but now X_2 does not display stochastic dominance of first degree over X_1. A way to force the result of the Corollary through, offsetting the lacking dominance of X_2 over X_1, is to have $\widetilde{D}_L^1 + \overline{D}_L^2$ display strong dominance over $\overline{D}_1^L + \widetilde{D}_2^L$ by letting parameter k_1 increase. Therefore we now concern:

Question 3: Given there is no stochastic dominance of first degree of X_2 over X_1, but $k_2 = 1$ how much do we need to increase k_1 in order to force through the result of Corollary 1?

Here we examine a case where $(s_1, \rho_1) = (1, 0.8)$ and $(s_2, \rho_2) = (2, 0.6)$, which means that $P(X_1 \leq x) \leq P(X_2 \leq x)$, except for $x = 1$ and 2. Again we let $L = 10$ and $\Gamma_1 = \Gamma_2 = 1.25$. Our results are summarized in Table 7, prepared very similar to Tables 3 – 6.

<Table 7 about here>

We see that although X_1 almost dominates X_2 we do not need to impose much regularity into the demand process of customer class 1 (that is to increase k_1) in order to regain the result of Corollary 1.

We finally address some analyses that would be of managerial interest

Question 4: How large can the difference in service levels be between the two customer classes?

First of all we should note that it can be as high as $1 - \beta$ when using optimization approach *(1)*. Consider the following example: let the phase parameters k_j, $j=1,2$ be very large, and let $\Gamma_1 = \Gamma_2$, so that in reality we have a deterministic arrival pattern where customers of each class arrive after rotational (and predictable) schemes. If we further have $P(X_2 > X_1) = 1$ and L is very small, then it is easy to see that the solution to (1) is a base stock level giving customers of class *1* a service level of *1* and customers of class *2* a service level of β. For this extremely artificial example it would of course be unwise to apply a constant base stock level. Instead, it should have two values, depending on the class belonging of the last arriving customer. We now investigate Question 2 numerically using a dataset, where the idea of having a constant base stock level is more meaningful. We consider a dataset where the assumptions of Corollary 1 are fulfilled. That is, we took the same customer order distributions as in Question1 and we let $k_2 = 1$. As in Question 1 we have $\beta = 0.9$. In order to get some insight into the importance of the arrival intensity as well as whether it matters if the demand pattern of customer class 1 is compound Poisson or not, we let the three remaining parameters be as specified by Table *8*. The results of our investigation are given in Table *9*.

<Tables *8* and *9* about here>

Because the assumptions of Corollary 1 are fulfilled we always see customer class 1 gets the best service. We see that the optimal S values are almost the same irrespective of which service measure is applied. It also turns out that the value of k_1 does not seem to have much significance when the optimal base stock level is determined. Returning to Question 4 we see that when the lead-time is small, the arrival pattern in customer class *1* is less erratic and the relative frequency of order arrivals of customer class *1* is high, then the difference in service levels can be about *5* percentage points for both service measures. In general the differences in service levels are a little bit higher for the order fill rate service measure than for the volume fill rate service measure. We find there is an interesting observation that can be drawn from this analysis. Namely that if customers of class 1 are “well behaved” in the sense they arrive after a regular pattern ($k_1 > 1$) and their orders are generally smaller than their counterparts of customer class *2*, then they will also be rewarded in terms of better service. This could be a “carrot” to be employed in some demand planning or customer relationship activities, for instance if customer class *1* consists of a single large customer. In the following we expand a little bit further on that. Assume that the result in row no. *3* of Table *9* represents the end result after some customer relationship activities have taken place by which the customer has changed his demand pattern. We assume that it could have followed two traces: one where the focus has been on achieving regularity and one where the focus has been on achieving smaller orders. The result of the first trace can be seen from Table *9* comparing row no. *12* to row no. *3*. Here we see a rather small increase in service levels when changing the regularity parameter k_1 from *1* to *3*. If we followed the other trace and assumed that the probability parameter ρ_1 is unchanged but originally had a higher value of s_1 we get the results of Table *10*. Note that in order to have the same demand volume we also adjust the arrival intensity Γ_1.

<Table *10* about here>

At least for the order fill rate service measure a significant improvement can be seen when going from $s_1 = 4$ or *5* and to $s_1 = 1$, maybe not that surprising when considering the definition of the order fill rate. But there is also an improvement when considering the volume based fill rate service measure though it is smaller. Supported from some other numerical analyses it seems like what matters most in terms of getting improved service is to change ones demand patterns in such a way that more frequently smaller orders are dispatched, rather than achieving more regularity.

5. Concluding remarks

We have shown how to derive two important customer-oriented service measures for a base stock system with heterogeneous demand, in our case two customer classes with distinct characteristics. The analysis could have been extended to several customer classes. Through our analysis we are able to predict which customer class will get the best service. This information could be very valuable for the management of the inventory system; because the management would then have knowledge about in which segments of their market they can expect to find the most satisfied/dissatisfied customers. It can therefore be a useful tool in order to direct scarce resources for customer relation activities to the right segment of the market. Our numerical investigations reveal that the conditions for Theorem 1 and Corollary 1 are quite robust. It can be seen from (11) that in order for customer class 1 to get the best order fill rate, it only requires $\tilde{D}_1^L + \bar{D}_2^L + X_2$ to display first order stochastic dominance over $\tilde{D}_2^L + \bar{D}_1^L + X_1$. So obviously, the conditions stated in this paper, for a customer class to get the best service, can be sharpened somewhat. It is of course a subject for further research.

Acknowledgement

Part of this research was done during a stay as Visiting Scholar at Technishe Universiteit, Eindhoven, The Netherlands. The stay was partly supported by a grant from Otto Moensteds Fond.

References

Aviv Y., Federgruen A: The benefits of design for postponement. In Quantitative models for supply chain management (editors Tayur S, Ganeshan R, Magazine M), Kluwer, 1999.

Copeland T.E., Weston J.F.: Financial theory and corporate policy. Addison Wesley, 1979.

Cox D.R.: Renewal theory. Methuen, London, 1962.

Despande V., Cohen M.A., Donohue K.: A threshold inventory rationing policy for service differentiated demand classes. **Management Science**, Vol 49, 2003, pp. 683-703.

Dekker R., Kleijn M.J., de Rooij P.J.: A spare parts stocking policy based on equipment criticality. **International Journal of Production Economics**, Vol 56-57, 1998, pp. 69-77.

Dekker R., Hill R.M., Kleijn M.J., Teunter R.H.: On the *(S-1,S)* lost sales inventory model with priority demand classes. **Naval Logistics Research**, Vol 49, 2002, pp. 593-610.

Evans R.V.: Sales and restocking policies in a single item inventory system. **Management Science**, Vol 14, 1968, pp. 463-472.

Garg A., Lee H.L.: Managing product variety: an operations perspective. In Quantitative models for supply chain management (editors Tayur S, Ganeshan R, Magazine M), Kluwer, 1999.

Hausman W.H., Lee H.L., Zhang A.X.: Joint demand fulfillment probability in a multi-item inventory system with independent order-up-to policies. **European Journal of Operational Research**, Vol 109, 1998, pp. 646-659.

Kaplan A.: Stock rationing**. Management Science**, Vol 15, 1969, pp. 260-267.

Larsen C., Thorstenson A.: A comparison between the order and the volume fill rates for a base-stock inventory control system under a compound renewal demand process. Working Paper L-01, Logistics/SCM Research Group, Department of Accounting, Finance and Logistics, Aarhus School of Business, 2006. http://www.hha.dk/afl/wp/log/L_2006_01.pdf

Levy H.: Stochastic dominance and expected utility: survey and analysis. **Management Science**, Vol 38, 1992, pp 555-593.

Melchiors P.: Restricted time-remembering policies for the inventory rationing problem. **International Journal of Production Research**, Vol 81-82, 2003, pp. 461-468.

Melchiors P., Dekker R., Kleijn M.J.: Inventory rationing in an *(s,Q)* inventory model with two demand classes and lost sales. **Journal of the Operational Research Society**, Vol 51, 2000, pp. 111-122.

Mendelson H., Whang S.: Optimal incentive-compatible priority pricing for the M/M/1 queue. **Operations Research**, Vol 38, 1990, pp. 870-883.

Rosling K.: Inventory cost rate functions with nonlinear shortage costs. **Operations Research**, Vol 50, 2002, pp. 1007-1017.

Silver E.A., Pyke D.F., Petersen R.: Inventory management and production planning and scheduling. John Wiley, 3rd edition, 1998.

Simchi-Levi D., Kaminski P., Simchi-Levi E.: Designing & managing the supply chain: concepts strategies and case studies. McGraw-Hill, 2nd edition, 2003.

Song J-S.: On the order fill rate in a multi-item base-stock inventory system. **Operations Research**, Vol 46, 1998, pp. 831-845.

Topkis D.M.: Optimal ordering and rationing policies in a nonstationary dynamic inventory model with *n* demand classes. **Management Science**, Vol 15, 1968, pp. 160-176.

Van Mieghem J.A.: Price and service discrimination in queuing systems: incentive compatibility of $Gc\mu$ scheduling. **Management Science**, Vol 46, 2000, pp. 1249-1267.

Veinott A.F.: Optimal policy in a dynamic, single product, nonstationary inventory model with several demand classes. **Operations Research**, Vol 13, 1965, pp. 761-778.

Zipkin P.H.: Foundations of inventory management. McGraw-Hill, Boston, USA, 2000.

Tables

k_1	k_2	S	$OFR_1(S)$ computed	$OFR_1(S)$ simulated	$OFR_2(S)$ computed	$OFR_2(S)$ simulated
2	2	195	92.48	92.42 (0.11)	90.35	90.27 (0.13)
2	4	188	92.11	92.05 (0.10)	90.17	90.11 (0.13)
2	6	186	92.19	92.19 (0.10)	90.37	90.36 (0.13)
2	8	185	92.26	92.25 (0.11)	90.50	90.51 (0.08)
4	2	194	92.38	92.32 (0.08)	90.07	90.01 (0.14)
4	4	188	92.44	92.39 (0.09)	90.39	90.30 (0.10)
4	6	185	92.11	92.00 (0.12)	90.07	89.95 (0.13)
4	8	184	92.18	92.13 (0.09)	90.20	90.14 (0.09)
6	2	194	92.47	92.35 (0.09)	90.12	89.98 (0.13)
6	4	188	92.55	92.50 (0.11)	90.46	90.39 (0.14)
6	6	185	92.23	92.18 (0.11)	90.15	90.09 (0.09)
6	8	184	92.31	92.26 (0.14)	90.29	90.25 (0.13)
8	2	194	92.51	92.49 (0.09)	90.15	90.14 (0.12)
8	4	188	92.61	92.52 (0.11)	90.49	90.36 (0.12)
8	6	185	92.29	92.24 (0.11)	90.19	90.14 (0.13)
8	8	184	92.38	92.33 (0.11)	90.33	90.30 (0.14)

Table 1: **Comparison between computed and simulated service levels when the service measure is order fill rate**. X_1 and X_2 are both (delayed) negative binomial distributed with probability parameter $\rho_1 = 0.6$ and $\rho_2 = 0.8$ and shape parameters $s_1 = 1$ and $s_2 = 2$, respectively. The other parameter values are $\Gamma_1 = \Gamma_2 = 1.25$, $L = 10$ and $\beta = 0.9$. In the two columns with simulated values, are also stated in parenthesis: the half length of the *95%* confidence interval (*10* replications with run-length *100000* time units with the initial condition that the net-inventory is equal to the base stock level).

k_1	k_2	S	$VFR_1(S)$ computed	$VFR_1(S)$ simulated	$VFR_2(S)$ computed	$VFR_2(S)$ simulated
2	*2*	*193*	*91.73*	*91.67 (0.11)*	*90.29*	*90.22 (0.17)*
2	*4*	*186*	*91.21*	*91.15 (0.14)*	*90.11*	*90.03 (0.15)*
2	*6*	*184*	*91.25*	*91.26 (0.13)*	*90.31*	*90.28 (0.13)*
2	*8*	*183*	*91.29*	*91.29 (0.10)*	*90.44*	*90.46 (0.11)*
4	*2*	*192*	*91.61*	*91.54 (0.10)*	*90.01*	*89.97 (0.12)*
4	*4*	*186*	*91.57*	*91.51 (0.12)*	*90.33*	*90.22 (0.10)*
4	*6*	*183*	*91.15*	*91.06 (0.16)*	*90.00*	*89.90 (0.15)*
4	*8*	*182*	*91.20*	*91.14 (0.11)*	*90.15*	*90.06 (0.10)*
6	*2*	*192*	*91.71*	*91.57 (0.09)*	*90.06*	*89.92 (0.12)*
6	*4*	*186*	*91.68*	*91.65 (0.15)*	*90.40*	*90.31 (0.13)*
6	*6*	*183*	*91.28*	*91.22 (0.12)*	*90.09*	*90.03 (0.10)*
6	*8*	*182*	*91.33*	*91.27 (0.16)*	*90.23*	*90.17 (0.13)*
8	*2*	*192*	*91.75*	*91.76 (0.07)*	*90.09*	*90.06 (0.14)*
8	*4*	*186*	*91.74*	*91.68 (0.12)*	*90.44*	*90.30 (0.12)*
8	*6*	*183*	*91.35*	*91.29 (0.13)*	*90.13*	*90.08 (0.15)*
8	*8*	*182*	*91.40*	*91.35 (0.13)*	*90.27*	*90.23 (0.13)*

Table 2: **Comparison between computed and simulated service levels when the service measure is volume fill rate**. Parameter values as well as the simulation set-up as in Table *1*.

k_2	$\{S : OFR_1(S) \geq OFR_2(S)\}$	$\{S : VFR_1(S) \geq VFR_2(S)\}$
2	*1 –* 252 *(99.62)*	*1 – 252 (99.67)*
4	*1 – 238 (99.63)*	*1 – 238 (99.68)*
6	*1 – 233 (99.62)*	*1 – 233 (99.68)*
8	*1 – 230 (99.62)*	*1 – 230 (99.68)*
10	*1 – 229 (99.64)*	*1 – 229 (99.69)*

Table 3: Base stock levels where customer class *1* gets the best service. The case strong dominance of X_2 over X_1 and k_1 = *3*. In parenthesis the service level is stated (scaled by *100*) for customer class *2* that corresponds to the upper value of *S* (all lower values of *S* give service level *0* to customer class *2*).

k_2	$\{S : OFR_1(S) \geq OFR_2(S)\}$	$\{S : VFR_1(S) \geq VFR_2(S)\}$
2	*1 – 254 (99.63)*	*1 – 254 (99.62)*
4	*1 – 241 (99.65)*	*2 – 241 (99.70)*
6	*1 – 236 (99.65)*	*8 – 236 (99.70)*
8	*1 – 234 (99.67)*	*15 – 234 (99.71)*
10	*1 – 232 (99.65)*	*22 - 232 (99.71)*

Table 4: Base stock levels where customer class *1* gets the best service. The case strong dominance of X_2 over X_1 and k_1 = *1*. The same comments as for Table *3* apply here.

k_2	$\{S : OFR_1(S) \geq OFR_2(S)\}$	$\{S : VFR_1(S) \geq VFR_2(S)\}$
2	*1 – 303 (99.67)*	*1 – 302 (99.69)*
4	*1 – 291 (99.70)*	*1 – 290 (99.72)*
6	*1 – 287 (99.71)*	*5 – 286 (99.73)*
8	*1 – 285 (99.72)*	*17 – 284 (99.74)*
10	*1 – 283 (99.70)*	*34 – 282 (99.72)*

Table 5: Base stock levels where customer class *1* gets the best service. The case weak dominance of X_2 over X_1 and k_1 = *3*. The same comments as for Table *3* apply here.

k_2	$\{S>0: OFR_1(S) \geq OFR_2(S)\}$	$\{S>0: VFR_1(S) \geq VFR_2(S)\}$
2	*1 (0.00) – 315 (99.71)*	*63 (0.00) – 314 (99.73)*
3	*28 (0.00) – 298 (99.73)*	*Ø*
4	*226 (84.42) -295 (99.74)*	*Ø*
5	*Ø*	*Ø*

Table 6: Base stock levels where customer class *1* gets the best service. The case weak dominance of X_2 over X_1 and $k_1 = 3$. In parenthesis the service level is stated (scaled by *100*) of customer class *2* that corresponds to the values of S . Ø denotes the empty set.

k_1	$\{S>0: OFR_1(S) \geq OFR_2(S)\}$	$\{S>0: VFR_1(S) \geq VFR_2(S)\}$
1	*1 – 10 (0,00)*	*Ø*
2	*1 – 176 (98,03)*	*1 – 52 (0,25)*
3	*1 – 198 (99,74*	*1- 143 (85,78)*
4	*1 – 196 (99,73)*	*1 – 196 (99,75)*

Table 7: Base stock levels where customer class *1* gets the best service. No dominance of X_2 over X_1 and $k_2 = 1$. Ø denotes the empty set. Otherwise see the comments in Table *3*.

Γ_1	Γ_2	k_1	L
2	0.5	3	10
2	0.5	3	5
2	0.5	3	2
0.5	2	3	10
0.5	2	3	5
0.5	2	3	2
1.25	1.25	3	10
1.25	1.25	3	5
1.25	1.25	3	2
2	0.5	1	10
2	0.5	1	5
2	0.5	1	2
0.5	2	1	10
0.5	2	1	5
0.5	2	1	2
1.25	1.25	1	10
1.25	1.25	1	5
1.25	1.25	1	2

Table 8: Parameters to be varied when examining Question 4.

	OFR optimization			*VFR* optimization		
Row	S	$OFR_1(S)$	$OFR_2(S)$	S	$VFR_1(S)$	$VFR_2(S)$
1	*141*	*93.38*	*90.40*	*139*	*93.15*	*90.35*
2	*83*	*94.64*	*90.10*	*81*	*93.78*	*90.05*
3	*46*	*96.04*	*90.14*	*44*	*95.14*	*90.14*
4	*267*	*92.28*	*90.11*	*265*	*91.76*	*90.06*
5	*152*	*92.91*	*90.02*	*151*	*92.59*	*90.37*
6	*78*	*94.06*	*90.02*	*77*	*93.67*	*90.54*
7	*206*	*92.76*	*90.17*	*204*	*92.17*	*90.12*
8	*119*	*93.47*	*90.06*	*117*	*92.75*	*90.01*
9	*63*	*94.77*	*90.15*	*61*	*93.92*	*90.11*
10	*143*	*93.53*	*90.48*	*141*	*92.79*	*90.42*
11	*85*	*94.47*	*90.58*	*83*	*93.62*	*90.53*
12	*47*	*95.68*	*90.39*	*45*	*94.73*	*90.38*
13	*267*	*91.99*	*90.03*	*266*	*91.72*	*90.28*
14	*153*	*92.89*	*90.36*	*151*	*92.24*	*90.30*
15	*79*	*94.07*	*90.54*	*77*	*93.26*	*90.49*
16	*207*	*92.55*	*90.25*	*205*	*91.95*	*90.20*
17	*120*	*93.30*	*90.29*	*118*	*92.57*	*90.23*
18	*64*	*94.67*	*90.61*	*62*	*93.80*	*90.56*

Table 9: Exploration of the difference in service levels between the two customer classes. Row no. in the table corresponds to the row number in Table *8*. X_1 and X_2 are both (delayed) negative binomial distributed with probability parameter $\rho_1 = 0.6$ and $\rho_2 = 0.8$ and shape parameters $s_1 = 1$ and $s_2 = 2$, respectively. $\beta = 0.9$.

	OFR optimization			*VFR* optimization		
s_1	S	$OFR_1(S)$	$OFR_2(S)$	S	$VFR_1(S)$	$VFR_2(S)$
2	*47*	*95.75*	*90.61*	*45*	*95.11*	*90.60*
3	*47*	*95.02*	*90.26*	*45*	*94.67*	*90.25*
4	*48*	*94.75*	*90.76*	*46*	*94.75*	*90.75*
5	*48*	*93.91*	*90.43*	*46*	*94.31*	*90.41*

Table 10: The effect of a changed order size behaviour. The parameter $\Gamma_1 = 2/(0.6s_1 + 0.4)$. The remaining parameters are set as in row no. *3* of Table *9*.

Working Papers from Logistics/SCM Research Group

L-2006-03 Christian Larsen: Computation of order and volume fill rates for a base stock inventory control system with heterogeneous demand to investigate which customer class gets the best service.

L-2006-02 Søren Glud Johansen & Anders Thorstenson: Note: Optimal base-stock policy for the inventory system with periodic review, backorders and sequential lead times.

L-2006-01 Christian Larsen & Anders Thorstenson: A comparison between the order and the volume fill rates for a base-stock inventory control system under a compound renewal demand process.

L-2005-02 Michael M. Sørensen: Polyhedral computations for the simple graph partitioning problem.

L-2005-01 Ole Mortensen: Transportkoncepter og IT-støtte: et undersøgelsesoplæg og nogle foreløbige resultater.

L-2004-05 Lars Relund Nielsen, Daniele Pretolani & Kim Allan Andersen: *K* shortest paths in stochastic time-dependent networks.

L-2004-04 Lars Relund Nielsen, Daniele Pretolani & Kim Allan Andersen: Finding the *K* shortest hyperpaths using reoptimization.

L-2004-03 Søren Glud Johansen & Anders Thorstenson: The (r,q) policy for the lost-sales inventory system when more than one order may be outstanding.

L-2004-02 Erland Hejn Nielsen: Streams of events and performance of queuing systems: The basic anatomy of arrival/departure processes, when focus is set on autocorrelation.

L-2004-01 Jens Lysgaard: Reachability cuts for the vehicle routing problem with time windows.

Handelshøjskolen i Århus

Aarhus School of Business

ISBN 87-7882-119-3

Department of Business Studies

Aarhus School of Business
Fuglesangs Allé 4
DK-8210 Aarhus V - Denmark

Tel. +45 89 48 66 88
Fax +45 86 15 01 88

www.asb.dk