WORKING PAPER L-2008-07

WP

Bisheng Du & Christian Larsen

Base stock policies with degraded service to larger orders

CORAL - Centre for Operations
Research Applications in Logistics

Base stock policies with degraded service to larger orders

Bisheng Du
Christian Larsen

CORAL[1], Department of Business Studies
Aarhus School of Business
Aarhus University
DK-8210 Aarhus V
Denmark

Abstract: We study an inventory system controlled by a base stock policy assuming a compound renewal demand process. We extend the base stock policy by incorporating rules for degrading the service of larger orders. Two specific rules are considered, denoted *Postpone(q,t)* and *Split(q)*, respectively. The parameter q distinguishes between regular orders (of size less than or equal to q) and larger orders. We develop mathematical expressions for the performance measures: order fill rate of the regular orders and average on-hand inventory level. We make numerical experiments where the postpone parameter t and the base stock levels of each rule are such that all customers (of both order types) are indifferent between the two rules. When comparing the difference in the average on-hand inventory levels, we can then make an assessment of the threshold value of the cost of splitting an order (which may otherwise be hard to quantify) in the rule *Split(q)*. Our numerical results indicate that this threshold value is increasing in the variance of the order sizes. Based on the numerical experiment our conclusion is therefore that when the variance of the order sizes is low, then *Postpone(q,t)* seems to be a good option, while when the variance is high, then *Split(q)* is more competitive.
Keywords: Base stock policy, compound renewal process, order fill rate, differentiated service.

1. Introduction

It is well known that the larger demand variation is, the higher inventory levels are needed in order to secure adequate service from an inventory system. One reason for large demand variation may be that the system occasionally receives large customer orders. Furthermore, these large orders may have negative effects further upstream in the supply chain. Therefore, as a manager of an inventory system, you would prefer to receive smaller orders at a more frequent pace rather than receive orders with large variation at a less frequent pace. However, assuming that in the short run you cannot do anything to change the order behaviour of your customers, it may be sensible to introduce a policy implying degraded service to larger orders. Customers submitting larger orders may also very well be aware of the inconvenience that they cause and be willing to accept degraded service. Furthermore, such an initiative may lead the customers to change their ordering behaviour such that in the long run one will in fact encounter less variation in the demand pattern.

The considerations raised here are inspired by a discussion that the authors had recently with the logistics personnel in a larger Danish company where the second author has been involved in an inventory control project. The aim of this project was to decide optimal base stock levels when the service measure is the order fill rate, that is, the fraction of orders received where the whole order is delivered instantaneously from the inventory. This project was reported in Larsen et al. (2008) and some theoretical aspects of the project in Larsen and Thorstenson (2008). Obviously, larger orders

[1] Centre for Operations Research Applications in Logistics, e-mail addresses: bisd@asb.dk and chl@asb.dk

can have very negative effects on the order fill rate service measure (thus, using this service measure, it is implicitly assumed that smaller orders are just as important as larger ones).

As the source of inspiration is this company project, it is natural that the mathematical model developed in this paper takes its point of departure in the model of this project. Therefore, we study a base stock system with control parameter S and where all replenishment orders, issued instantaneously upon receipt of an order, have a constant lead-time L. Our extension of the model is the introduction of a given positive integer q, based on which orders are distinguished as being either regular orders (of a size less than or equal to q) or larger orders (of a size larger than q). We introduce two rules, *Split(q)* and *Postpone(q,t)*, which both seek to serve regular orders as well as possible while degrading the service of larger orders. *Postpone(q,t)* operates as follows: the parameter t is in the interval between *0* and L. When receiving a larger order, say at time point $\overline{\tau}$, the order is deliberately backlogged for t time units before it is attempted to be served. This implies that all regular orders arriving in the time interval $\left(\overline{\tau},\overline{\tau}+t\right)$ are served ahead of this larger order. One could interpret *Postpone(q,t)* not as a rule for degrading service but as a special case of an advance demand information system where all receipts of larger orders are known t time units in advance. For studies of inventory systems in the presence of advance demand information, see among others Hariharan and Zipkin (1995), Gallego and Őzer (2001) and Marklund (2006). As shown in Hariharan and Zipkin (1995), one could therefore make the interpretation that the effective replenishment lead-time of any larger order is L - t. The other rule *Split(q)* operates as follows: each time a larger order is received, it is split into a suborder of size q and a suborder of the remaining size. The first suborder is then treated similarly as a regular order while the second suborder is handled outside the inventory system, directly from the supply system, thus having a lead-time L. Therefore, the inventory system only faces orders that are less than or equal to q and all replenishment orders are the original orders eventually truncated by q.

We believe that both rules could be reasonable choices for handling a situation where one wants to give degraded service to larger orders. As the two rules are quite different, determining which of the two to use is not immediately obvious. However, we here present an approach for accomplishing this. When choosing between the two rules, we propose that one should examine the following *4* elements; *1*: the inconvenience caused to larger orders, *2*: the service offered to regular orders, *3*: the inventory investment and *4*: the additional costs of splitting an order in the *Split(q)* rule. Our approach involves all *4* elements. We believe that for a given q, a threshold value of t exists making the customers of the larger orders indifferent between the two rules. Later in the paper, we provide a reasonable method for establishing this threshold value (though the method is somewhat approximate it is convenient for computations). When this threshold t value is determined, we compute base stock levels enabling the customers of the regular orders to reach the same order fill rate level, irrespective of which rule is used. So when going through these two steps, all customers should be indifferent between to the two rules. The choice of which rule to apply is then based on which of the two performs best with respect to the last two elements. In our opinion the rule *Split(q)* is a bit harder to administrate as it involves splitting and monitoring two suborders. Specifically, we imagine there is a cost incurred when doing a split. However, as this cost of splitting an order may be hard to quantify, we instead identify its threshold value, making the cost performance of the two rules equal (as can be seen later, the threshold value is actually measured relative to the inventory holding cost rate). Therefore, our numerical analysis is focused on identifying values of the input data where this threshold value is low (which suggests that *Postpone(q,t)* is the best choice), and where it is high (which suggests that *Split(q)* is the best choice). As a by-product, our numerical analysis also provides some insights into how the preference between the rules may alter upon observing a change in the behaviour of the customers of the larger orders due to the downgrading of their order requests, for instance, if gradually one begins to see an increased arrival intensity of the order requests but in general a smaller average order size. Here our results seem to suggest that initially one might prefer the *Split(q)* rule. However, due to a possible change in customer behaviour, this could be altered to a preference for the *Postpone(q,t)* rule.

There has been a large number of studies of inventory control in the presence of several (most often two) customer classes. For a good overview of the literature, see Teunter and Haneveld (2008). The aim of these studies is to design rationing rules concerning when to backlog (or reject) the demand of the least important customer class when the inventory level is critically low. In our work, we do not introduce any critical numbers on the inventory levels for when to deny service of (some) orders. Furthermore we do not explicitly model several customer classes. Therefore, our work is not so closely related to the traditional studies of rationing policies in inventory control systems. When considering the rule *Postpone(q,t)*, our work is more in line with the paper of Wang et al. (2002) which is motivated by a company analysis reported in Cohen et al. (1999). They consider a base stock system with two customer classes where the service of one of the classes is first attempted after a given time period (in the paper denoted demand or delivery lead times) has elapsed since the receipt of the order. As the demand model therein is a Poisson process (thus without a compound element), the concern about degrading service of larger orders is obviously outside the scope of this paper. Furthermore, we also generalize the demand model by considering a renewal process instead of a Poisson process. As an aside, we note that recently a paper by Kocuga and Sen (2007) has been published which extends the model of Wang et al. (2002) by introducing critical number rules as seen in the rationing literature, featured in Teunter and Haneveld (2008). When considering the rule *Split(q)*, our work has some relation to studies of order splitting and multiple sourcing. For a review of these studies, see Thomas and Tyworth (2006). However, as we exclusively use *Split(q)* in comparisons with another rule, *Postpone(q,t)*, which cannot be cast into the framework of the splitting/sourcing literature, our work cannot be compared to results obtained in this field.

In Section 2, we derive mathematical expressions for the order fill rate of the regular orders and the average on-hand inventory for the rules *Split(q)* and *Postpone(q,t)*. For each rule, this is done in two stages: first for a general compound renewal process and then in the case of a compound Erlang process, making the mathematical expressions more computable. We finish this section by deriving a reasonable threshold value of t, which we find will make customers of larger orders indifferent between the two rules, and by deriving the threshold value of the split cost. Then in Section 3, we present the results of a numerical study. Finally, we state some concluding remarks in Section 4.

2. Mathematical model

2.0 Preliminaries

The demand process is a compound renewal process where the time between order arrivals is specified by a positive continuous random variable T. The size of a customer order is specified by a positive integer valued random variable X. As stated in the previous section, we assume a given positive integer q that distinguishes between regular and larger orders. We imagine that q is reasonably large, for instance the *90%* or the *95%* quantile of X, as seen later in our numerical experiments.

Let the random variable X^{Reg} denote the size of a regular order. It has probability distribution

$$P(X^{\mathrm{Reg}} = x) = \frac{P(X = x)}{P(X \le q)} \qquad x=1,2,..,q \tag{1}$$

For any non-negative integer m, define the random variable $Q(m)$ as

$$Q(m) = \sum_{i=1}^{m} X_i \tag{2}$$

where X_i, $i=1,..,m$ are independent and identically distributed as X. Similarly, for any non-negative integer m, define the random variable $Q^{Reg}(m)$ as

$$Q^{\mathrm{Reg}}(m) = \sum_{i=1}^{m} X_i^{\mathrm{Reg}} \tag{3}$$

where X_i^{Reg}, $i=1,..,m$ are independent and identically distributed as X^{Reg}. Per definition, $P(Q(0)=0) = 1$ and $P(Q^{Reg}(0)=0) = 1$.

For later use, we state

Lemma 1
When $x \le q$, it holds that $P(Q(m) = x) = \left(P(X \le q)\right)^m P(Q^{\mathrm{Reg}}(m) = x)$
Proof: See Appendix.

Let τ be an arbitrarily chosen time point which in our paper can either be an arrival point of a customer with a regular order or a randomly chosen time point. For any non-negative real number s, the random variable $N(\tau)_s$ is the number of customer arrivals in the time interval $[\tau - s, \tau)$.

2.1 Rule *Split(q)*

We first develop an expression for the order fill rate. Let $\overline{\tau}$ be the time point of an arrival of a customer with a regular order. Let the random variable $\overline{D}_L$ denote the aggregate demand recorded in the inventory system in the time interval $[\overline{\tau} - L, \overline{\tau})$. As all larger orders are truncated by q (and the remaining order of any larger order is handled outside the inventory system), the probability distribution of $\overline{D}_L$ can be specified as follows.

When $x = 0$ or $q = 1$

$$P(\overline{D}_L = x) = P(N(\overline{\tau})_L = x) \tag{4a}$$

and when x > 0 and $q > 1$

$$P(\overline{D}_L = x)$$

$$= \sum_{m=Roof(x/q)}^{x} P(N(\bar{\tau})_L = m) \sum_{y=Roof((qm-x)/(q-1))}^{m} \binom{m}{y} \left(P(X \le q)\right)^y \left(P(X > q)\right)^{m-y} P(Q^{reg}(y) = x-(m-y)q)$$
(4b)

Roof(a) is the smallest integer greater than or equal to the real number *a* and $\binom{m}{y}$ is the binomial coefficient. The order fill rate service measure, abbreviated *OFR*, measuring the fraction of regular orders where the whole order is delivered instantaneously from the inventory, is then

$$OFR_{Split(q)}(S) = P(X^{\mathrm{Reg}} + \bar{D}_L \le S) \tag{5}$$

We now develop expressions for the average on-hand inventory level. Let $\tilde{\tau}$ be a randomly chosen time point. Let the random variable $\tilde{D}_L$ denote the aggregate demand recorded in the inventory system in the time interval $[\tilde{\tau} - L, \tilde{\tau})$. It then follows that the probability distribution of $\tilde{D}_L$ can be specified similarly as (4a-b) when replacing $N(\bar{\tau})_L$ with $N(\tilde{\tau})_L$. The average on-hand inventory level is

$$I_{Split(q)}(S) = \sum_{x=0}^{S-1} P(\tilde{D}_L = x)(S-x) \tag{6}$$

For the case where *T* is a *k*-phased Erlang distribution with mean *k/λ* (that is, *λ* is the intensity of the underlying Poisson process), we have (see for instance Cox (1962) p. 39)

$$P(N(\bar{\tau})_L = n) = e^{-\lambda L} \sum_{j=nk}^{(n+1)k-1} \frac{(\lambda L)^j}{j!} \qquad n=0,1,2,... \tag{7}$$

and

$$P(N(\tilde{\tau})_L = n) = \begin{cases} e^{-\lambda L} \sum_{j=0}^{k-1} \frac{k-j}{k} \frac{(\lambda L)^j}{j!} & n = 0 \\ e^{-\lambda L} \sum_{j=1-k}^{k-1} \frac{k-|j|}{k} \frac{(\lambda L)^{j+nk}}{(j+nk)!} & n = 1,2,... \end{cases} \tag{8}$$

One can thus get computable expressions in order to do numerical analysis of *Split(q)* for the case of a compound Erlang process.

2.2 Rule *Postpone(q,t)*

First, we derive a measure for the order fill rate. Let $\bar{\tau}$ be the arrival point of a customer with a regular order. Let the random variable $\bar{C}_L(t)$ denote what we define to be the committed aggregate demand in the time interval $[\bar{\tau}-L,\bar{\tau})$. This is all the recorded demand in the interval $[\bar{\tau}-L,\bar{\tau}-t)$ and all the recorded demand of the regular orders in the time interval $[\bar{\tau}-t,\bar{\tau})$. The reason why only demand of the regular orders is counted in the latter interval is that all larger orders received in this interval are still denied access to the inventory at time point $\bar{\tau}$. The probability distribution of $\bar{C}_L(t)$ can be specified as follows.

$$P(\bar{C}_L(t)=0)=P(N(\bar{\tau})_L=0)+\sum_{m=1}^{\infty}\left(P(X>q)\right)^m P(N(\bar{\tau})_L=m,N(\bar{\tau})_t=m) \tag{9a}$$

When x > 0

$$P(\bar{C}_L(t)=x)$$
$$=\sum_{m=1}^{\infty}\sum_{r=\max\{m-x,0\}}^{m} P(N(\bar{\tau})_L=m,N(\bar{\tau})_t=r)\sum_{y=0}^{\min\{x-m+r,r\}}\binom{r}{y}\left(P(X\le q)\right)^y\left(P(X>q)\right)^{r-y}P(Q(m-r)+Q^{\mathrm{Reg}}(y)=x) \tag{9b}$$

The two random variables $N(\bar{\tau})_L$ and $N(\bar{\tau})_t$ are positively correlated and $P(N(\bar{\tau})_L\ge N(\bar{\tau})_t)=1$. The order fill rate service measure, measuring the fraction of regular orders where the whole order is delivered instantaneously from the inventory, is then

$$OFR_{Postpone(q,t)}(S)=P(X^{reg}+\bar{C}_L(t)\le S) \tag{10}$$

We now derive a measure for the average on-hand inventory level. Let $\tilde{\tau}$ be a randomly chosen time point. Let the random variable $\tilde{C}_L(t)$ denote the committed aggregate demand in the time interval $[\tilde{\tau}-L,\tilde{\tau})$. As previously defined, this is all the recorded demand in the interval $[\tilde{\tau}-L,\tilde{\tau}-t)$ and all the recorded demand of the regular orders in the time interval $[\tilde{\tau}-t,\tilde{\tau})$. The probability distribution of $\tilde{C}_L(t)$ can be specified as in (9a-b) when the random variables $N(\bar{\tau})_L$ and $N(\bar{\tau})_t$ are replaced by $N(\tilde{\tau})_L$ and $N(\tilde{\tau})_t$, respectively. Therefore, the average on-hand inventory level is specified as

$$I_{Postpone(q,t)}(S)=\sum_{x=0}^{S-1}P(\tilde{C}_L(t)=x)(S-x) \tag{11}$$

A major problem in computing the expressions derived so far for *Postpone(q,t)* is the presence of the joint probabilities $P(N(\bar{\tau})_L=m,N(\bar{\tau})_t=r)$ and $P(N(\tilde{\tau})_L=m,N(\tilde{\tau})_t=r)$. We now consider the case where T is a k-phased Erlang distribution with mean k/λ. Note that a k-phased Erlang distribution can be subdivided into k phases, each having a duration that is exponentially distributed with mean $1/\lambda$. Each time a phase completes, we can interpret it as an "arrival", where it is only every kth arrival that is real while the others are fictitious. Thus, the process will at any time point be in one of the phases *1,2,..,k*. Being in phase *i* means that *k –i* fictitious arrivals have elapsed since the last real arrival, see Figure 1 (note that as we look backward in time when doing mathematical derivations, we also state the phase numbers accordingly).

<Figure 1 about here>

When we use the word arrival in the following, it should be understood as a real arrival. Let τ be an arbitrarily chosen time point. Denote by $F(i,r,t|j)$ the conditional probability that given that we are in phase *j* at time point τ, we are in phase *i* (*i=1,2.,,k*) at time point τ - *t* and the total number of arrivals in the time interval $[\tau\text{-}t,\tau)$ is *r*. Then

$$F(i,r,t|j) = e^{-\lambda t}\frac{(\lambda t)^{rk+i-j}}{(rk+i-j)!} \qquad i = 1,..,k;\ j = 1,..,k;\ r = I_{\{j>i\}},....\infty \tag{12}$$

where the function $I_{\{A\}}$ is *1* if condition *A* is true and *0* otherwise. For an explanation of the number *rk + i – j* in (12), see Figure 2.

<Figure 2 about here>

Let the random variable $\hat{N}(i)_{L-t}$ denote the total number of arrivals in the time interval $[\tau\text{-}L,\tau\text{-}t)$ given that we are in phase *i* at time point τ *–t*. This has probability distribution

$$P(\hat{N}(i)_{L-t} = u) = e^{-\lambda(L-t)}\sum_{v=\max\{uk+1-i,0\}}^{(u+1)k-i}\frac{(\lambda(L-t))^{v}}{v!} \tag{13}$$

Let the random variable $C_L(t|r,u)$ denote the aggregate committed demand (see the previous definition) in the time interval $[\tau\text{-}L,\tau)$ given that one has observed *r* arrivals in the time interval $[\tau\text{-}t,\tau)$ and *u* arrivals in the time interval $[\tau\text{-}L,\tau\text{-}t)$. This has probability distribution

$$P(C_L(t|r,u) = x) = \sum_{y=0}^{\min\{r,x-u\}}\binom{r}{y}\left(P(X \le q)\right)^{y}\left(P(X > q)\right)^{r-y}P(Q^{\mathrm{Reg}}(y) + Q(u) = x) \tag{14}$$

Using τ as an arrival point of a customer with a regular order, we get the probability distribution of $\bar{C}_L(t)$ can be specified as follows

$$P(\bar{C}_L(t) = x) = \sum_{i=1}^{k}\sum_{u=0}^{x}P(\hat{N}(i)_{L-t} = u)\sum_{r=0}^{\infty}F(i,r,t|1)P(C_L(t|r,u) = x) \tag{15}$$

Using τ as a randomly chosen time point, we will at this time point be in phase is *j* with probability *1/k*. Therefore, the probability distribution of $\tilde{C}_L(t)$ can be specified as follows

$$P(\tilde{C}_L(t)=x)=\frac{1}{k}\sum_{j=1}^{k}\sum_{i=1}^{k}\sum_{u=0}^{x}P(\hat{N}(i)_{L-t}=u)\sum_{r=0}^{\infty}F(i,r,t\,|\,j)P(C_L(t\,|\,r,u)=x) \tag{16}$$

This accomplishes that we can get computable expressions for doing numerical analysis of *Postpone(q,t)* in the case of a compound Erlang process.

2.3 Implementation

Some of the expressions presented so far can be further simplified when $k = 1$, that is the demand process is a compound Poisson process, by using the recursion scheme of Adelson (1966). Actually, we have developed computer codes, programmed in Visual Basic for Excel, both for the general Erlangean case (k any positive integer) and the case $k = 1$, in order to validate our computer codes as well as possible. This is done for both rules *Split(q)* and *Postpone(q,t)*.

2.4 The case $S \leq q$

For this case in particular, the rule *Split(q)* is superfluous as not even the truncated suborder of a larger order has any chance of receiving full service. Note that *Postpone(q,0)* represents the case where larger orders are not discriminated. We can prove

Proposition 1

When $S \leq q$, the performance measures (*OFR* and average on-hand inventory) are identical for the rules *Postpone(q,0)* and *Split(q)*.

Proof: see Appendix.

2.5 Choice of t

We derive here what we find is a reasonable threshold value of t, making customers of larger orders indifferent between *Postpone(q,t)* and *Split(q)*. Consider a larger order whose size can be specified by the random variable X^{Lar} which has probability distribution

$$P(X^{Lar}=x)=\frac{P(X=x)}{P(X>q)} \quad x=q+1,q+2,.... \tag{17}$$

In *Postpone(q,t)* a measure for accumulated waiting time of all units of a larger order is $E[X^{Lar}]t$. This measure is slightly optimistic as we here assume that any larger order is instantaneously served after t time units of postponement. Similarly, in *Split(q)* a measure for the accumulated waiting time can be specified as $E[\max\{X^{Lar}-q,0\}]L$. Also this measure is slightly optimistic as we here assume that the truncated suborder (of size q) is served instantaneously. We assume that both measures are equally optimistic. The value of t that equalizes these two expressions is

$$t = \frac{L \sum_{j=q+1}^{\infty} (j-q)P(X=j)}{\sum_{j=q+1}^{\infty} jP(X=j)} \tag{18}$$

The right hand side of (18) belongs to the interval between *0* and *L*. We find that the right hand side of (18) is a good estimate of a *t* value that makes customers of larger orders indifferent between *Postpone(q,t)* and *Split(q)*. Developing exact expressions would be very difficult.

2.6 Specification of a threshold split cost

Irrespective of which rule is applied, we assume that a holding cost rate *h* is charged on the average on-hand inventory. Furthermore, in the cost evaluation of *Split(q)*, we assume that a cost c_{Split} is incurred each time a larger order is split. As the expected time between successive occurrences of larger orders is *E[T]/P(X>q)*, the cost evaluation of *Split(q)* is

$$hI_{Split(q)}(S) + c_{Split}\frac{P(X>q)}{E[T]} \tag{19}$$

while the cost evaluation of *Postpone(q,t)* is $hI_{Postpone(q,t)}(S)$. Assume now that for a given *q*, *t* has been settled according to (18), such that customers of larger orders are indifferent between the two rules. Assume then that for each rule, the base stock levels $S_{Postpone(q,t)}$ and $S_{Split(q)}$, respectively, have been established such that customers of regular orders enjoys the same order fill rate and are therefore also indifferent between the two rules. When equalizing the two cost expressions stated above and isolating the cost parameters, we then get a threshold value of the split cost (relative to the inventory holding cost rate)

$$c_{Split}/h = \frac{(I_{Postpone(q,t)}(S_{Postpone(q,t)}) - I_{Split(q)}(S_{Split(q)}))E[T]}{P(X>q)} \tag{20}$$

In principle, this threshold value could be negative. However, as the demand volume faced by the inventory is less under *Split(q)* than under *Postpone(q,t)* (due to some demand being filtered away from the inventory and sent directly to the supply system) the result is that generally when the parameters are set such that customers are indifferent between the two rules, the average on-hand inventory in *Postpone(q,t)* is larger than the average on-hand inventory in *Split(q)*.

3. Numerical results

As indicated in the previous section, we have only done implementations of the models when *T* is *k*-phased Erlang distributed with mean *k/λ*. Throughout this section, we keep the lead-time *L* fixed at level *4*. We also assume that the random variable *X* is (delayed) geometrically distributed. This means that it has probability distribution

$$P(X=j)=(1-\rho)\rho^{j-1} \qquad j=1,2,\ldots \tag{21}$$

where the parameter ρ is in the interval between *0* and *1*. The use of the word "delayed" is due to Zipkin (2000, p. 451). Then (18) simplifies to

$$t=\frac{1}{q+1-\rho q}L \tag{22}$$

and the demand rate d, defined as $d = E[X]/E[T]$, is

$$d=\frac{\lambda}{k(1-\rho)} \tag{23}$$

When $k = 1$, this demand process is called a "stuttering" Poisson process (see Axsater (2006) p. 82). In Johnston et al. (2003), some empirical evidence is given for the relevance of such a demand model.

We now make a systematic evaluation of how the two rules compare to each other. We let the demand rate d attain the values *1.25, 2.5, 3.75, 5* (these values perhaps seem more obvious choices when multiplying by L as $dL = 5, 10, 15, 20$). For each choice of d, we then let $\rho = 0.5, 0.6, 0.7, 0.8, 0.9$. In order also to investigate the impact of deviating from the common Poisson process assumption, we let $k = 1$ and *2*. The parameter λ is then given from (23). We let q be the α quantile (the least integer value x making $P(X \le x) \ge \alpha$) of X where α is either *0.90* or *0.95*. We denote the required order fill rate of the regular orders β and let it be either *0.9* or *0.95*. Combining all these parameter values (d,ρ,k,α,β) gives a total of *160* data sets. Our numerical procedure can now be outlined as follows.

1. Given input (d,ρ,k,α,β).

2. Compute q as the least integer value x making $P(X \le x) \ge \alpha$ and compute t by (22).

3. Compute $S_{Postpone(q,t)}$ as the least integer value of S bringing $OFR_{Postpone(q,t)}(S) \ge \beta$ and compute $S_{Split(q)}$ as the least integer value of S bringing $OFR_{Split(q)}(S) \ge \beta$. Evaluate the average on-hand inventories $I_{Postpone(q,t)}(S)$ and $I_{Split(q)}(S)$ by setting $S = S_{Postpone(q,t)}$ and $S = S_{Split(q)}$, respectively.

4. Evaluate the threshold split cost relative to the inventory cost by

$$c_{Split}/h=\frac{(I_{Postpone(q,t)}(S_{Postpone(q,t)})-I_{Split(q)}(S_{Split(q)}))k}{\lambda P(X>q)} \tag{24}$$

Here we follow closely the outline given in the introduction. We refrain from making a formal statistical analysis of variance (ANOVA) on the data material on the threshold split costs, as this will probably lead us nowhere as all interaction effects are statistically significant. Instead we inspect the data material more informally. It turns out that c_{Split}/h very much depends on the parameter ρ and it is generally increasing in ρ. A typical illustration of this is given in Table 1, though there are also a few exceptions, as illustrated in Table 2. The reason is that we are forced to choose the base stock values as integers. For instance, the reason why the threshold value c_{Split}/h is about *50* when $\rho = 0.5$ in Table 2 is that the base stock value securing an order fill rate above *0.9* is actually well above, at level *0.93*. As we will demonstrate later, this phenomenon is most likely to occur when the demand rate *d* is low, that is with low-frequent demand. When *d* is increased, the selected values of *S* will in general ensure that the actual order fill rate is close to the required one. Because the pattern seen in Table 1 is the most typical for our data, which can also be verified aggregated from Table 3, we find it natural to group our observations of the c_{Split}/h values on the values of the ρ parameter. In the following Tables 4_{i} – 7_{ii} and Figures 3 – 6, we explore how the other input factors influence the relationship between c_{Split}/h and ρ. From Tables $4_{i\text{-}ii}$ and Figure 3, we see that *k* has some significance. When having an Erlangean arrival process (*k=2*), the threshold value is in general larger, making *Split(q)* more attractive for threshold values falling in between the threshold values of Table 4_{i} and Table 4_{ii}. Thus, the assumption about the demand process has significance for the conclusion. It emphasizes the importance of doing a careful analysis of the data of a demand process in order to determine the appropriate demand model, and not just automatically choose a compound Poisson process because it is most convenient. From Tables $6_{i\text{-}ii}$ and Figure 5, it also seems that the threshold values c_{Split}/h are larger when having a high required order fill rate (*0.95*) on the regular orders than when the required order fill rate is only *0.90*. One explanation may be that as more attention to regular orders is required (by increasing the order fill rate), it is of more advantage to direct some demand away from the inventory system by using the rule *Split(q)*. It does not seem as if the requirement on α is of great significance, see Tables $7_{i\text{-}ii}$ and Figure 6. Finally, when examining Tables $5_{i\text{-}iv}$ and Figure 4, we see, as stated above, that as *d* is increased, it appears more clearly that c_{Split}/h is increasing in ρ. Our results also give some indication of what might happen if customers began to react on our rules for degraded service to larger orders, for instance if customers of larger orders began subdividing their orders into several smaller ones. As an example, examine Tables 1 from bottom ($\rho = 0.9$) to top ($\rho = 0.5$). When following this trace, we initially have a high variation on the order sizes ($\rho = 0.9$), and as the c_{Split}/h is high, it suggests that the *Split(q)* rule might be a good option. As ρ is decreased, the variation decreases and thereby the definition of larger orders is also redefined by decreasing *q*, and *t* is changed as well. In this process it turns out that c_{Split}/h decreases considerably, now making *Postpone(q,t)* more attractive.

<Tables 1 – 7_{ii} about here>
<Figures 3 – 6 about here>

4. Concluding remarks

Using a very general demand model, we have developed mathematical expressions for the average on-hand inventory level and order fill rate service measure of the regular orders in a base stock inventory system when degraded service is enforced for customers of larger orders. The models (and corresponding computer codes) may be valuable tools for an organization to explore the impact of various strategies for degrading the service of larger orders; particularly in recognition of the fact that these orders may have a very disruptive effect upstream in the supply chain. We have refrained from doing a cost optimization with respect to service requirements. This is partly due to the fact that this is a quite complex operation as we have two service level constraints to consider, and partly due to the fact that we believe that a cost parameter such as the cost of splitting a larger order may be hard to quantify. Therefore, our approach is to specify the parameters t, $S_{Postpone(q,t)}$ and $S_{Split(q)}$ such that all customers (irrespective whether they issue a regular or a larger order) will consider the two rules to be equally good. Additionally, we explore the threshold value of the split cost, thereby establishing when each rule seems to perform best. Our numerical investigations reveal that the choice between *Split(q)* and *Postpone(q,t)* seems to depend on the variance of the order sizes. The lower it is, the better option is the rule *Postpone(q,t)*. It also seems that deviating from the common Poisson process assumption as well as the choice of order fill rate offered to the regular orders does have some impact on this conclusion. Though we examine a fairly large number of data sets, we look exclusively at the geometric distribution to describe order sizes. So obviously, further numerical studies could be devoted to examining other distributions, for instance the negative binomial, the Poisson or the logarithmic distribution. As noted in the introduction, the rule *Postpone(q,t)* has resemblance to an advance demand information system. Therefore, the mathematical model concerning this rule can easily be adapted to a more general study about aspects of advance demand information where it need not be solely the customers of larger orders that provide this information.

Acknowledgement

This work is supported by grant no. 275-07-0094 from the Danish Social Science Research Council. We are grateful for comments received from two anonymous reviewers and Johan Marklund, Lund University, Sweden, that considerably improved the paper.

References

Adelson RM (1966). Compound Poisson distributions. *Operational Research Quarterly*, **17**: 73-75.

Axsäter S (2006). *Inventory Control*, *2*nd ed. Springer's International Series. Springer: New York.

Cohen MA, Zheng Y-S, Wang Y (1999). Identifying opportunities for improving Teradyne's service-parts logistics system. *Interfaces*, **29** (4): 1-18.

Cox DR (1962). *Renewal Theory*. Methuen: London.

Gallego G, Őzer Ő (2001). Integrating replenishment decisions with advance demand information. *Management Science*, **47**: 1344-1360.

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

Johnston FR, Boylan JE and Shale EA (2003). An examination of the size of orders from customers, their characterization and the implication for inventory control of slow moving items. *Journal of the Operational Research Society*, **54**: 833-837.

Kocaga YL, Sen A (2007). Spare parts inventory management with demand lead times and rationing. *IIE Transactions*, **39**: 879-898.

Larsen C, Seiding CH, Teller C and Thorstenson A (2008). An inventory control project in a major Danish company using compound renewal demand models. *IMA Journal of Management Mathematics*, **19**: 145-162.

Larsen C, Thorstenson A (2008). A comparison between the order and the volume fill rate for a base-stock inventory control system under a compound renewal demand process. *Journal of the Operational Research Society*, **59**: 798-804.

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

Teunter R, Haneveld WKK (2008). Dynamic inventory rationing strategies for inventory systems with two demand classes, Poisson demand and backordering. *European Journal of Operational Research*, **190**: 156-178.

Thomas DJ, Tyworth JE (2006). Pooling lead-time risk by order splitting: a critical review. *Transportation Research Part E*, **42**: 245-257.

Wang Y, Cohen MA and Zheng Y-S (2002). Differentiating customer service on the basis of delivery lead-times. *IIE Transactions*, **34**: 979-989.

Zipkin PH (2000). *Foundations of inventory management*. McGraw-Hill: Boston.

Appendix

Proof of Lemma 1

We prove by induction in m. When $m = 1$, the result is given from (1). Consider now an integer $m' \geq 1$ and assume that we have shown the result for $m = m' - 1$. For any $x=0,1,\ldots,q$, it holds that

$$P(Q(m')=x)=\sum_{y=0}^{x}P(Q(m'-1)=y)P(X=x-y)=\left(P(X\leq q)\right)^{m'}\sum_{y=0}^{x}P(Q^{\text{Reg}}(m'-1)=y)P(X^{\text{Reg}}=x-y)$$

$$=\left(P(X\leq q)\right)^{m'}P(Q^{\text{Reg}}(m')=x)$$

□

Proof of Proposition 1

We claim that for any $x = 0,1,..,q-1$, it holds that $P(\bar{D}_L = x) = P(\bar{C}_L(0) = x)$. First, we consider the random variable $\bar{C}_L(0)$. When $t = 0$, the random variable $\bar{N}(\bar{\tau})_t$ is identical to zero. This means that the summation on the right hand side of (9a) vanishes to zero. Also, the right hand side of (9b) only gives non-zero terms if index $r = 0$. This again forces index m to be les than or equal to x and index y to be zero. Therefore, we get

$$P(\bar{C}_L(0)=x)=\begin{cases}P(N(\bar{\tau})_L=0) & x=0\\ \sum_{m=1}^{x}P(N(\bar{\tau})_L=m)P(Q(m)=x) & x=1,2,...\end{cases} \tag{A1}$$

Now we consider the random variable $\bar{D}_L$. If $q = 1$, our claim follows by comparing (A1) to (4a). Assume $q > 1$. When $x \leq q - 1$, the inequality $(qm-x)/(q-1) > m-1$ is equivalent to the inequality $q + m > x + 1$, which is true when $q > x$, implying $Roof((qm-x)/(q-1)) = m$ (note $(qm-x)/(q-1)$ is always less than or equal to m). When $q > x$, it also holds that $Roof(x/q) = 1$. Therefore, we get from (4a-b) that

$$P(\bar{D}_L=x)=\begin{cases}P(N(\bar{\tau})_L=0) & x=0\\ \sum_{m=1}^{x}P(N(\bar{\tau})_L=m)\left(P(X\leq q)\right)^{m}P(Q^{\text{Reg}}(m)=x) & x=1,2,...,q-1\end{cases} \tag{A2}$$

Lemma 1 now verifies our claim. Similarly, we can prove that for any $x = 0,1,..,q-1$, it holds that $P(\tilde{D}_L = x) = P(\tilde{C}_L(0) = x)$. As it is the probabilities $P(\bar{D}_L = x) = P(\bar{C}_L(0) = x)$ for $x = 0,1,.., S-1$ that specify the *OFR* expressions in (5) and (10) and it is the probabilities $P(\tilde{D}_L = x) = P(\tilde{C}_L(0) = x)$ for $x = 0,1,.., S-1$ that specify the average on-hand inventory expressions in (6) and (11), the result follows from $S \leq q - 1$.

□

Figures

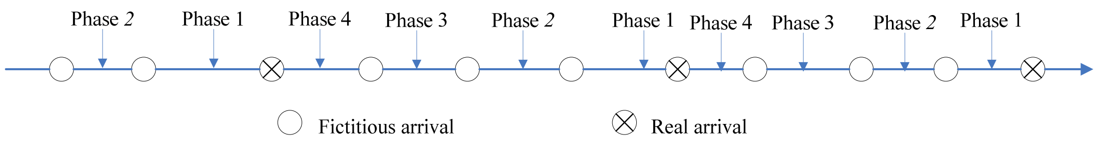

Figure 1: Illustration of an Erlang process with $k=4$.

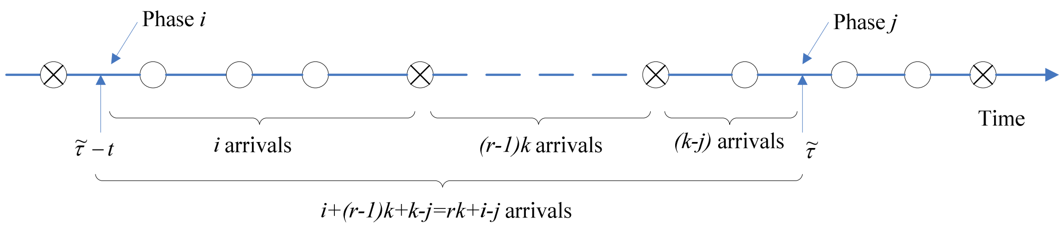

Figure 2: Explanation for (12). Note the word arrival here means real as well as fictitious arrivals.

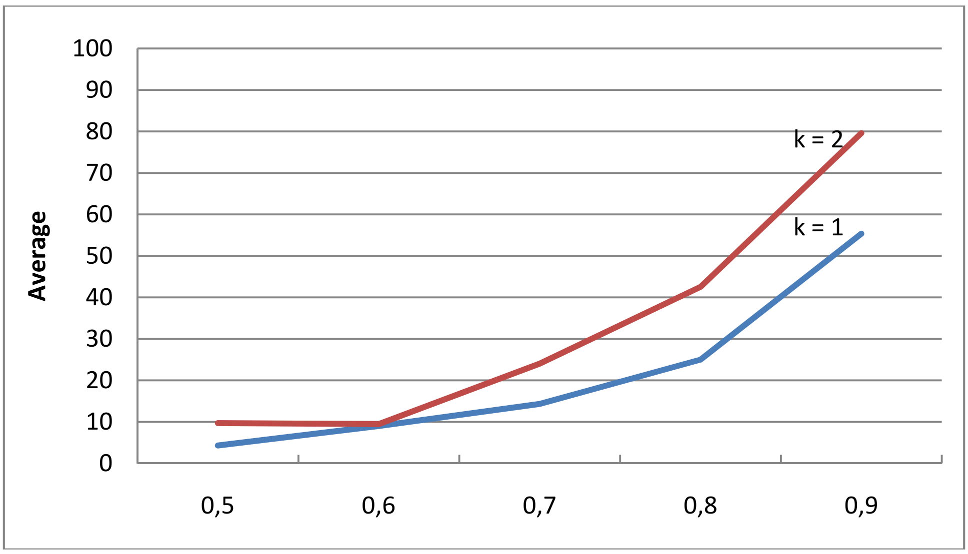

Figure 3: Average threshold split costs from Tables $4_{\text{i-ii}}$ depicted as a function of ρ.

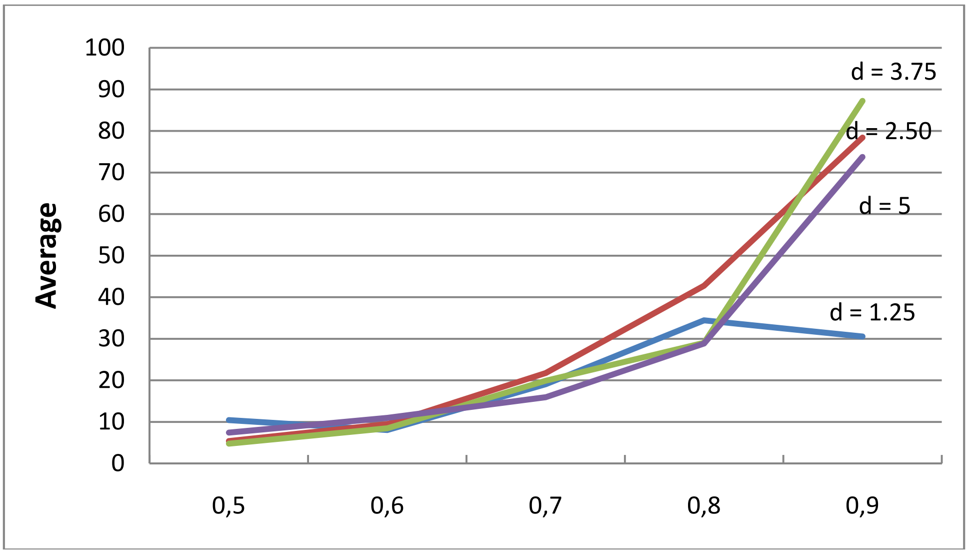

Figure 4: Average threshold split costs from Tables $5_{\text{i-iv}}$ depicted as a function of ρ.

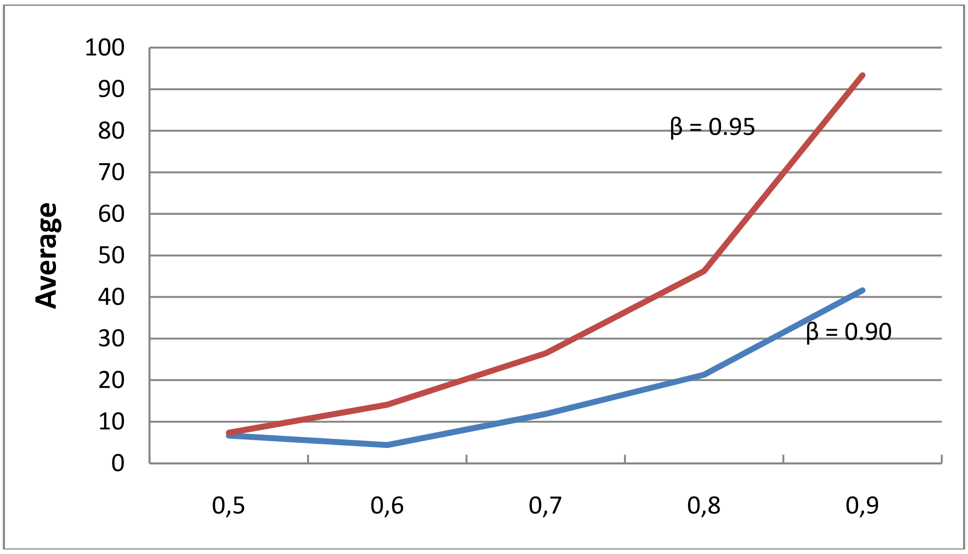

Figure 5: Average threshold split costs from Tables $6_{\text{i-ii}}$ depicted as a function of ρ.

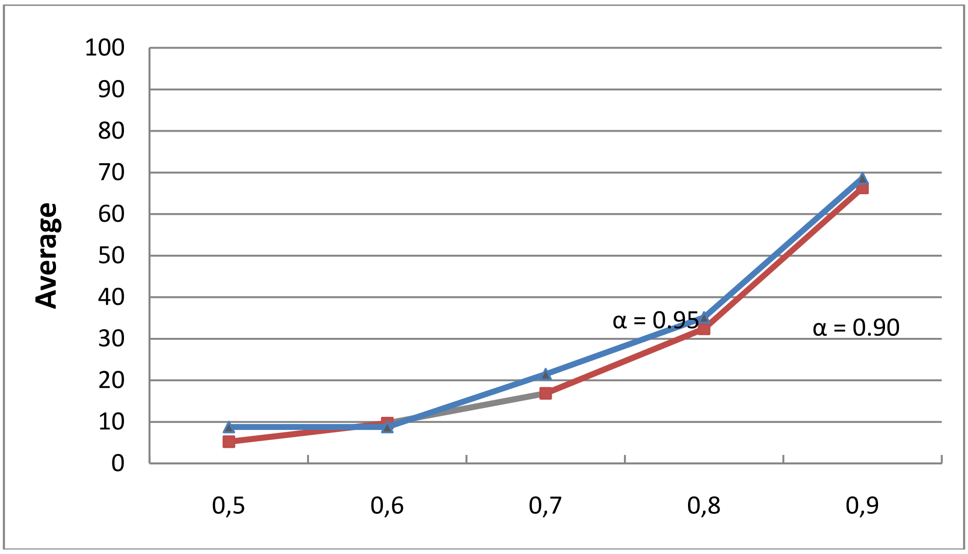

Figure 6: Average threshold split costs from Tables $7_{\text{i-ii}}$ depicted as a function of ρ.

Tables

Input			*Postpone(q,t)*				*Split(q)*			
ρ	λ	q	t	S	I	OFR	S	I	OFR	c_{Split}/h
0.5	0.625	4	1.333333	13	8.385382	0.951239	13	8.354707	0.960539	0.785279
0.6	0.5	5	1.333333	15	10.46213	0.958584	14	9.44159	0.955505	26.24843
0.7	0.375	7	1.290323	17	12.51784	0.95224	16	11.47872	0.950908	33.64719
0.8	0.25	11	1.25	22	17.55818	0.956467	21	16.49419	0.95826	49.54626
0.9	0.125	22	1.25	32	27.72768	0.951631	31	26.59003	0.950894	92.41964

Table 1: Numerical results when $d = 1.25, k=1, \alpha = 0.90, \beta = 0.95$. For simplicity, various subscripts are suppressed.

Input			*Postpone(q,t)*				*Split(q)*			
ρ	λ	q	t	S	I	OFR	S	I	OFR	c_{Split}/h
0.5	1.25	5	1.142857	11	6.250177	0.932796	10	5.264566	0.907758	50.46328
0.6	1	6	1.176471	11	6.40433	0.902063	11	6.343367	0.913418	2.613286
0.7	0.75	9	1.081081	13	8.401099	0.906259	13	8.332111	0.914233	4.558888
0.8	0.5	14	1.052632	16	11.48464	0.906994	16	11.37934	0.910106	9.576759
0.9	0.25	29	1.025641	24	19.64693	0.902384	24	19.51507	0.90118	22.39568

Table 2: Numerical results when $d = 1.25, k=2, \alpha = 0.95, \beta = 0.90$. For simplicity, various subscripts are suppressed.

ρ	Average	St.deviation
0.5	6.993599	10.78944
0.6	9.233533	7.785681
0.7	19.17503	14.76924
0.8	33.76842	23.39682
0.9	67.48941	43.94978

Table 3: All data on threshold split costs grouped on their ρ value. In all *32* observations in each group.

ρ	Average	St.deviation
0.5	4.296584	7.062121
0.6	9.007948	7.643195
0.7	14.31746	11.49414
0.8	24.99271	17.41206
0.9	55.36688	35.18928

Table 4_i: All data on threshold split costs with $k = 1$ grouped on their ρ value. In all *16* observations in each group.

ρ	Average	St.deviation
0.5	*9.690614*	*13.23607*
0.6	*9.459119*	*8.169931*
0.7	*24.03261*	*16.38137*
0.8	*42.54413*	*25.76513*
0.9	*79.61193*	*49.39785*

Table 4_{ii}: All data on threshold split costs with $k = 2$ grouped on their ρ value. In all 16 observations in each group.

ρ	Average	St.deviation
0.5	*10.42467*	*18.3465*
0.6	*8.032891*	*11.30724*
0.7	*19.07961*	*23.93596*
0.8	*34.43772*	*40.52619*
0.9	*30.54452*	*25.24033*

Table 5_{i}: All data on threshold split costs with $d = 1.25$ grouped on their ρ value. In all *8* observations in each group.

ρ	Average	St.deviation
0.5	*5.367868*	*9.022767*
0.6	*9.466754*	*9.227178*
0.7	*21.77455*	*13.88*
0.8	*42.78129*	*15.67816*
0.9	*78.40797*	*35.95885*

Table 5_{ii}: All data on threshold split costs with $d = 2.50$ grouped on their ρ value. In all *8* observations in each group.

ρ	Average	St.deviation
0.5	*4.766321*	*6.171197*
0.6	*8.452221*	*5.101997*
0.7	*19.90063*	*12.41371*
0.8	*29.00739*	*16.949*
0.9	*87.24183*	*54.59416*

Table 5_{iii}: All data on threshold split costs with $d = 3.75$ grouped on their ρ value. In all *8* observations in each group.

ρ	Average	St.deviation
0.5	7.415543	6.091241
0.6	10.98227	4.84963
0.7	15.94535	5.105444
0.8	28.84729	10.09324
0.9	73.76331	38.16384

Table 5_{iv}: All data on threshold split costs with $d = 5$ grouped on their ρ value. In all *8* observations in each group.

ρ	Average	St.deviation
0.5	6.637132	13.3852
0.6	4.385126	3.837437
0.7	11.87145	9.553983
0.8	21.31465	16.05706
0.9	41.61732	23.68623

Table 6_{i}: All data on threshold split costs with $\beta = 0.90$ grouped on their ρ value. In all 16 observations in each group.

ρ	Average	St.deviation
0.5	7.350067	7.819857
0.6	14.08194	7.771774
0.7	26.47862	15.67569
0.8	46.2222	23.2941
0.9	93.3615	44.75401

Table 6_{ii}: All data on split threshold costs with $\beta = 0.95$ grouped on their ρ value. In all *16* observations in each group.

ρ	Average	St.deviation
0.5	5.227948	7.14656
0.6	9.691002	7.663387
0.7	16.85802	10.71873
0.8	32.38406	23.19559
0.9	66.29964	39.85611

Table 7_{i}: All data on threshold split costs with $\alpha = 0.90$ grouped on their ρ value. In all *16* observations in each group.

ρ	Average	St.deviation
0.5	8.759251	13.5226
0.6	8.776064	8.130241
0.7	21.49205	18.01274
0.8	35.15279	24.2733
0.9	68.67917	48.99397

Table 7_{ii}: All data on threshold split costs with $\alpha = 0.95$ grouped on their ρ value. In all 16 observations in each group.

[Working Papers from CORAL – Centre for Operations Research Applications in Logistics]

L-2008-07 Bisheng Du & Christian Larsen: Base stock policies with degraded service to larger orders.

L-2008-06 Karina H. Kjeldsen: Classification of routing and scheduling problems in liner shipping.

L-2008-05 Daniele Pretolani, Lars Relund Nielsen, Kim Allan Andersen & Matthias Ehrgott: Time-adaptive versus history-adaptive strategies for multicriterion routing in stochastic time-dependent networks.

L-2008-04 Christian Larsen: Comments to Özkaya et al. (2006).

L-2008-03 Christian Larsen: Derivation of confidence intervals of service measures in a base-stock inventory control system with low-frequent demand.

L-2008-02 Jens Lysgaard: The Pyramidal Capacitated Vehicle Routing Problem.

L-2008-01 Jens Lysgaard & Janni Løber: Scheduling participants of Assessment Centres.

L-2007-03 Christian Larsen: Note: Comments on the paper by Rosling (2002).

L-2007-02 Christian Larsen, Claus Hoe Seiding, Christian Teller & Anders Thorstenson: An inventory control project in a major Danish company using compound renewal demand models.

L-2007-01 Christian Larsen: The *Q(s,S)* control policy for the joint replenishment problem extended to the case of correlation among item-demands.

L-2006-11 Daniele Pretolani, Lars Relund Nielsen & Kim Allan Andersen: A note on "Multicriteria adaptive paths in stochastic, time-varying networks".

L-2006-10 Lars Relund Nielsen, Kim Allan Andersen & Daniele Pretolani: Bicriterion a priori route choice in stochastic time-dependent networks.

L-2006-09 Christian Larsen & Gudrun P. Kiesmüller: Developing a closed-form cost expression for an *(R,s,nQ)* policy where the demand process is compound generalized Erlang.

L-2006-08 Eduardo Uchoa, Ricardo Fukasawa, Jens Lysgaard, Artur Pessoa, Marcus Poggi de Aragão, Diogo Andrade: Robust Branch-Cut-and-Price for the Capacitated Minimum Spanning Tree Problem over a Large Extended Formulation.

L-2006-07 Geir Brønmo, Bjørn Nygreen & Jens Lysgaard: Column generation approaches to ship scheduling with flexible cargo sizes.

L-2006-06 Adam N. Letchford, Jens Lysgaard & Richard W. Eglese: A Branch-and-Cut Algorithm for the Capacitated Open Vehicle Routing Problem.

L-2006-05 Ole Mortensen & Olga W. Lemoine: Business integration between manufacturing and transport-logistics firms.

L-2006-04 Christian H. Christiansen & Jens Lysgaard: A column generation approach to the capacitated vehicle routing problem with stochastic demands.

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.

ISBN 9788778823588

Department of Business Studies

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

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

www.asb.dk