CEJOR (2009) 17:95–110
DOI 10.1007/s10100-008-0077-8

ORIGINAL PAPER

A queuing approach for inventory planning with batch ordering in multi-echelon supply chains

Sandeep Jain · N. R. Srinivasa Raghavan

Published online: 27 November 2008

Abstract This paper presents stylized models for conducting performance analysis of the manufacturing supply chain network (SCN) in a stochastic setting for batch ordering. We use queueing models to capture the behavior of SCN. The analysis is clubbed with an inventory optimization model, which can be used for designing inventory policies . In the first case, we model one manufacturer with one warehouse, which supplies to various retailers. We determine the optimal inventory level at the warehouse that minimizes total expected cost of carrying inventory, back order cost associated with serving orders in the backlog queue, and ordering cost. In the second model we impose service level constraint in terms of fill rate (probability an order is filled from stock at warehouse), assuming that customers do not balk from the system. We present several numerical examples to illustrate the model and to illustrate its various features. In the third case, we extend the model to a three-echelon inventory model which explicitly considers the logistics process.

Keywords Order batching · Queuing systems · Inventory optimization · Multi-echelon

List of symbols

Q number of units in one bucket
K total number of buckets at warehouse
Z maximum inventory at warehouse ($K\,Q$)

S. Jain (✉) · N. R. S. Raghavan
Department of Management Studies, Indian Institute of Science, Bangalore 560012, India
e-mail: j.sandeep@yahoo.com

N. R. S. Raghavan
e-mail: dr.nrsraghavan@gmail.com

λ	demand arrival rate at warehouse
$A(t)$	number of orders arrived up to time t at manufacturer
μ	service rate of manufacturing plant units/unit time
$I(t)$	inventory at warehouse at time t
$N(t)$	number of orders at manufacturing plant being processed at time t
$B(t)$	number of back orders in the system at time t
$R(t)$	number of orders arrived up to time t, but after the last batch was released for processing
h	inventory holding cost ($ per unit per unit time)
b	back order cost ($ per unit per unit time)
C_s	order set up cost ($ per set up)
α	desired service level for orders at warehouse i.e. probability an order is filled from stock at warehouse
ρ	intensity of the system ($\lambda/\mu < 1$)
M	number of retailers served by the warehouse
λ_m	demand arrival are at retailer m units/time unit
λ	$\sum_{m=1}^{M} \lambda_m$, net demand rate at the warehouse units/time unit
l_m	lead time for logistics for retailer m to receive items from warehouse, $l_1 = l_2 = \cdots = l_M = l$ (Local haul assumed negligible), as exponential random variable
X_m	demand during replenishment lead time, a Poisson random variable
$p_m(a)$	probability mass function of demand, $P\{X_m = a\}$
$F_{Xm}(x)$	cumulative distribution function of demand during lead time
Γ	expected number of orders in the queue $M^C/M/\infty$ in steady state
ξ	service rate of logistics process (exponentially distributed)
ρ'	intensity of the the logistics hub ($\lambda/\xi < 1$)
W	expected waiting time at warehouse due to back ordering alone
L_m	mean lead time (including back ordering delay) for an order of items from retailer m to be filled from warehouse, $L_1 = L_2 = \cdots = L_M = L$
θ_m	expected demand during replenishment lead time for item at retailer ($\theta_m = \lambda_m L_m$)
β_m	probability of satisfying customer demand at retailer m from available stock
r_m	reorder point for item at retailer m (decision variable)
R_m	$r_m + 1$, base stock level for item at retailer m

1 Introduction

The supply chain activities constitute a mega process and numerous decisions are involved in their successful design and operation. Decisions regarding stocking and control of inventory of physical goods are a problem common to all enterprises. Asset managers of large enterprises have the responsibility of determining the approximate inventory level in the form of components and finished goods to hold at each level of supply chain in order to guarantee specified end customer service levels. Given the size and complexity of the supply chain, a common problem for these asset managers

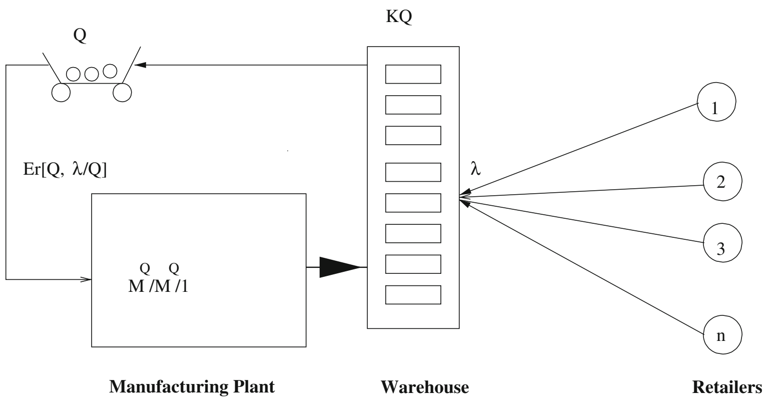

Fig. 1 Three-stage supply chain network

is to know how to quantify the trade-off between service level and investment in inventory required to support these service levels. The problem is made even more difficult because the supply chains are highly dynamic with uncertainty in demand, variability in processing times at each stage of the supply chain, multiple dimensions for customer satisfaction, finite resources, etc.

We model a three-stage linear supply chain. We assume the supply chain consists of one manufacturer with one warehouse who supplies for n identical retailers as shown in Fig. 1. We assume all retailers have identical requirements. Here, identical requirement means different retailers require similar items both in configuration and in quantities. We do not explicitly model the outbound logistics process from the warehouse (we present this in a later section). This can albeit be modeled as part of the manufacturing process itself, at the cost of being inexact. We model the warehouse inventory as the input control mechanism for the manufacturer which itself is modeled as a single stage queueing system.

The orders for retailers arrive at the warehouse as a Poisson process with a constant rate λ. Here we assume that each order is placed for a constant batch Ω of items across all retailers. The Poisson assumption is appropriate if the orders are placed from a population consisting of a large group of individuals acting independently. It also makes the analysis tractable. There are several papers on inventory management with Markovian arrivals assumed (see for instance Lin et al. 2000; Kim and Tang 1997). The warehouse keeps finished goods inventory in buckets, each of which holds exactly Q number of batches of size Ω. Q is an integer, which is multiple of Ω. There are a total of K buckets at the warehouse. Thus, the warehouse can keep a maximum of KQ units in inventory. Arriving orders from different retailers deplete the on-hand inventory at warehouse, if any. Otherwise, (in a stock-out situation) the arriving orders have to wait to be fulfilled; the waiting process consists of manufacturing the desired units to be produced at the manufacturing plant and being shipped to the warehouse. We refer this situation as back order. In our model, we assume that back orders can be infinite. The warehouse places orders to the manufacturing plant when one bucket (Q units) is depleted from the inventory at the warehouse. Thus, when the retailers orders are Poisson, the inter-arrival time of the orders at the manufacturing plant is Erlang distributed with Q phases and rate λ/Q. We assume that the manufacturing plant has infinite waiting line capacity. At the manufacturer plant, orders arrive in batches of Q units. In the first model, the processing time of orders includes setup, manufacturing and logistics time. The orders are processed based on the first-come-first-serve policy. The manufactured items are shipped in batch of Q units. We assume processing time for each lot forming Q units, is exponentially distributed with rate μ. Markovian assumption can be justified if setup time is highly indeterministic (see for instance Mandelbaum et al. 1988). In this paper, in the first model we also assume that the setup cost is negligible but we relax this assumption in the next model.

Our work represents the application of queueing models to supply chain analysis and design. Given the values K and Q, we can compute the distribution of the number of orders/inventory in the system or the waiting times. Such computations are useful in determining the available to promise (ATP) quantities and lead times at the supplier's facilities. They are also useful in deriving delivery reliability value for a given supply chain. Since this is a simple application of certain existing results, we do not present the same in this paper. We instead dwell on the converse problem where we design the supply chain by optimally choosing K and Q. A total cost function is formulated based on inventory carrying costs and back order costs subjected to the constraint of meeting set bounds on fill rate (probability an order is filled from stock at warehouse).

This paper is organized as follows. In the next section, we review related work followed by a presentation of our model in Sect. 3. In Sect. 4 we present main performance measures of interest and their characteristics. We present an alternate formulation in Sect. 5 where we impose service level constraint. The design problem and numerical examples are illustrated in Sects. 6 and 7, where several observations on optimal inventory are discussed. We include one more stage of inventory and logistics process in Sect. 8 followed by conclusions in Sect. 9.

2 Literature review

In this section, we briefly survey the literature on mathematical models for inventory control in supply chains. There is a large body of literature available on make to stock inventory models. For an overview of various inventory models, the reader may refer Hadley and Whitin (1963), Pyke et al. (1998).

Inventory control under a periodic review policy with demand arriving from stochastic and deterministic sources has been discussed by Sobel (2001). The authors use dynamic programming to formulate an optimization problem and prove that modified (s, S) is the best policy under general conditions if there is setup cost.

For a single product assembly system, Song and Yao (2002) present a model where final product is assembled to order and the sub-assemblies are built-to-stock. Greedy type algorithms are developed to solve the inventory/service trade-off.

Matheus and Gelders (2000) consider an inventory system subjected to probabilistic non-unit sized demand pattern, and propose an exact and an approximate reorder point calculation method for the (R, Q) inventory policy. Simulation results are presented for various distributions under different service levels.

Karmakar (1987) examines the effect of lot sizing on work-in-process inventory and lead time. The queueing models are used to capture the effect of intensity and scale. The model was extended to multi-products also.

Kim and Tang (1997) focus on the inventory control in pull production systems with single warehouse, and single production facility. Production authorization (PA) system in the form of PA card is used to control the inventory in the system. The authors highlight the trade-off between manufacturing lead time and response time and computed optimal inventory at the warehouse by using heuristics. A single stage $E_K/M/1$ (continuous time) queueing model is used for analyzing the system.

Performance analysis of make to stock supply chains presented by Sandeep and Raghavan (2003) use discrete time queueing models for the first time.

Our work in this paper considers the effect of batching at the manufacturer and the retailer in a queueing setting. We consider the single class case with no bill of materials involved. Our models are hence stylized yet yield rich insight into stochastic inventory control, and may be viewed as an extension to the work by Kim and Tang (1997) albeit, using a different approach for analysis and optimization.

3 Analysis

At the warehouse, there is some cost associated with keeping inventory and there will be some back order cost associated with orders in backlog. We assume that both inventory carrying cost and back order cost are linear in nature. We would like to minimize the expected total cost at the warehouse, i.e.

Minimize Total Cost=Expected inventory holding cost+Expected back ordering cost

Mathematically, we can express,

$$\text{Minimize TC}(K, Q) = h\ E[I] + b\ E[B] + C_{\text{s}}\frac{\lambda}{Q} \tag{1}$$

subject to,

$$K, Q \in \mathbf{Z}^{+} \tag{2}$$

We consider the case in which orders arrive at the manufacturing plant in batches of Q units. The manufacturing plant itself processes these orders in the batches of Q. We assume that there is only one processing center at the manufacturing plant. Hence any new batch of orders has to wait in the waiting buffer until the manufacturing plant finishes the processing of the entire current batch. The dynamics reflected above can be modeled using an $E_Q/E_Q/1$ queueing system. This is a bulk arrival, batch processing system. For computing inventory and back orders, we need to develop stochastic equations, which capture the properties of the system as in Buzacott and Shanthikumar (1993). Observe that,

$$R(t) = A(t) - \left\lfloor \frac{A(t)}{Q} \right\rfloor Q, \qquad t \geq 0 \tag{3}$$

$$B(t) = \{N(t)Q + R(t) - KQ\}^{+}, \qquad t \geq 0 \tag{4}$$

which means

$$B(t) = \max\,[N(t)\ Q + R(t) - KQ, 0], \qquad t \geq 0$$

Similarly for the inventory, the equation is,

$$I(t) = \{KQ - N(t)\ Q - R(t)\}^{+}, \qquad t \geq 0 \tag{5}$$

which means

$$I(t) = \max\,[KQ - N(t)\ Q - R(t), 0], \qquad t \geq 0$$

The corresponding steady state probability distribution for R, N, B, I are as follows: R is uniformly distributed from 0 to $Q - 1$. Thus,

$$P\{R = n\} = 1/Q, \qquad n = 0, 1, \ldots, Q - 1 \tag{6}$$

The distribution of N can be approximated as:

$$P\{N = n\} = p_N(n) \approx \begin{cases} 1 - \rho & n = 0 \\ \rho(1 - \sigma)\sigma^{n-1} & n = 1, 2, \ldots \end{cases} \tag{7}$$

$$P\{B = n\} = \frac{1}{Q}\ p_N\left(\left\lfloor \frac{Z + n}{Q} \right\rfloor\right), \qquad n = 1, 2, \ldots \tag{8}$$

$$P\{I = n\} = \frac{1}{Q}\ p_N\left(\left\lfloor \frac{Z - n}{Q} \right\rfloor\right), \qquad n = 1, 2, \ldots, KQ \tag{9}$$

Where σ is expressed as:

$$\sigma = \frac{\hat{N} - \rho}{\hat{N}} \tag{10}$$

where $\hat{N}$ is the expected number of orders in the system. Further $\hat{N}$ can be simplified as,

$$\hat{N} = \lambda w_0 + \rho \tag{11}$$

where w_0 is the waiting time of jobs in a $GI/G/1$ queue. We use the approximation provided in Buzacott and Shanthikumar (1993) wherein:

$$w_0 = \hat{W}_{GI/GI/1}\left[\frac{\lambda}{Q}, \mu Q, \frac{1}{Q}, \frac{1}{Q}\right] \tag{12}$$

The above functional form can be further simplified as,

$$w_0 = \left\{\frac{\lambda(1 + Q)}{(\mu^2 Q^5 + \lambda^2)}\right\}\left\{\frac{\mu^2 Q^4 + \lambda^2}{2\mu Q^2(\mu Q^2 - \lambda)}\right\} \tag{13}$$

Finally, we get

$$\hat{N} = \frac{\lambda}{\mu}\left\{\frac{\lambda(1+Q)}{\mu^2Q^5+\lambda^2}\right\}\left\{\frac{\mu^2Q^4+\lambda^2}{2\mu Q^2(\mu Q^2-\lambda)}\right\}+\rho \tag{14}$$

where $\rho = \lambda/\mu$ and $\rho < 1$ for stability.

We can simplify $E[I]$ and $E[B]$ as,

$$E[I] = \left[\left(\sum_{i=1}^{K-1}\sum_{j=Q(i-1)}^{Qi-1} i\ p_N\{K-(j-1)\}\right) + KQ\ p_N\{0\}\right] \tag{15}$$

$$E[B] = \left[\left(\sum_{i=1}^{\infty}\sum_{j=Q(i-1)}^{Qi-1} i\ p_N\{K-(j-1)\}\right)\right] \tag{16}$$

Hence, we can express our objective function as:

$$\begin{aligned}\mathrm{TC}_{\{K,Q\}} = &\left[\left(\sum_{i=1}^{K-1}\sum_{j=Q(i-1)}^{Qi-1} i\ p_N\{K-(j-1)\}\right) + KQ\ p_N\{0\}\right] b\\ &+\left[\left(\sum_{i=1}^{\infty}\sum_{j=Q(i-1)}^{Qi-1} i\ p_N\{K-(j-1)\}\right)\right] h + C_s\lambda/Q\end{aligned} \tag{17}$$

4 Characteristics of performance measures

It can be easily seen from Eq. (17) that TC is a complicated function of Q and K. Hence it is not easy to solve the optimization problem directly. This motivates us to examine some characteristics of TC, which enable us to proceed further for optimization. For any given size of the bucket Q, TC has the following properties:

Effect of K:

1. For any given size of bucket Q, TC is convex in K.

Proof The proof can be established by showing the necessary condition $\frac{\partial(TC)}{\partial K} = 0$ and by sufficient condition $\frac{\partial^2(TC)}{\partial K^2} \geq 0$. It is shown in Figs. 2 and 3. See Appendix A for details.

Property 1 can be interpreted as follows: first, as the number of buckets K increases the number of orders in the system increases. This causes the increase in inventory in the system, which leads to the total cost, TC, to increase. Second, as the number of buckets K increases, the stock-out probability decreases.

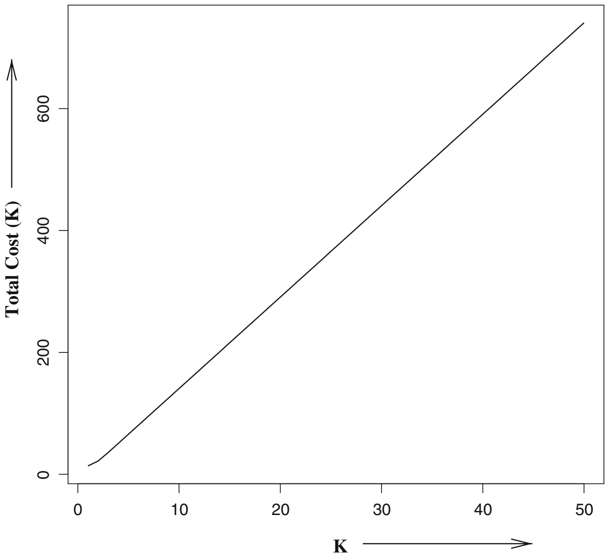

Fig. 2 Total cost as function of the number of buckets for the case when $\rho = 0.50, \alpha = 0.95, b/h = 1$

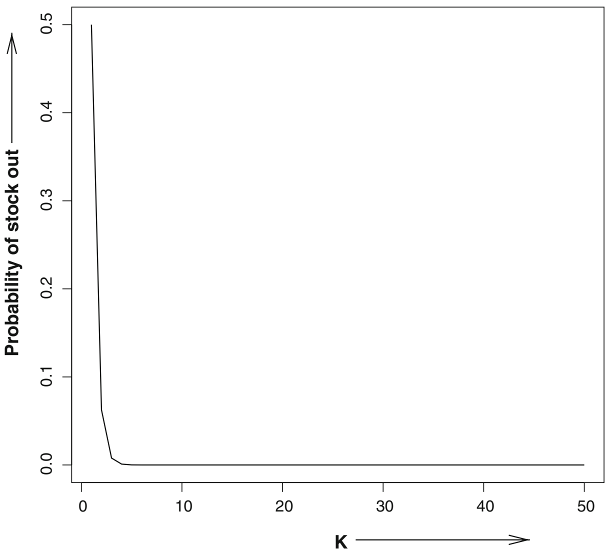

Fig. 3 Probability of stock out as function of buckets for the case when $\rho = 0.50, \alpha = 0.95, b/h = 1$

5 Alternate model

In the current section, we present an alternative formulation of the problem. In our first model, we assumed that the back order cost is given. When back order cost is not explicitly known then the penalty on back orders can be imposed in terms of a stock out constraint.

Now we express the objective function as,

$$\text{Minimize } TC(K, Q) = h\ E[I] + C_s \frac{\lambda}{Q} \tag{18}$$

subject to ,

$$P[I = 0] \leq 1 - \alpha \tag{19}$$

$$K, Q \in \mathbf{Z}^+ \tag{20}$$

For our formulations, it is obvious that when there is inventory at the warehouse, the demand will be satisfied immediately and if there is no inventory, the orders will be back ordered. We define the stock out probability as:

$$P\{I = 0\} = P\{Z \leq N(t)Q + R(t)\}, \quad t \geq 0 \tag{21}$$

$$= \frac{1}{Q}\rho(1-\sigma)\sigma^{K-1} + \rho\sigma^K \tag{22}$$

6 The design problem

In this section, we discuss the design part of our problem. We consider the optimization problem with objective function equation 1, which is subjected to constraint equation 2. In our model, for obtaining the values of the objective function, we input values of λ, μ, Q, b, and h and get the function for total cost. In real life, μ is known to the designer based on the history and this is a system control parameter but λ has to be estimated and is an uncontrolled parameter. Likewise, for the constraints, we input the value of Q, and obtain the probability of an order being filled from stock at warehouse. We generate objective functions for different b and h values. Of course, λ and μ are constant for one set of observations. We compute the optimum value of K by enumeration. The output of our model is the optimum K (K^*), which helps us to compute optimum inventory level and total cost.

7 Discussion

In this section, we present interpretation of the results from our experiments. It is evident from the results that as we increase Q, for the same b/h, optimum number of buckets (K^*) is non-increasing. This happens in both unconstrained and constrained models. See Figs. 4, 5, 6, 7, 8, 9, 10, and 11.

As b/h increases (for same ρ and Q), total inventory and total cost at the warehouse increases. It implies that if back orders are expensive than holding the inventory then the system tends to hold more inventory. From Fig. 4, for $Q = 1$, $b/h = 1$ gives lower optimum cost than $Q = 1$, and $b/h = 10$.

As Q increases, for the same traffic intensity (work load) and same b/h, total inventory at the warehouse increases. Hence, it is always better to keep Q as low as possible if setup cost is negligible (see Figs. 4, 5, 6, 7, 8, 9, 10, and 11).

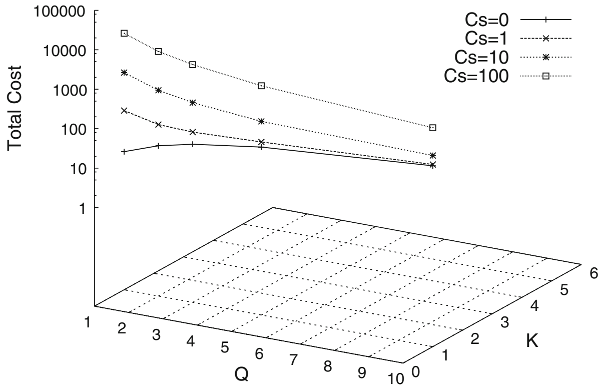

Fig. 4 Optimum inventory level versus K and Q when $\lambda = 100$, $\rho = 0.5$, $h = \$10$, $b/h = 1$

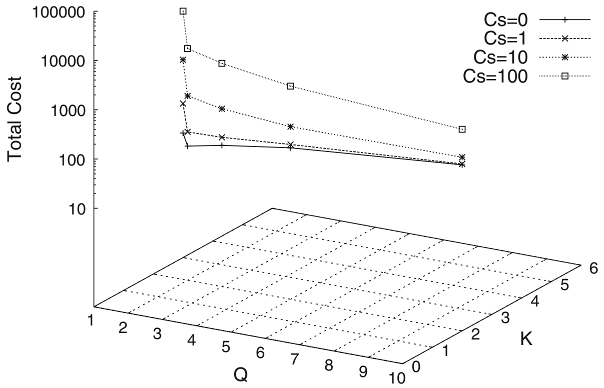

Fig. 5 Optimum inventory level versus K and Q when $\lambda = 100$, $\rho = 0.5$, $h = \$10$, $b/h = 10$

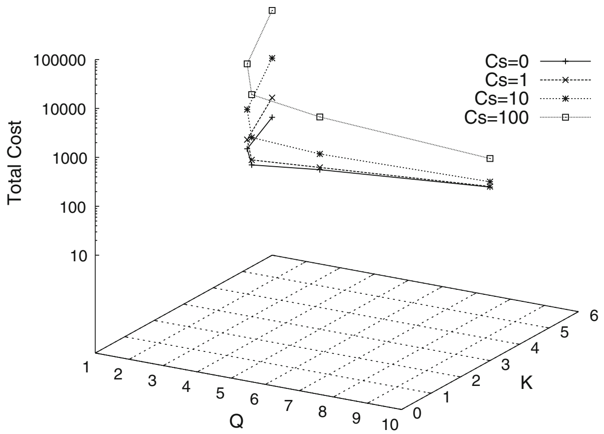

Fig. 6 Optimum inventory level versus K and Q when $\lambda = 100$, $\rho = 0.5$, $h = \$10$, $b/h = 100$

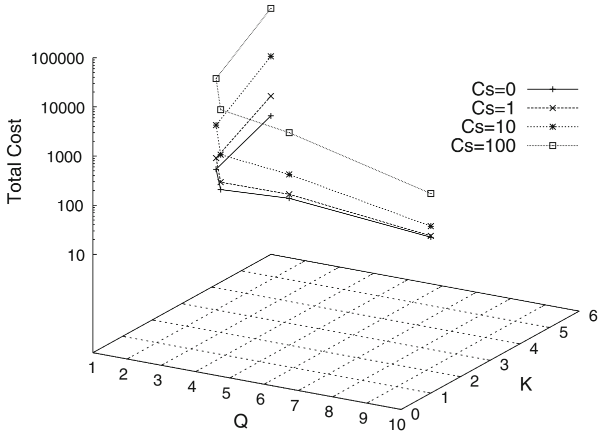

Fig. 7 Optimum inventory level versus K and Q when $\lambda = 100$, $\rho = 0.9$, $h = \$10$, $b/h = 1$

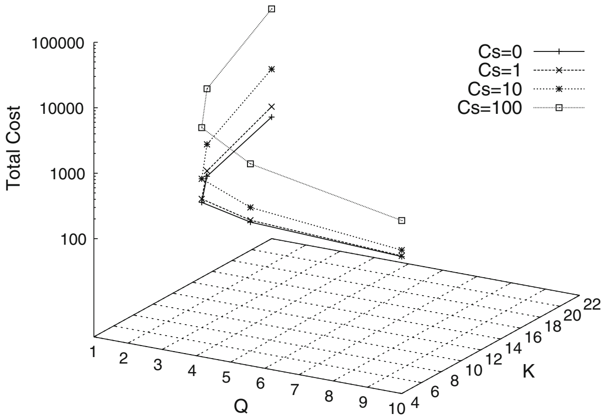

Fig. 8 Optimum inventory level versus K and Q when $\lambda = 100$, $\rho = 0.9$, $h = \$10$, $b/h = 10$

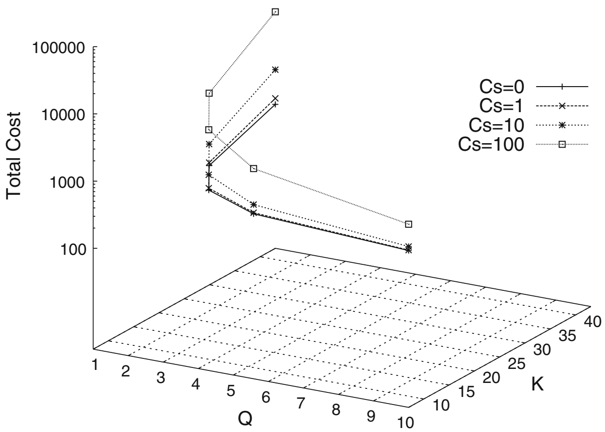

Fig. 9 Optimum inventory level versus K and Q when $\lambda = 100$, $\rho = 0.9$, $h = \$10$, $b/h = 100$

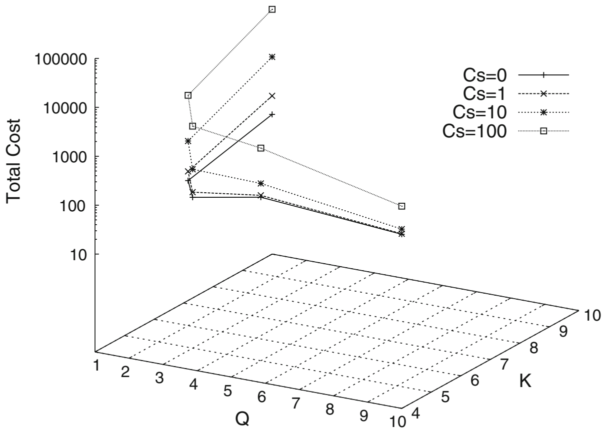

Fig. 10 Optimum inventory level versus K and Q when $\lambda = 100$, $\rho = 0.75$, $h = \$10$, $\alpha = .95$

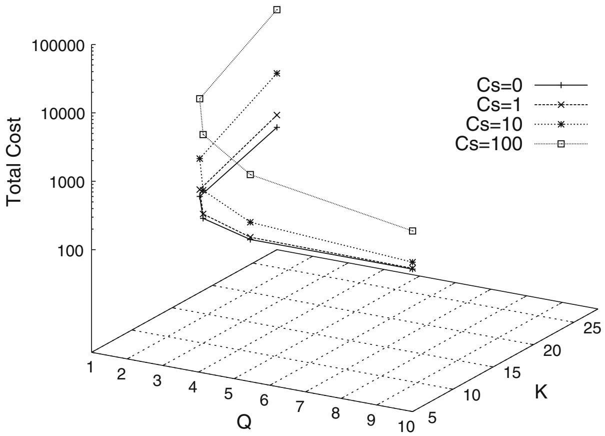

Fig. 11 Optimum inventory level versus K and Q when $\lambda = 100$, $\rho = 0.95$, $h = \$10$, $\alpha = .95$

From Figs. 4, 5, 6, 7, 8, and 9, it is evident that as the intensity of the system increases, the total inventory required at the warehouse also increases for the same b/h and Q. This means that at high utilization, the total cost is also high for the same customer fill rate.

We see that for $\rho = 0.50$ and $b/h = 10$, $Q = 3$ is the best policy for ordering with set up costs, $C_s = \$10$ per set up but when set up cost is $C_s = \$100$ per set up then $Q = 10$ is a superior policy. The results are explicitly shown in Figs. 4, 5, and 6. Thus the order batching is desirable in the presence of increased set up costs. When setup cost(C_s) is zero then $Q = 1$ is the best policy. It validates our model that base stock policy is best policy when setup cost is negligible.

In the model 2 for attaining the same service level (α), required inventory increases as we increase ρ. See Figs. 10 and 11.

8 Three stage supply chain with base stock policy at retailer end

In this section, we extend our model for the three-echelon case (see Fig. 12). We assume that both warehouse and retailers make use of continuous review policies, where retailers follow the base stock policy. At the retailer end, retailers follow one-at-a-time replenishment policy. It means, if demand is a Poisson process, then demand at the warehouse is also Poisson. Here, we assume that there is some logistics time to supply items from warehouse to retailers. We model the logistics process by using $M/M^C/\infty$ queue in continuous time, where C is vehicle capacity which is deterministic and service rate (logistics time) is exponentially distributed. We also assume that to supply the order, infinite vehicles are available. Thus assumption holds good for third party logistics.

We assume that the logistics process is independent of the manufacturing plant and is dependent only on the customer orders for its arrival process. This subsumes that as soon as orders arrive at the logistics hub, they are serviced. Nevertheless, customer orders wait for inventory only at the manufacturing plant. observe that customer orders are Poisson with rate λ and hence with the assumption that service at hub is also exponential, $M/M^C/\infty$ is a good model to start with.

For the performance analysis of $M/M^C/\infty$ queue, we follow the results of Kashyap et al. (1990) and Purdue and Linton (1981). All arrived units are processed individually, by an infinite pool of servers so that all jobs are processed and processing times (logistic times) are Markovian Lin et al. (2000). In our model, we assume that all retailers are co-located. Hence, the time required to serve retailers can be assumed to be *iid* because the time spent in logistics process is only in line haul. Finished goods transfer from warehouse to co-located retailers is called as *line haul*. *Local haul* is distribution of finished goods among retailers and time in local haul is assumed negligible. Now for the given service level (fill rate) at retailers end we wish to compute the base stock level at retailers end and the optimal inventory at the warehouse. The latter is computed as in the earlier sections and we do not present that here. We need the following notation for further analysis.

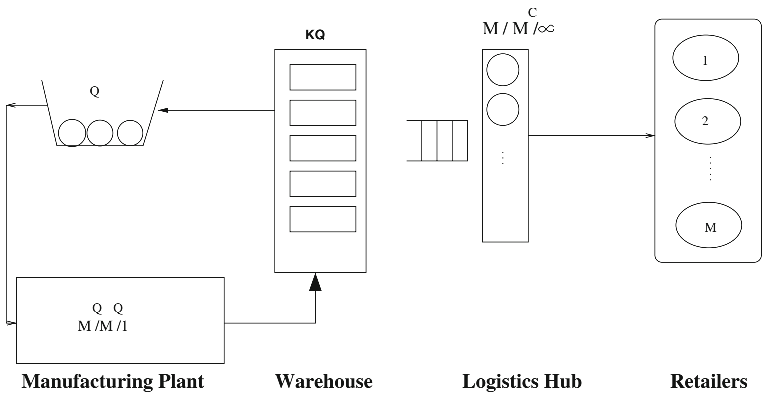

Fig. 12 Three-stage supply chain network with logistics hub and co-located retailers

We can compute the expected number of outstanding back orders at warehouse from Eq. (16). The expected time an order from retailer waits at the warehouse due to back ordering can we expressed using Little's law as

$$W = E[B]/\lambda \tag{23}$$

Hence, the total mean lead time at retailer end is

$$L_m = l_m + W \tag{24}$$

We obtain l_m (which is same for all the retail outlets i.e. $l_1 = l_2 = \cdots = l_m = l$) from the $M/M^C/\infty$ queue analysis Kashyap et al. (1990). At steady state mean number of items in the system can be given as

$$\Gamma = \frac{\lambda}{\xi} C \tag{25}$$

and by applying Little's law, we can get logistics lead time as

$$l = C/\xi \tag{26}$$

Hence L_m is at hand. From the mean lead time we can compute expected demand during replenishment lead time as

$$\theta_m = \lambda_m L_m \tag{27}$$

and cumulative distribution during demand replenishment lead time at retailer m can be expressed as

$$F_{Xm}(x) = \sum_{a=0}^{x} P\{X = a\} \tag{28}$$

For the given r_m values, we can find $F_{Xm}(r_m)$. We vary value of r_m until it crosses the desired service level β_m. The corresponding r_m value is the optimum value of reorder point and base stock value is $R_m = r_m + 1$.

Now, we present a numerical example to compute base stock level of retailers. Let there be two retailers, then $M = 2$ and let $\lambda_1 = 1.0$ unit/day; $\lambda_2 = 2.0$ unit/day; $\lambda = \lambda_1 + \lambda_2 = 3.0$ unit/day, $C = 5$ units, $\xi = 5$ unit/day, $\rho' = 0.60$, $\rho = 0.75$, $\beta_1 = 0.95$, $\beta_2 = 0.98$, $l = 1.0$ day

- For $Q = 1$, $W = 0.04$ days, $L = 1.04$, $\theta_1 = 1.04$, $\theta_2 = 2.08$, $r_1 = 3$, $R_1 = 4$ and $r_2 = 5$, $R_2 = 6$
- For $Q = 2$, $W = 0.10$ days, $L = 1.10$, $\theta_1 = 1.10$, $\theta_2 = 2.20$, $r_1 = 3$, $R_1 = 4$ and $r_2 = 6$, $R_2 = 6$
- For $Q = 5$, $W = 0.21$ days, $L = 1.21$, $\theta_1 = 1.21$, $\theta_2 = 2.42$, $r_1 = 3$, $R_1 = 4$ and $r_2 = 6$, $R_2 = 7$

- For $Q = 10$, $W = 0.67$ days, $L = 1.67$, $\theta_1 = 1.67$, $\theta_2 = 3.34$, $r_1 = 4$, $R_1 = 5$ and $r_2 = 8$, $R_2 = 9$
- For $Q = 15$, $W = 0.67$ days, $L = 1.67$, $\theta_1 = 1.67$, $\theta_2 = 3.34$, $r_1 = 4$, $R_1 = 5$ and $r_2 = 8$, $R_2 = 9$
- For $Q = 20$, $W = 0.68$ days, $L = 1.68$, $\theta_1 = 1.68$, $\theta_2 = 3.36$, $r_1 = 4$, $R_1 = 5$ and $r_2 = 8$, $R_2 = 9$

We again observe that batching at the warehouse does not increase the base stock levels of the retailers greatly.

9 Conclusions

This paper focuses on developing analytical models for inventory optimization in supply chains. We consider a three-stage supply chain where inventory is maintained at the warehouse only. Numerical results are presented for different workloads of the system. These results can be used to design the optimal inventory policy of the system. We assume that the arrival of orders from the retailers to the warehouse is Poisson. This is a limitation of the model, although it can model several real world settings. In addition, the service process at the manufacturer is assumed to be exponentially distributed. This captures essentially a high variance process, which may not be the practical case. A more generic model will need estimating the actual distributions and using approximations for formulating optimization problems. Our models nevertheless, can be used to illustrate the impact of variance in the procurement process, as also designing inventory control systems at the warehouse.

This research also enables to understand the effect of order batching in supply chains. Moreover, we believe that the model used can be extended to other problems like available-to-promise, due date setting etc. Thus, this research has a significant impact on overall inventory management. We summarized our results below:

- As we increase the bulk size of orders, for achieving the same service level, required inventory at the warehouse is higher. This increases the total inventory carrying cost for the same back order to inventory cost ratio. Higher bulk size of the order increases variability in the process. Therefore, it is always advisable to place orders in small batches if setup cost is negligible.
- In a low workload system, inventory required to satisfy the same service level, is less as compared to a high workload system.
- As back order to inventory cost ratio increases, the optimum inventory level remains the same but the total cost increases.

The current work can be extended for multi products models, supply chains with more than two echelons, cases where Inbound and outbound logistics are included in the model, other inventory policies like (Q, r), etc.

Acknowledgments We are thankful to anonymous reviewers and associate editor whose comments helped us to posit the better work.

Appendix A: Convexity of the objective function

To get the global minima of the objective function i.e. Eq. (18), under constraints (19) and (20), should satisfy $\frac{\partial^2(\mathrm{TC})}{\partial K^2} \geq 0$. We present below two cases because for a general case expression is very complex.

For $Q = 5,\ b = 1,\ h = 1,\ \rho = 0.8$

$$\frac{\partial^2(\mathrm{TC})}{\partial K^2} = 12.5469\ (0.411028)^K \tag{A-1}$$

For $Q = 10,\ b = 10,\ h = 5,\ \rho = 0.6$

$$\frac{\partial^2(\mathrm{TC})}{\partial K^2} = 277.6141\ (0.3113)^K \tag{A-2}$$

This can be verified for any values of $Q,\ b,\ h,\ \rho$.

References

Buzacott JA, Shanthikumar GJ (1993) Stochastic models of manufacturing systems. Prentice Hall, New Jersey

Hadley G, Whitin TM (1963) Analysis of inventory systems. Prentice-Hall Inc., Englewood Cliffs

Jain S, Raghavan NRS (2003) Performance analysis of make to stock supply chains using discrete time queueing models. Oper Res Proc 2002 1:65–70

Karmakar US (1987) Lot sizes, lead times and in-process inventories. Manage Sci 33(3):409–418

Kashyap BRK, Liu L, Templeton JGC (1990) On the $GI^X/G/\infty$ system. J Appl Probab 27:671–683

Kim I, Tang CS (1997) Lead time and response time in a pull production control system. Eur J Oper Res 101(3):474–485

Lin GY, Ettl M, Feigin GE, Yao DD (2000) A supply network model with base-stock control and service requirements. Oper Res 48(2):216–232

Mandelbaum A, Ackere AV, Chen H, Harrison JM, Wein LM (1988) Empirical evaluation of a queueing network model for semiconductor wafer fabrication. Oper Res 26(2):202–215

Matheus P, Gelders L (2000) The (R,Q) inventory policy subject to compound poisson demand pattern. Int J Prod Econ 68(3):307–317

Purdue P, Linton D (1981) An infinite-server queue subject to an extraneous phase process and related models. J Appl Probab 18:236–244

Pyke DF, Silver AE, Peterson R (1998) Inventory management and production planning and scheduling. Wiley, New York

Sobel MJ (2001) Inventory policy for systems with stochastic and deterministic demand. Oper Res 49(1):157–162

Song JS, Yao DD (2002) Performance analysis and optimization of assemble -to-order systems with random lead times. Oper Res 50(5):889–903