A Comparison Study of Stochastic- and Guaranteed-Service Approaches on Safety Stock Optimization for Multi Serial Systems

Peng Li∗, Di Wu
Faculty of Economics and Management, Xian University of Technology, Xian 710048, China

*Corresponding author e-mail: lipenggn@xaut.edu.cn

Abstract. Two competing approaches have been developed over the years for multi-echelon inventory system optimization, stochastic-service approach (SSA) and guaranteed-service approach (GSA). Although they solve the same inventory policy optimization problem in their core, they make different assumptions with regard to the role of safety stock. This paper provides a detailed comparison of the two approaches by considering operating flexibility costs in the optimization of (R, Q) policies for a continuous review serial inventory system. The results indicate the GSA model is more efficient in solving the complicated inventory problem in terms of the computation time, and the cost difference of the two approaches is quite small.

1. Introduction
Effective management of inventories in a supply chain is critical for the firms in the chain to assure a high service level to their customers at the minimal costs. As the supply chain can be modeled as a multi-echelon inventory system, one important issue of this management is to find an optimal inventory policy for such system. Two approaches are often used in the optimization of inventory policies for multi-echelon inventory systems: stochastic-service approach (SSA) (Clark and Scarf, 1960, 2000) and guaranteed-service approach (GSA) (Graves and Willems, 1996). The two approaches differ in their demand treatment and service time characterization. In the SSA, each stock maintains a certain level of safety stock in order to cope with the variability of its demand. When the on-hand inventory of the stock is not sufficient to meet its demand, unsatisfied part of the demand will be backlogged. This implies that the stock may have a stochastic delay to fulfill the unsatisfied demand. Contrarily, the GSA assumes that each stock can use operating flexibility measures such as expending and overtime to fulfill excessive customer demand superior to a pre-specified bound as a supplement to its safety stock. With this assumption, the approach can guarantee that each stock assures a pre-specified service time to its customer. Although the two approaches have been used in inventory theory for some time, only Klosterhalfen and Minner (2010) compared the costs of the two base-stock policies obtained by the GSA and the SSA, respectively, and observed that the GSA outperforms the SSA for systems with long warehouse processing time and high retailer service level.

This paper provides a detailed comparison of the two safety stock optimization approaches by considering operating flexibility costs in the optimization of (R, Q) policies for a continuous review serial inventory system with Poisson final demand and fixed order costs at each stock (Li et al, 2013). The consideration of operating flexibility costs makes the optimization problem much more complex.
but more practical. Firstly, we describe the two inventory optimization approaches for serial systems and outline two properties for the order size $Q$ in the SSA; Secondly, we propose two methods for solving the models based on a line search for finding the optimal reorder point $R$ and order size $Q$ when the target cycle service level is given. Numerical experiments on randomly generated instances show the comparison results of the two approaches.

2. Mathematical model formulations

2.1. Common Assumption and Characteristics

A continuous review serial inventory system with $N$ ($N>2$) stocks is considered, where stock $N$ orders from an external supplier with unlimited stock, stock $N-1$ orders from stock $N$, stock $N-2$ orders from stock $N-1$, and so on. Finally, at the lowest stock, stock 1, customer demand occurs. A serial inventory system with $N$ stocks can be depicted as in Figure 1. A practical example of this type of inventory system can be found in the mechanical industry, for instance, where a metal material passes through several processing operations such as cutting, drilling, grinding before it becomes a final product.

![Figure 1. A serial inventory system with $N$ stocks](image)

The customer demand is assumed to be stationary and independent Poisson distribution with the average demand rate $\lambda$. All demands not being immediately satisfied from stock are backordered. All stocks in the system operate a continuous review echelon ($R$, $Q$) policy. That is, when its inventory position (the stock on hand plus its outstanding orders minus all backorders at stock 1) of a stock declines to or below a reorder point $R$, an order of $Q$ units is placed. In addition, each stock is assumed to have a deterministic production time, $T_i$, which is the production lead time, given that all necessary components are available. No capacity constraints exist at any of the stocks.

The objective is to minimize the expected total costs of the system per time unit, while at the same time achieving a predefined non-stockout probability service level at the end stock. For the system considered, the total costs consist of three types: inventory holding cost, fixed order costs and operating flexibility costs for fulfilling excessive demand. Specifically, the inventory holding costs will be evaluated based on the echelon on-hand inventory of each stock; the fixed order costs are evaluated based on the number of orders that each stock places to its supplier (immediate upstream stock). That is, the placement of each order incurs a fixed order cost. As for the third cost, it is assumed to depend linearly on the amount of demand fulfilled by using operating flexibility. All parameters to be used in the formulation of the total cost are given as follows:

- $c_i$: fixed cost for placing each order by stock $i$ to its supplier, i.e., stock $i+1$,
- $h_{i}\epsilon$: echelon holding cost per unit of product per time unit for stock $i$, $i=1,2,...,N$,
- $p$: cost for using operating flexibility to fulfill each unit of excessive customer demand,
- $d[t, t+L)$: the sum of demands that occur during the time interval $[t, t+L)$,
- $R_i$: reorder point of each stock $i$, $i=1,2,...,N$,
- $Q_i$: order size of each stock $i$, $i=1,2,...,N$,
- $\beta$: the fill rate of the system, which is the fraction of customer demand satisfied directly from the stock.
2.2. Stochastic-service Approach

SSA model specifies the demand as a stochastic process is more realistic compared with their deterministic counterparts. In the case of stockout, each stock fully backorders the unsatisfied demand and fills the demand later when its on-hand inventory becomes available. Thus, the delay of demand fulfilled due to stockout is random and the committed service time of each stock cannot be 100% guaranteed. The service time of each stock \( i \), \( ST_i \), defined as the time it takes until materials ordered by its downstream stock are made available for processing by the stock, is thus stochastic. In case of immediate material availability, the service time is 0. Consequently, the entire replenishment time of each stock \( i \), \( Li = ST_i + Ti \), is stochastic as it consist of the stochastic service time of the stock and the deterministic processing time.

For the serial system considered, Zheng (1992) provided stochastic inventory models and got some key results, which will be used in this paper. In their model, they combine the inventory holding costs and backorder penalty costs into one, and referred as inventory costs. Denote \( G_i(y) \) as the expected inventory costs of stock \( i \) at time \( t+Li \) when its inventory position at time \( t \) equals \( y \). Therefore, the expected total system costs can be formulated as the following function:

\[
C(Q, R) = \sum_{i=1}^{N} c_i \lambda \beta + \int_{R_i}^{R_i+Q} \frac{G_i(y)dy}{Q_i}
\]

Where \( G_i(y) = E[h_i(y-d)^+ + p(d-y)^+] \).

Equation (1) is valid when the inventory position is uniformly distributed over the interval \([R_i+1, R_i+Q]\) in steady state and is independent of the lead time demand. Furthermore, when the customer demands are discrete and face stationary distribution process, the cost function can be approximate as the following equation:

\[
\tilde{C}(Q, R) = \sum_{i=1}^{N} c_i \lambda \beta + \sum_{y=R_i}^{R_i+Q} \frac{G_i(y)}{Q_i}
\]

Moreover, for echelon \((R, Q)\) inventory policy considered, we restrict it satisfying the integer-ratio constraint, that is, the order size of each stock \( i \) is a multiple of the order size of its immediate successor, i.e., stock \( i-1 \). Specifically, \( Q_i \) is a multiple of \( Q_{i+1} \), i.e., \( Q_i = m_{i+1} Q_{i+1} \), where \( m_{i+1} \) is a positive integer for \( i=1,2,\ldots,N \). This assumption is natural since stock \( i-1 \) always places an order of \( Q_{i+1} \) units to stock \( i \) and each inventory replenishment order of stock \( i \) is used to fulfill the demands from stock \( i-1 \).

The problem is to find the order quantity \( Q^* \) and reorder point \( R^* \) for each stock that minimize the total system costs. The two-dimensional minimization problem \( C(Q, R) \) can be solved sequentially (Gallego, 1997). Let \( R_i(Q_i) \) denote the optimal \( R_i \) for fixed \( Q_i \) for stock \( i \). Then, \( R_i^* = R_i(Q_i) \) if and only if \( G_i(R_i) = G_i(R_i+Q_i) \). In order to fixed \( Q_i \), we present two properties to determine the possible values of \( Q_i \), \( i=1,2,\ldots,N \).

Property 1: For serial inventory system with \( N \) stocks, an upper bound of \( Q_1 \) is given by:

\[
Q_1 = \sqrt{\frac{2\lambda \beta (c_1 + c_2 + \cdots + c_N)}{h_1^p + 3h_2^p + \cdots + (2N-1)h_N^p}} = \sqrt{\frac{2\lambda \beta \sum_{i=1}^{N} c_i}{\sum_{i=1}^{N} (2i-1)h_i^p}}
\]
Property 2: For stock $i$ ($i=2, \ldots, N$) of the serial inventory system, if the state of stock $i-1$ is $Q_{i-1}$, then an upper bound of $m_{i-1}$ can be derived by:

$$m_{i-1} = \frac{1}{Q_{i-1}} \cdot \sqrt{\frac{2\lambda \beta \sum_{j=1}^{N} e_j}{\sum_{j=1}^{N+1} (2j-1) \cdot h_{j-1+j}^e}}$$  \hspace{1cm} (4)$$

With the above properties, we can derive all possible values of $Q_i$ for each stock $i$, $i=1,2,\ldots,N$. Then, fixed each possible $Q_i$, get the optimal $R_i(Q)$ for each stock $i$ (Gallego, 1998).

2.3. Guaranteed-service Approach

One key assumption of the GSA is that a maximum reasonable lead time demand level is specified for lead time demand of the customer and excessive part of the lead time demand beyond the levels is fulfilled by using operating flexibility. Here, the maximum level is not defined directly on the demand of each time unit but the lead time demand. Since the lead time is a decision variable in the GSA model, this level is usually defined as a function of the lead time.

Let us denote the lead time demand over $\tau$ units of time ($\tau \geq 0$) from time $t-\tau$ to time $t$ ($t \geq \tau$) by $d[t-\tau, t]$ and the maximum reasonable lead time demand level over the lead time demand by $D(\tau)$. The bounded lead time demand assumption can be described as follows:

$$D(\tau) \geq d[t-\tau, t]$$  \hspace{1cm} (5)$$

Note that the bounded demand assumption and the GSA were adopted by Graves and Willems (1996, 2000) in the context of setting the safety stock in a supply chain. In their work, it is assumed that if the lead time demand exceeds the upper bound, the stock might resort to extraordinary measures such as expediting and overtime to fulfill the excessive part of the demand. Since the safety stock is strongly related to the service level of the stock, the lead time demand can be given according to the event-oriented service level $\alpha$ to final customer. For the serial system depicted in Figure 1, the customer demand only occurs at stock 1, that is, the maximum reasonable lead time demand level at stock 1 is defined as the minimum number $D(\tau)$ satisfying the following condition

$$p\{d[t-\tau, t] \leq D(\tau)\} \geq \alpha$$  \hspace{1cm} (6)$$

Since the customer demand at stock 1 follows a Poisson process with average demand rate $\lambda$, $D(\tau)$ can be calculated by

$$\sum_{k=0}^{D(\tau)} \frac{[\lambda \tau]^k}{k!} e^{-\lambda \tau} \geq \alpha$$  \hspace{1cm} (7)$$

In the GSA, it is assumed that each stock $i$ has an outbound service time $S_i$, a production time $T_i$, and an inbound service time $S_{i+1}$, $i=1,2,\ldots,N$. The outbound service time $S_i$ is the time required by stock $i$ to satisfy the demand of its immediate downstream stock $i-1$. This means that when stock $i$ receives a demand at time $t$ from stock $i-1$, it has to fulfill the demand by time $t+S_i$. The inbound service time $S_{i+1}$ is the time required by stock $i$ to receive the corresponding delivery from its immediate upstream stock $i+1$ after it places an order. According to GSA, stock $i$ has to keep a certain amount of on-hand inventory to satisfy the demand occurring between $t+S_i$ and $t+S_{i+1}+T_i$, the time span $S_{i+1}+T_i$ is called the net lead time of stock $i$. The net lead time plays an important role in the GSA.
In our previous work, the safety stock optimization problem can be formulated as the following problem:

\[
P: \text{Minimize } \sum_{i=1}^{N} \left( \frac{c_i \lambda}{Q_i} + \sum_{j=1}^{N} \left[ \sum_{j=1}^{N} D(S_j + T_j - S_{j-1}) - \lambda \beta(S_i + T_i - S_{i-1}) + \frac{1+Q_i}{2} - i \right] + \sum_{j=i}^{N} h_{ij}^{e} \cdot Q_{i-1} \right) + p \lambda (1 - \beta)
\]

Subject to:

\[
Q_{i+1} = m_i Q_i \text{ for } i = 1, 2, \ldots, N - 1
\]

\[
SI_i + T_i - S_i \geq 0 \text{ for } i = 1, 2, \ldots, N
\]

\[
SI_i \geq S_{i-1} \text{ for } i = 1, 2, \ldots, N
\]

\[
0 \leq S_i \leq s_i
\]

\[
Q_i \geq 0 \text{ and integer for } i = 1, 2, \ldots, N
\]

\[
SI_i, S_i \geq 0 \text{ and integer for } i = 1, 2, \ldots, N
\]

3. Experiment Results

In this section, the performance of the two approaches is evaluated by computational experiments on randomly generated instances. All algorithms are coded in C++ with Visual Studio 6.0 compiler and the parameters are given in Table 1. In the case of the SSA, optimal \(R\) and \(Q\) have been calculated by the algorithm described before. For the GSA, results have been obtained by the algorithm presented in Li et al (2013). The computation results of the instances are given in Table 2.

Table 1. The parameter setting

| parameter | description | values |
|-----------|-------------|--------|
| \(N\)     | the number of stocks | 10 |
| \(\lambda\) | demand rate of the system | \(U \in [1,20]\) |
| \(h_{i}^{e}\) | echelon inventory holding cost of stock \(i\) | \(U \in [1,5]\) |
| \(c_i\) | fixed order cost of stock \(i\) | \(U \in h_{i}^{e} \ *[1,5]\) |
| \(p\) | operating flexibility cost of the system | 15,20,30 |
| \(T_i\) | processing time of stock \(i\) | \(U \in [1,10]\) |
| \(a\) | customer service level | 0.85, 0.95 |
| \(s_1\) | lower bound of outbound service time for stock 1 | \(U \in [1,3]\) |

Table 2. Results for solving \(R\)-problem and \(Q\)-problem

| Model     | Max/average computation time in seconds |
|-----------|-----------------------------------------|
|           | \(R\)-problem | \(Q\)-problem |
| SSA model | 0.8311/0.4567 | 1.1517 / 0.1775 |
| GSA model | 0.0041/0.0024 | 0.0011 / 0.0007 |

From table 2, we can observe that the GSA model has obviously advantages than SSA model in terms of the computation time. In order to gain further insights into the performances of the two approaches, we identify an important value, operating flexibility costs. The results are given in table 3.
Table 3 Comparison results of the operating flexibility cost for the two approaches

| $a$ | $p$ | SSA model | GSA model |
|-----|-----|-----------|-----------|
| 0.85 | 15  | 152.39    | 157.673   |
|      | 20  | 146.872   | 147.593   |
|      | 30  | 138.274   | 126.901   |
| 0.9  | 15  | 96.423    | 233.449   |
|      | 20  | 73.458    | 72.876    |
|      | 30  | 71.758    | 69.436    |

From table 3, we can demonstrate that with an increase in the parameter $p$, the operating flexibility costs of the two models will decrease. In addition, the cost differences between the SSA and GSA are quite small.

4. Conclusion

In this paper, we have studied a serial inventory system with Poisson demand, fixed order costs, and operating flexibility costs, the system is controlled by an $(R,Q)$ inventory policy. We compared two competing approaches, the stochastic-service approach (SSA) and the guaranteed-service approach (GSA). Our study has demonstrated that some advantages of the GSA model for the optimization of serial inventory systems with fixed order costs.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (Program No.71401133), National Science Basic Research Plan in Shaanxi Province of China (Program No. 2015JQ7280).

References

[1] Clark A., Scarf H., Optimal policies for a multi-echelon inventory problem, Management Science, 1960, 6: 475-490.
[2] Graves S. and Willems S., Strategic safety stock placement in supply chain Proceedings of the MSOM Conference, Hanover, NH, 1996.
[3] Graves S., Willems S., Optimizing strategic safety stock placement in supply chain, Manufacturing and Service Operations Management, 2000, 2: 68–83.
[4] Gallego, G., 1997. New bounds and heuristics for (Q, R) policies. Management Science, 1997,44, :219-233.
[5] Li Peng, Chen Haixun and Che Ada, Optimal Batch Ordering Policies for Assembly Systems with Guaranteed Service, International Journal of Production Research, 2013(51): 6275–6293.
[6] Klosterhalfen, S., Minner, S., Safety stock optimization in distribution systems: a comparison of two competing approaches, International Journal of Logistics: Research and Applications, 2010, 13: 99–120.
[7] Simchi-Levi D., Kaminsky P., The generic methods for evaluating stochastic multi-echelon inventory systems, Working Paper, Rutgers University, Neward, NJ, 2007.
[8] Zheng, Y.S. On Properties of Stochastic Inventory Systems. Management Science, 1992, 38: 87–103.