Safety Stock Placement in Capacitated Supply Chains

Carles Sitompul, El-Houssaine Aghezzaf, Wout Dullaert, Hendrik van Landeghem

To cite this version:


HAL Id: hal-00512978
https://hal.science/hal-00512978
Submitted on 1 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Safety Stock Placement in Capacitated Supply Chains

<table>
<thead>
<tr>
<th>Journal:</th>
<th>International Journal of Production Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>TPRS-2006-IJPR-0392.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Manuscript</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>05-Dec-2006</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Sitompul, Carles; Ghent University, Industrial Management Aghezzaf, El-Houssaine; Ghent University, Industrial Management Dullaert, Wout; University of Antwerp, Institute of Transport and Maritime Management Antwerp Van Landeghem, Hendrik; Ghent University, Industrial Management</td>
</tr>
<tr>
<td>Keywords:</td>
<td>SAFETY STOCKS, SUPPLY CHAIN MANAGEMENT</td>
</tr>
<tr>
<td>Keywords (user):</td>
<td>safety stock, capacity constraint, periodic review, supply chain</td>
</tr>
</tbody>
</table>
Safety Stock Placement Problem in Capacitated Supply Chains

CARLES SITOMPUL, EL-HOUSSAINE AGHEZZAF*, WOUT DULLAERT, HENDRIK VAN LANDEGHEM
† University of Ghent, Department of Industrial Management
Technologiepark 903, B9052 Zwijnaarde, Belgium
‡ University of Antwerp, Institute of Transport and Maritime
Keizerstraat 64, 2000 Antwerp, Belgium
(released December 2006)

Today’s highly competitive business environment forces supply chain managers to maintain high service levels while keeping inventory related costs as low as possible. Therefore, placing the right amount of safety stock at the right places in the supply chain is an important aspect of effective inventory management. This safety stock placement problem, for which some solution strategies were proposed in case of uncapacitated supply chains, becomes much more complicated when, in addition to the variability of the demand, capacity constraints also come into play. In this paper, we propose a model to locate safety stocks in a capacitated supply chain with the objective of maintaining the required service level. The underlying relationships linking excess capacity, demand variability, and service levels are analyzed to gain deeper understanding of the safety stock placement problem in capacitated supply chains. Based on these relationships a solution approach for the problem is proposed and is tested with Monte Carlo simulation.

Keywords: safety stock, capacity constraint, periodic review, supply chain

1 Introduction

Managing inventory, especially safety stocks, is one of the most challenging tasks facing supply chain managers. Decisions related to inventory locations and their corresponding levels throughout the supply chain has a fundamental impact on the service level, response time, delivery lead-time and the total cost of the supply chain. These interactions, present at each single link of the chain, render the analysis at the supply chain level much more difficult and complex. Traditionally, safety stock is exclusively stored in the final stage of the supply chain, i.e. at the retailers dealing with customer demand. As a result, the variation effect in demand is only addressed at this final stage. In reality, this variability cannot fully be addressed at the retailer stage, its effect breaks into the upstream stages of the supply chain through to the production stages up to raw material’s supply stages. Therefore, it makes sense to consider placing safety stock at some critical stage of the supply chain. Naturally, this safety stock may then consist of finished goods, semi-finished goods and raw materials.

A more efficient way to tackle the effect of demand variability in a supply chain consists of locating safety stock at a number of upstream stages. The problem then is to determine the right location (i.e. the right stage) and the right amount of stock that must be kept to ensure the required overall service level at the lowest cost. The majority of the existing models for the problem reported in the literature assume infinite capacity. Under the infinite capacity assumption, production lead-time is assumed to be independent of the production-batch size. In practice this is hardly ever the case. Actually, the problem of locating safety stocks becomes a lot more complicated if the production capacity constraint is taken into account.

In this paper we first investigate the relationships between demand variability, capacity constraints and...
safety stocks to gain some better insights on the issue of the safety stock placement in a supply chain. Based on these relationships, we propose an alternative model for safety stocks placement in a capacitated supply chain. There is an evident relationship linking capacity constraint with manufacturing and delivery lead-times. In a capacitated supply chain, lead-times depend on the ordered quantity. Thus, when the order quantity increases (or decreases) manufacturing lead-times also increase (or decrease) accordingly. To ensure a given service level, safety stock is needed to put up with lead-time demand variability. Once these relationships are determined, an appropriate strategy can be deployed to efficiently place safety stock.

The remainder of this paper is structured as follows. Section 2 provides a brief literature review on the problem of safety stock placement in supply chains. Section 3 shows the effect of capacity in a single-stage capacitated chain through simulation. Based on the results obtained, a modified base stock strategy is proposed. In section 4 a capacitated n-stage serial chain problem is formulated and discussed. Computational results and analysis are provided in Section 5. Section 6 presents some conclusions and some directions for further research.

2 Literature Review

The existing models for safety stock placement in uncapacitated supply chains can be categorized according to the structure of the supply chain. A supply chain can be modelled as a serial chain, a divergent system, a convergent system or a general multi-echelon system. A divergent system has a single central stage and several successors. A serial chain is a special case of the divergent systems as each stage has a single successor. A convergent system has a one-end stage with several predecessors. A general network or a general multi-echelon system is a combination of the above structures.

Early research focused mainly on serial systems (Simpson 1958, Clark and Scarf 1960). Simpson (1958) modelled a serial production chain as a sequence of in-process inventories. The objective of the model is to determine the inventory service time at each production stage to satisfy the demand of its downstream stage at minimum inventory costs. The inventory service time is defined as the time guaranteed to fill the demand requested by the downstream stage. The model assumes ‘infinite capacity’ and ‘maximum reasonable demand’. The maximum reasonable demand is used to express the fact that a company’s inventory is designed to handle any demand up to a certain level. The maximum reasonable demand can be defined as the level of inventory that will satisfy demand for a certain fixed percentage of the time intervals. Under the maximum reasonable demand assumption, Simpson (1958) shows that the cost function of the model is concave and that the minimum occurs on a vertex of the solution domain. In serial chains, the time to get materials from an upstream stage (predecessor) equals the service time guaranteed by that upstream stage. Then, the lead-time for any stage is the time to get materials (i.e. the service time guaranteed by upstream stage) plus the processing time. The net replenishment time of any stage is then equal to the service time guaranteed by the upstream stage plus the processing time minus the service time to fill demand from the downstream stage. Under this serial network structure, Simpson proves that the minimum cost occurs when the inventory at each stage is either full or empty. The inventory is full if the service time is zero, which means that safety stock is set to satisfy the maximum reasonable demand during the lead-time. The inventory is empty (no safety stock is needed) if the service time is greater or equal to the lead-time. Also, Clark and Scarf (1960) study the safety stock placement problem in serial chains. Unlike Simpson (1958), they do not assume a ‘maximum reasonable demand’. Unmet demand is charged with a penalty or shortage cost. The problem is to determine the size of the safety stock at every stage such that the total holding costs and shortage costs for the entire chain are minimized. They also extend the model to a divergent system. However, the solution approach stays the same, i.e. a minimization problem over the feasible region of safety stocks.

Sherbrooke (1968) considers a divergent system consisting of a single one base to serve multiple depots (base-depot supply system) for which item demand follows a Poisson distribution. A mathematical model is developed to determine the base and depot stock levels subject to a constraint on system investment (budget constraint) and system performance. The objective is to minimize the expected number of backorders.
Rosling (1989) studies the inventory policy for assembly systems under random demands. Assembly systems are convergent systems due to the fact that one stage can have several predecessors or upstream stages. Rosling (1989) demonstrates that assembly systems can be remodelled as serial systems. His model is a generalization of the serial model of Clark and Scarf (1960) considering random demands and proportional production, and holding costs to minimize the average cost or the total discounted cost of the network. The optimal placement of safety stocks is also studied by Lagodimos and Anderson (1993). They consider multi-echelon production networks operating in an MRP system. Supporting Chakravarty and Shtub (1986), they demonstrate that the optimal positioning of safety stock is system-specific. They argue that the network structure, the demand variability and the service measure chosen are interdependent factors that may affect the optimal policy. Lagodimos and Anderson (1993) provide a theoretical basis of the optimal positioning policy for serial and divergent networks. Using numerical analysis, they claim that putting all safety stock at the upper echelon is always optimal for serial networks and some divergent networks under certain conditions. They also suggest to pay more attention to the issue of lot sizing as it may affect the optimal safety stock policy.

Inderfurth and Minner (1998) formulate the safety stock location problem for a system with a normally distributed demand, without internal delays and every stage satisfying the service level constraint. It is shown that the optimal policy depends on the structure of the multi-stage system and the service measure used. Two service measures are considered in this paper, i.e. \( \alpha \)-type and \( \gamma \)-type service levels. The \( \alpha \)-type service level is defined as the probability of demand that can be met from inventory. Thus, safety stock factor is the \( \alpha \)-quantile of the standard normal distribution which is independent of the coverage time. Safety stock coverage time denotes the time span for which protection against uncertain demands is guaranteed. The \( \gamma \)-type service level is an extension of well-known fill-rate measure (Silver et al. 1998). It is defined as 1 - (expected backorders at the end of a period/expected demand of a period). The \( \gamma \)-type service level depends on the coverage time and its computation is quite difficult. The \( \alpha \)-type service level leads to a class of optimization problems that can be solved by the ‘either full or empty’ inventory policy (Simpson 1958). When using the \( \gamma \)-type service level, the problem is to determine the optimal coverage time such that the total cost is minimized. Inderfurth and Minner (1998) demonstrate some properties of the problem for serial, convergent and divergent systems that allowed them to formulate for each system a simplified optimization model. As a result, for system type only a finite number of coverage time combinations need to be taken into account.

Graves and Willems (2000) formulate the safety stock placement problem as a network optimization problem and develop a dynamic programming algorithm for its solution. Their work is closely related to Simpson (1958) and Inderfurth and Minner (1998). The demand is bounded by the maximum possible demand over the net replenishment time for the stage. For a demand smaller or equal to this bounded demand, every stage provides 100% service to its customers. Service levels to both external and internal stages are expressed by means of guaranteed service times. Thus, the safety stock placement problem is to determine the guaranteed service time that minimizes the total holding costs subject to the precedence constraints of the supply chain. This optimization problem consists of minimizing a concave objective function over a bounded convex set where the optimum occurs at an extreme point of the feasible region. For different network structures of the supply chain, solution algorithms are provided. For serial chains, the problem is equivalent to a shortest path problem. Dynamic programming is suggested to solve the problem when the supply chain is an assembly network or, as is generally the case, a spanning tree. Lesnaia et al. (2004) formulate the safety stock problem in supply chains as a general network problem. The safety stock problem in a general network is shown to be NP-hard, limiting exact solution techniques only to small-scale problems. Branch and Bound is suggested as a solution technique, but no dedicated approach is offered for large-scale problems.

Early attempts at solving the safety stock problem with capacitated production can be attributed to the work of Federgruen and Zipkin (1986a,b). They model the inventory problem with limited production capacity and uncertain demands. The problem is to determine the optimal base-stock level under a periodic-review policy. The objectives of the model are to minimize the average cost Federgruen and Zipkin (1986a) and the discounted cost Federgruen and Zipkin (1986b) over the planning horizon. The average cost is the long-run expected cost of production and holding costs. Thus, the model assumes an infinite planning horizon.
horizon. The discounted cost criterion is used for both infinite and finite horizon of planning. Future costs are discounted by a parameter, \( \alpha \) where \( 0 < \alpha < 1 \). Both papers suggest a modified base stock policy, which is stationary and optimal. The modified base stock policy includes the capacity constraint. When the inventory level falls below a critical number, stock is produced to bring total stock up to the base stock level or as close to it as possible, given the limited capacity. It is shown that the average costs and discounted costs are convex, thus optimality can be guaranteed. However, the base stock policy suggested solves the safety stock problem for one stage only. The model for a two-stage chain is developed by Anily and Federgruen (1991). They consider a two-stage production/inventory system. In the first stage, a common intermediate product is produced and possibly stored. In the second stage, the intermediate product is processed into a number of distinct finished products. The problem is to determine the replenishment interval time for the intermediate product and for each finished product such that the long-run average cost is minimized. An optimal procedure to solve the problem is presented. The model is related to the safety stock placement problem discussed in this paper, in the sense that the replenishment interval for each stage is actually coherent with the net replenishment time (Graves and Willems 2000, Lesnaia et al. 2004). Safety stock is placed to cover variation of demand during the replenishment interval.

Recent research aimed at solving the capacitated safety stock placement problem includes Mapes (1992), Roundy & Muckstadt (2000), Bouhia and Abernathy (2003), Paschalidis et al. (2004), and Wijngaard & Karaesmen (2005). Mapes (1992) shows the effect of capacity limitations on safety stock using simulation. The effect of capacity limitations is noticeable when the excess capacity is less than twice the standard deviation of demand. The results indicate that the amount of safety stock should be significantly increased to provide a given service level. Roundy & Muckstadt (2000) study the problem of determining production quantities in each period for a single item over an infinite horizon. They propose a heuristic to compute the periodic review interval for a periodic-review base stock policy. Their work is closely related to the work of Federgruen and Zipkin (1986a,b), because they modify the base stock policy to minimize the long-run costs in one stage systems. Using simulation, they show that their heuristic works as long as the coefficient of variation of demand is less than two. Bouhia and Abernathy (2003) formulate production policies to maximize one stage system performance (measured in terms of profit) under capacitated production. The production policies considered are base-stock policy and look-forward policy. The look-forward policy predicts the finished goods inventory in the future before determining the production amount. A simulation-based model is proposed to compare both ordering policies aimed at increasing the overall performance of the system. The results show that both ordering policies are statistically equivalent. The authors claim nevertheless that the look-forward policy is more advantageous if forecasting information is available.

Paschalidis et al. (2004) model an n-stage serial supply chain that produces a single product when demand is stochastic. The model is a stochastic optimization problem to set the safety stock level for a base stock policy at each stage to minimize the expected inventory cost subject to capacity constraints and the constraints of ensuring a stock out probability below a given level. A combination of large deviation and perturbation analysis is used to solve the problem. Large deviation analysis is an asymptotic-based approach to evaluate an objective measure involving ‘rare events’, e.g. for asymptotically obtaining a tight approximation of the stock out probability. This approach requires knowledge of the characteristics of the stochastic process of demand and production. In contrast, perturbation analysis estimates gradients of the objective measure to control parameters by observing an actual system or by using simulation.

Miragliotta and Staudacher (2004) propose another approach to manage uncertain demand. They recommend to determining the amount of safety stock and the capacity over sizing simultaneously. Thus, there is a clear trade-off between the reduced stock and the reduced investment in idle overcapacity. To do so, the system needs an information sharing (data set exchange) among many departments. They argue that the approach is simple and effective as a management tool. However, the issue for multi stage systems (i.e. where to keep the stock) is not yet addressed in this study.

As noted earlier, the safety stock placement model in supply chains is an NP-hard problem. From a computation point of view, exact solution approaches are not promising for large-scale problems. The single stage capacitated production problem has been studied thoroughly. However, its extension into capacitated multi-stage systems still needs some attention. This paper’s objective is to explore new approaches to solve the safety stock placement problem in a capacitated supply chain, closely related to the work of Simpson.

3 The Effect of Capacity in a Single-Stage Capacitated Supply Chain

We first start the discussion with the analysis of the effect of capacity constraint on the service level for a single-stage capacitated supply chain. Mapes (1992) discussed the effect of capacity limitations on safety stock using simulation. The results showed that the effect of limiting production capacity is to (significantly) increase the safety stock to maintain the prespecified service level. However, the study did not show how the safety stock is to be increased when capacity limitations are present. In this section we will show that to achieve a certain service level (i.e. the stock out probability), the amount of safety stock needs to be increased by a constant which is dependent on the capacity limitation, the average and the standard deviation of the demand. We simulate a single-stage capacitated supply chain under a base stock policy. At the beginning of each period, the amount to be produced is decided to be the smallest among the capacity and the quantity necessary to bring back the inventory up to the base stock level. Thus, the policy. At the beginning of each period, the amount to be produced is decided to be the smallest among

\[ t \]

the capacity and the quantity necessary to bring back the inventory up to the base stock level. Thus, the production at time \( t \) is then

\[ x_t = \max(0, \min(c, B - I_{t-1})) \]

where \( c \) is the production capacity, \( B \) is the base stock level and \( I_{t-1} \) is the inventory position at the beginning of period \( t \). Assume that the production is realized at that period, the inventory level at the end of period \( t \) is then

\[ I_t = I_{t-1} + x_t - d_t \]

where \( d_t \) is the demand in period \( t \). It is assumed that backordering is not allowed, thus unmet demand then turn into lost sales. If production capacity is not limited, the production is always able to replenish the stock up to the base stock level. In this case, if the demand follows a normal distribution, using \( z_\alpha = 2.33 \) a base stock equals to \( \mu + z_\alpha \sigma \) gives a 0.01 stock out probability, where \( \mu \) is the average of the demand and \( \sigma \) is its standard deviation. The term \( z_\alpha \) refers to the standard normal value such that the probability of meeting demand is \( (1 - \alpha) \).

In a single-stage capacitated supply chain, production cannot bring the inventory level up to the base stock level immediately. Several periods of full production may be required, during which further large demand might occur, requiring still more time to return to a normal inventory level (Federgruen and Zipkin 1986a,b). However, Glasserman and Tayur (1994) show that the base stock policy is stable if the mean demand per period is smaller than the production capacity per period. Since the stock out probability for capacitated single-stage is quite difficult to determine beforehand, simulation is first used to observe the effect of capacity limitation on the stock out probability. We know that the production is dependent on the characteristic of demand (the average and the standard deviation) and the production capacity. For ease of use, the system behaviour can be described then by a single parameter \( \rho \), where

\[ \rho = \frac{c - \mu}{\sigma} \]

(1)

The parameter \( \rho \) represents the excess capacity over the standard deviation of demand. From Equation 1, the system becomes an uncapatcitated system as \( \rho \) increases to infinity. Using safety stock equals to \( z_\alpha \sigma \) guarantees the stock out probability of \( \alpha \) (\( \alpha = 1\% \) in the simulation case, i.e. \( z_\alpha = 2.33 \)). If \( \rho \) is too small the system will ‘explode’ because the production will not be able to bring the inventory level back to the base stock level. In this case, the stock out probability will steeply increase and the service level will plummet. To achieve the same stock out probability as in the uncapatcitated system, the capacitated system would require an infinite amount of safety stock.

We simulate three cases which correspond to demand processes with the same average (i.e. 100 units per period) and three levels of variability (i.e. standard deviation = 10, 20 and 40 units). Each simulation run consists of 4000 periods and repeated four times. The stock out probability is then averaged. Figure 1 shows the effect of \( \rho \) on stock out probability when using the same base stock level as in the uncapatcitated system. The figure indicates that a system where \( \rho > 1.5 \) behaves as an uncapatcitated system because the stock out probability of the system is the same as the stock out probability for the uncapatcitated system.

The stock out probability increases significantly when \( \rho \) falls below 1 or when the excess capacity cannot cover one time standard deviation.
Given that a system with $\rho$ above 1.5 behaves as an uncapacitated system, it is useful to further analyse the system with $\rho$ less than 1.5. For each simulation with the same setting as above, the safety stock level (hence the base stock level) is gradually increased such that the stock out probability approaches the stock out probability for uncapacitated system, i.e. 0.01 stocks out in the simulated cases. Each case is simulated four times and the corresponding safety stock is averaged and plotted in Figure 2. Figure 3 shows the ratio of these safety stocks over the safety stock under the uncapacitated system, $z_\alpha \sigma$.

The amount of safety stock has to be increased by a correction factor such that the stock out probability is the same as in the uncapacitated single stage system, i.e. the system will experience a stock out probability 100$\alpha$%. If $B_u = \mu + z_\alpha \sigma$ is the base stock level in an uncapacitated stage then increasing the safety stock by a correction factor gives the base stock level for a capacitated stage as follows:

$$B_c = \mu + \theta z_\alpha \sigma$$

(2)

where $B_c$ is the base stock level for a capacitated single stage and $\theta$ is the correction factor given the parameter $\rho$. The rational behind this heuristic approach is that to maintain a certain service level when capacity limitation exists, it is necessary to produce more, if possible, so that it can be used to cover some of the variation of demand in the future.

Table 1 enlists the values of the correction factor obtained by averaging the value in Figure 3 for the three demand variability levels. For $\rho$ values that are not in the table, exponential interpolation can be used to estimate the required correction factor. We use a non linear regression, i.e. the exponential function to determine the correction factor given the parameter $\rho$. The exponential regression gives a high correlation between values in Table 1 and the regression value (R-squared = 0.996). The correction factor is given as follow

$$\theta = 1 + 5.25e^{-5.25(\rho-0.075)}.$$%

(3)

The correction factor as shown in Equation 3 is bounded by the value 1 as the parameter $\rho$ increases, i.e. when the capacity is unlimited.
Figure 2. Safety stock for a stock out probability of 1%

Figure 3. Capacitated safety stock over the safety stock for uncapacitated system
Table 1. Correction factor, θ

<table>
<thead>
<tr>
<th>ρ</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>1.0</td>
<td>1.1</td>
<td>1.2</td>
<td>1.3</td>
<td>1.4</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

θ* = 5.7251 3.5694 2.5832 2.0288 1.7015 1.4869 1.3385 1.2402
1.1767 1.1418 1.1043 1.0882 1.0703 1.0569 1.0479

*Accuracy is subject to numerical error.

Figure 4. An n-stage serial supply chain.

4 N—Stage Capacitated Serial Supply Chain

4.1 Problem Description

Consider an n-stage serial chain in Figure 4, where stage 1 is facing an external demand while stage n has an abundant supply of raw materials. In the terminology of Graves and Willems (2000), the capacitated safety stock placement problem can be formulated as follows. The outbound service time, $S_{j}$, is the time guaranteed by stage $j$ to satisfy its downstream demand. We assume that we want to fill customers demand immediately, thus $S_1 = 0$. The inbound service time, $SI_j$, is the time for stage $j$ to get input material from its upstream stage. In a serial chain, inbound service time stage $j$ equals the outbound service time of its upstream stage. Since stage $n$ has unlimited supply of raw materials then the inbound service time for stage $n$, $SI_n = 0$. Assume that Stage 1 faces normally distributed demand with average $μ$ and standard deviation $σ$ and that one unit of production in any stage requires one unit input from its upstream stage.

It is also assumed that there is no delay on demand information.

If the production time for stage $j$ is constant $T_j$, then the replenishment time equals $SI_j + T_j$. Thus, the net replenishment time, $τ_j$, is equal to $SI_j + T_j - S_j$ which under ‘maximum allowable demand’ assumption, the base stock is equal to $B_j = μτ_j + z_ασ\sqrt{T_j}$, i.e. to cover demand exposure during this time interval (Simpson 1958, Graves and Willems 2000). If $τ_j \leq 0$ then a 100 % service level can be provided even without placing any safety stock. In this case, the stage would delay each order on its supplier by $S_j - SI_j - T_j$, so that the supplies will arrive when needed (Graves and Willems 2000). It is not surprising that when capacity is limited, such strategy will cause the stock out probability to increase steeply.

We study the capacitated inventory system under a common review period policy. At the beginning of each period (review period = 1), a production decision is taken to bring the stock up to the base-stock level when it is possible given the capacity constraint. In each period, the decision to produce for stage $j$ is defined by $x_{j,t} = \max(0, \min(c_j, B_j - I_{j,t-1}))$, where $B_j$ is the base stock level for stage $j$ and $I_{j,t-1}$ is the inventory level of stage $j$ at the beginning of period $t$. By definition of $x_{j,t}$, production time required to produce $x_{j,t}$ is one period ($x_{j,t} \leq c_j$). Actually, $x_{j,t}$ will be effectively produced after $SI_j + 1$ period ($SI_j$ is the inbound service time for stage $j$, i.e. the time to get input material from its upstream stage). Inventory at the end of period $t$ for stage $j$ is the inventory at the beginning of period $t$ plus the realization of production at period $t$ minus the demand delivery at period $t$. Demand at period $t$ is observed at the end of the period. The realization of production at period $t$ is the production decision taken at the beginning of period $t - SI_j - 1$. The demand delivery at period $t$ is the demand observed at period $t - S_j$, where $S_j$ is the time guaranteed by stage $j$ to fill demand of its downstream stage (outbound service time). There are two possibilities regarding the outbound service time ($S_j$) and the effective production time ($SI_j + 1$), i.e.

(i) $S_j < SI_j + 1$ corresponding with the positive net replenishment time,
(ii) $S_j \geq SI_j + 1$ corresponding with the non positive net replenishment time.

http://mc.manuscriptcentral.com/tprs Email: ijpr@lboro.ac.uk
Figure 5 shows the stock replenishment and demand delivery process, where (i) shows the process for the strictly positive net replenishment time and (ii) for the non positive net replenishment time. We define the net replenishment time for stage $j$, $\tau_j$ as follows:

$$\tau_j = 1 + SI_j - S_j.$$  \hfill (4)

The net replenishment for stage $j$ is corresponding with the time-lag between the production realization and the demand delivery. If the outbound service time of stage $j$ is strictly smaller than the effective production time of stage $j$, or $S_j < SI_j + 1$ then we get a strictly positive time-lag between the production realization and the demand delivery (see Figure 5(i)). This time-lag is thus corresponding to the positive net replenishment time. If the net replenishment time is positive then the base stock policy is to bring the stock (on hand and on order) up to the base stock level. The base stock level for stage $j$ is set to satisfy demand during the net replenishment time,

$$B^c_j = \mu (1 + SI_j - S_j) + \theta_j z_{\alpha} \sigma \sqrt{1 + SI_j - S_j},$$  \hfill (5)

where the safety stock for the capacitated stage $SS^c_j$ is given by

$$SS^c_j = \theta_j z_{\alpha} \sigma \sqrt{1 + SI_j - S_j}. $$  \hfill (6)

Recall that during the net replenishment time, the standard deviation of demand is $\sigma \sqrt{\tau_j}$ and the excess capacity during the net replenishment time is equal to $\tau_j(c - \mu)$, thus Equation 1 becomes

$$\rho_j = \frac{\tau_j(c - \mu)}{\sigma \sqrt{\tau_j}}.$$
or replacing \( \tau_j \) by \( 1 + SI_j - S_j \),

\[
\rho_j = \frac{(c_j - \mu) \sqrt{1 + SI_j - S_j}}{\sigma}.
\]  

(7)

Substituting \( \rho_j \) into Equation 3, the correction factor for stage \( j \) as a function of the parameter \( \rho_j \) is then given as follow

\[
\theta_j = 1 + 5.25e^{-5.25(\rho_j - 0.075)}.
\]  

(8)

The stock out probability is approximately \( \alpha \) if we use the base stock level for capitacitated stage \( j \) as given in Equation 5 with the parameter \( \rho_j \) as defined in Equation 7 and the correction factor \( \theta_j \) as given in Equation 8.

If the outbound service time of stage \( j \) is equal to or larger than the effective production time of stage \( j \), or \( S_j = SI_j + 1 \) then it is better to delay order so that the supplies arrive when needed. In Figure 5(ii), delaying order as much as \( S_j = SI_j - 1 \) periods brings the production realization coincides with the demand delivery. In an uncapacitated stage, the safety stock is not needed because this production realization is not limited by the capacity. However, in a capitacitated stage delaying order may be useful if there is enough stock such that the capacity plus the stock would meet demand \( 100(1 - \alpha)\% \) of the time, otherwise stock out probability will increase highly. Thus, we want the safety stock plus the capacity is able to meet the maximum demand for one period, \( SS_j = c_j + \mu + z_\alpha \sigma \) or \( SS_j = z_\alpha \sigma - (c_j - \mu) \). If the excess capacity in one period \( (c_j - \mu) \) is greater than \( z_\alpha \sigma \) then there is no need to put safety stock. Therefore, the base stock policy for non positive net replenishment time is to produce as much as the demand needed to be delivered plus an amount to bring current stock up to the safety stock given the capacity limitation. Again, we need a correction factor to maintain the stock out probability to \( \alpha \) level. In this case, the correction factor is needed only for one period of production as given in Equation 1, or

\[
\rho_j = \frac{c_j - \mu}{\sigma}.
\]  

(9)

The base stock level for non positive net replenishment time is therefore set to

\[
B^c_j = \mu + \theta_j(max(0, z_\alpha \sigma - (c_j - \mu))),
\]  

(10)

where the safety stock for a capitacitated stage \( SS^c_j \) equals to the base stock. Replacing \( c_j - \mu \) with \( \rho_j \sigma \) (see Equation 9), the safety stock for a capitacitated stage with a non positive net replenishment time is given by

\[
SS^c_j = \theta_j \sigma(max(0, z_\alpha - \rho_j)),
\]  

(11)

where \( \theta_j \) is also given as in Equation 8. The probability of stock out is \( \alpha \) if we use the base stock level as in Equation 10 for non positive net replenishment time, where \( \rho_j \) and \( \theta_j \) are given by Equation 9 and Equation 8 respectively.

The operating chart of the safety stock as given in Equation 6 and Equation 11 is shown in Figure 6. The figure is based on the average demand, \( \mu = 100 \), standard deviation, \( \sigma = 10 \), and the standard normal value, \( z_\alpha = 2.33 \). In uncaptacitated system (see for \( c = 240 \)), the safety stock depends on the net replenishment time, i.e. the square root function of the net replenishment time. In capitacitated systems, the safety stock is a function of net replenishment time and capacity. The benefit of delaying an order (i.e the net replenishment time is zero or negative) is significant when the capacity is larger. If capacity is small, we still need a certain amount of safety stock to maintain the stock out probability at a prespecified level. The effect of capacity is also noticeable when the net replenishment time is small because the parameter \( \rho_j \) increases as the net replenishment time decreases (see Equation 7), hence the correction factor \( \theta_j \) is also increasing.
4.2 The Model

Based on Section 4.1, we can formulate the safety stock placement problem for an $n$–stage capacitated serial supply chains as follows

Minimize

$$\sum_{j=1}^{N} (h_{j} SS_{j})$$

subject to

$$\rho_{j} = \begin{cases} \frac{(c_{j}-\mu)}{\sqrt{1+SI_{j}-S_{j}}} & \text{if } 1 + SI_{j} - S_{j} > 0, \forall j = 1, 2, \ldots, N, \\ \frac{(c_{j}-\mu)}{\sigma} & \text{if } 1 + SI_{j} - S_{j} \leq 0, \forall j = 1, 2, \ldots, N, \end{cases}$$

$$\theta_{j} = 1 + 5.25e^{−5.25(\rho_{j}−0.075)}, \forall j = 1, 2, \ldots, N,$$

$$SS_{j} = \begin{cases} \theta_{j} z_{\alpha} \sigma \sqrt{1+SI_{j}−S_{j}} & \text{if } 1 + SI_{j} - S_{j} > 0, \forall j = 1, 2, \ldots, N, \\ \theta_{j} \sigma (\max(0, z_{\alpha} - \rho_{j})) & \text{if } 1 + SI_{j} - S_{j} \leq 0, \forall j = 1, 2, \ldots, N, \end{cases}$$

$$S_{1} = 0,$$

$$S_{n+1} = 0,$$

$$S_{j} \leq M_{s}, \forall j = 2, \ldots, N,$$
where

\[ h_j \] denotes the holding cost per unit inventory at stage \( j \),
\[ c_j \] is the capacity of stage \( j \),
\[ \mu \] is the average of demand,
\[ \sigma \] is the standard deviation of demand,
\[ \rho_j \] denotes the excess capacity over standard deviation during net replenishment,
\[ \theta_j \] is the correction factor for capacitated stage \( j \) given parameter \( \rho_j \),
\[ z_\alpha \] is the standard normal distribution value (i.e. \( z_\alpha = 2.33 \) for 0.01 stock out probability),
\[ M_s \] is the maximum service time for stage \( j = 2, ..., N \),
\[ S_j \] is the outbound service time guaranteed by stage \( j \),
\[ SI_j \] is the inbound service time of stage \( j \) (for a serial chain, \( SI_j = S_{j+1} \)),
\[ SS_j \] is the amount of safety stock placed for stage \( j \).

Recall that in a serial chain, the inbound service time of stage \( j \) is equal to the outbound service time of its downstream stage or \( SI_j = S_{j+1} \). Thus, the notation \( SI_j \) at Equation 13 and Equation 15 is replaced by \( S_{j+1} \). Equation 12 shows the objective function, i.e. minimizing the expected cost of the inventory due to holding safety stock. The parameter \( \rho_j \) is given by Equation 13 which depends on the net replenishment time. As we see in the previous section, we need to increase the safety stock by a correction factor \( \theta_j \).

Equation 14 denotes the function of the correction factor given a parameter \( \rho_j \) for stage \( j \). In our problem, since the parameter \( \rho_j \) and \( \theta_j \) depend on the service times, they also become decision variables. Equation 16 and Equation 17 shows the need to immediately fill the customers demand from the inventory of stage 1, and the time to get raw material for stage \( j \) is also zero. Equation 18 shows that the outbound service time \( (j = 2, ..., N) \) is limited to a certain period.

From Equation 13 to 15, we know that if \( c_j \to \infty \) then the correction factor is one and the safety stock is then \( z_\alpha \sigma \sqrt{1 + SI_j - S_j} \) if \( 1 + SI_j - S_j > 0 \) and 0 otherwise. This is as expected for uncapacitated systems.

We model the above problem as a shortest path problem because of the special structure of the serial chain and its solution domain (as given by Equation 16 to 18). Let the set of nodes \( V = \{ S_j = l, \forall j = 2, ..., n, \forall l = 0, 1, ..., M_s \} \), where \( S_j = l \) means that the outbound service time of stage \( j \) is equal to \( l \) period. The origin node is given by \( S_{n+1} = 0 \), i.e. the outbound service time for stage \( n + 1 \) equals to zero. The destination node is given by \( S_1 = 0 \), i.e. the outbound service time for stage 1 equals to zero. The arc cost \( \text{cost}_j(k, l) \) between two nodes represents the costs of holding safety stock for stage \( j \) when the inbound service time of stage \( j \), \( S_{j+1} = k \) and the outbound service time \( S_j = l \) are chosen in the path, where \( (k, l) \in \{ 0, 1, ..., M_s \} \). The cost between two nodes for stage \( j \), \( \text{cost}_j(k, l) \) is defined by

\[
\text{Cost}_j(k, l) = \begin{cases} 
    h_j(1 + 5.25e^{-5.25\frac{(\mu - \mu_j)^2}{\sigma^2} - 0.075})z_\alpha \sigma \sqrt{1 + k - l} & \text{if } 1 + k - l > 0, \\
    h_j(1 + 5.25e^{-5.25\frac{(\mu - \mu_j)^2}{\sigma^2} - 0.075})\sigma(\max(0, z_\alpha - \frac{\mu - \mu_j}{\sigma})) & \text{if } 1 + k - l \leq 0.
\end{cases}
\]

Equation 19 is obtained from substituting parameter \( \rho_j \) in Equation 13 into \( \theta_j \) in Equation 14, substituting \( \theta_j \) into \( SS_j \) in Equation 15, and substituting \( SS_j \) into the cost function.

The problem is then to find a path from the origin node \( S_{n+1} = 0 \) to the destination node \( S_1 = 0 \) which gives the minimum total cost. Figure 7 shows the shortest path model for a 3-stage serial chain.

5 Computational Results

We test the model for a 3-stage serial chain. Stage 1 is facing a normally distributed demand with average \( \mu = 100 \) and standard deviation \( \sigma = 10 \). The capacity for each stage is set to 3 different levels, i.e. 102, 110, and 124 unit per period. We are using \( z_\alpha = 2.33 \) to determine the service level (i.e. the stock out probability = 0.01). The holding costs are \( \varepsilon 30, 20, \) and 10 for stage 1, 2, and 3 respectively. The outbound service time for stage 1, \( S_1 \) is zero, i.e. demand from external customers is delivered immediately. The outbound service time for the other stages \( (j = 2, 3, ..., n) \) is assumed to be not larger than 3 periods. The inbound service time for stage 3, \( S_3 \) is also zero because it is assumed to have an abundant supply of raw
material. The results are shown in Table 2.

There are some conclusions that can be drawn from the computational results in Table 2. In Cases (1-9), the capacity of stage 1 is very limited \((c_1 = 102)\). Because the outbound service time of stage 1 is equal to zero \((S_1 = 0)\) and the inbound service time \(S_1 \geq 0\) then the net replenishment time of stage 1, \(\tau_1\) is never less than one period \((\tau_1 = 1 + S_1 - S_1)\). Increasing the net replenishment time for stage 1, \(\tau_1\) gives the opportunity for stage 2 to reduce its safety stock through increasing the outbound service time for stage 2 \((S_1 = S_2)\). Increasing the net replenishment time for stage 1 is a sound choice because the required increase of the safety stock is quite small, in particular when the capacity is very limited (see Figure 6). Furthermore, the need of safety stock in stage 2 is reduced because its net replenishment time is decreasing. For Cases (1-9), stage 2 and stage 3 have non positive replenishment time, therefore their safety stocks are defined only by the capacity. The safety stock for stage 2 and stage 3 in Case 9 is zero because their production capacity can be considered unlimited \((c_j \geq \mu + z_\alpha \sigma)\).

The capacity of stage 3 in Case 10 is limited \((c_3 = 102)\), while in Case 12 can be considered unlimited \((c_2 = 124)\). In Case 10, all stages need to have a certain level of safety stocks to maintain a prespecified stock out probability. In Case 12, it is better not to put any stock at stage 3 if the net replenishment time

<table>
<thead>
<tr>
<th>Case</th>
<th>(c_1)</th>
<th>(c_2)</th>
<th>(c_3)</th>
<th>(S_1)</th>
<th>(S_2)</th>
<th>(S_3)</th>
<th>(\tau_1)</th>
<th>(\tau_2)</th>
<th>(\tau_3)</th>
<th>(\theta_1)</th>
<th>(\theta_2)</th>
<th>(\theta_3)</th>
<th>(SS_1)</th>
<th>(SS_2)</th>
<th>(SS_3)</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>102</td>
<td>102</td>
<td>102</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1.9531</td>
<td>3.7237</td>
<td>3.7237</td>
<td>91</td>
<td>79</td>
<td>79</td>
<td>5109.87</td>
</tr>
<tr>
<td>2</td>
<td>102</td>
<td>102</td>
<td>110</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1.9531</td>
<td>3.7237</td>
<td>1.0408</td>
<td>91</td>
<td>79</td>
<td>14</td>
<td>4455.16</td>
</tr>
<tr>
<td>3</td>
<td>102</td>
<td>102</td>
<td>124</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1.9531</td>
<td>3.7237</td>
<td>1.0408</td>
<td>91</td>
<td>79</td>
<td>0</td>
<td>2868.88</td>
</tr>
<tr>
<td>4</td>
<td>102</td>
<td>110</td>
<td>102</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1.9531</td>
<td>1.0408</td>
<td>3.7237</td>
<td>91</td>
<td>14</td>
<td>79</td>
<td>3800.46</td>
</tr>
<tr>
<td>5</td>
<td>102</td>
<td>110</td>
<td>110</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1.9531</td>
<td>1.0408</td>
<td>1.0408</td>
<td>91</td>
<td>14</td>
<td>14</td>
<td>3145.75</td>
</tr>
<tr>
<td>6</td>
<td>102</td>
<td>110</td>
<td>124</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1.9531</td>
<td>1.0408</td>
<td>1.0000</td>
<td>91</td>
<td>14</td>
<td>0</td>
<td>2730.45</td>
</tr>
<tr>
<td>7</td>
<td>102</td>
<td>124</td>
<td>102</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1.9531</td>
<td>1.0408</td>
<td>1.0000</td>
<td>91</td>
<td>0</td>
<td>79</td>
<td>3523.59</td>
</tr>
<tr>
<td>8</td>
<td>102</td>
<td>124</td>
<td>110</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1.9531</td>
<td>1.0408</td>
<td>1.0000</td>
<td>91</td>
<td>0</td>
<td>14</td>
<td>2865.88</td>
</tr>
<tr>
<td>9</td>
<td>102</td>
<td>124</td>
<td>124</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1.9531</td>
<td>1.0408</td>
<td>1.0000</td>
<td>91</td>
<td>0</td>
<td>0</td>
<td>2345.75</td>
</tr>
</tbody>
</table>

Figure 7. The shortest path model.
for stage 3, \( \tau_3 \) is non positive. A non positive net replenishment time for stage 3 is resulted from increasing the outbound service time for stage 3, \( S_3 \) (hence the inbound service time for stage 2 is also increasing). Therefore, stage 2 has to put more safety stock than that in Case 10 because the net replenishment time for stage 2 is increasing (see \( \tau_2 \) in Case 10 and 12). Thus, the decision of no safety stock in stage 3 is beneficial if the overall cost is minimum.

The capacity of stage 1 in Case 18 is less than the capacity of stage 1 in Case 27. However, there is only a small difference between the total costs of these cases. From Figure 6, the operating chart for \( c = 110 \) and \( c = 240 \) (unlimited) is almost the same for a positive net replenishment time. We can conclude that if the net replenishment time increases then the capacity impact into the safety stock becomes less noticeable.

In Case 21, the capacity of stage 1 is equivalently unlimited (see Case 3 for comparison). Reducing the number of safety stock to be kept in stage 1 can be achieved through decreasing the net replenishment time \( \tau_1 \). Decreasing the net replenishment time can be achieved using two ways, i.e. increasing the outbound service time or decreasing the inbound service time. Since increasing outbound service time in stage 1 is not possible (we want to serve external customer immediately), the only way to do this is decreasing the inbound service time for stage 1. Decreasing inbound service time for stage 1 means increasing the net replenishment time for stage 2, thus increasing the safety stock in stage 2. Therefore, if the capacity increases, reducing the number of safety stock by reducing the net replenishment time is a sound choice. However, reducing the net replenishment time for one stage changes the net replenishment time for the other stage. Therefore, reducing the number of safety stock for one stage influences the need of safety stock for the other stage. In general, the optimal solution is achieved when overall costs for the supply chain is minimized.

In Case 2, the capacity of stage 2, \( c_2 \) is 102 unit per period and the capacity of stage 3, \( c_3 \) is 110 unit per period. Meanwhile in Case 4, the capacity of stage 2, \( c_2 \) is 110 unit per period and the capacity of stage 3, \( c_3 \) is 102 unit per period. From the total costs in Table 2, the minimum cost is reached when the capacity is larger at the stage near the customer (i.e. stage 2). The same phenomenon also appears in Cases (11,13) and (20,22). However, Cases (17,23) and (18,24) show a contrasting phenomenon. In Case 17 and 18, the capacity of stage 1, \( c_1 \) is 110 unit per period and the capacity of stage 2, \( c_2 \) is 124 unit per period. Meanwhile in Case 23 and 24, the capacity of stage 1, \( c_1 \) is 124 unit per period and the capacity of stage 2, \( c_2 \) is 124 unit per period. The minimum cost is reached when the capacity is smaller at the stage near the customer (i.e. stage 1). This means that having a larger capacity at the stages near the customer does not necessarily result in smaller total costs.

We can conclude that the process of placing safety stock in a capacitated serial supply chain depends on both capacity and net replenishment time. The net replenishment time of a stage in a serial chain is influenced by the outbound service time guaranteed by its upstream stage. The need of safety stock in that stage is affected by its capacity and net replenishment time. A sequence of outbound service time (for each stage) is optimal if it results the minimum of total holding costs. In general, the optimality of the solution depends on the interaction between the capacity, the net replenishment time, and the holding costs.

From the solution results, we simulate Case 6 for 4000 periods. Each run is repeated four times of different random demands (D). Table 3 shows the actual inventory on hand (\( I_j \)) and the probability of stock out \( O_j \) for each stage. We choose Case No.6 because it has three characteristics of our interest. First, it has an uncapacitated stage with a non positive net replenishment time (i.e. stage 3). Second, it has a capacitated stage with a non positive replenishment time (i.e. stage 2). Last, it has a capacitated stage with a positive net replenishment time (i.e. Stage 1). It is shown that even when the replenishment time is non positive, the safety stock is still needed to give an \( \alpha \) stock out probability if capacity is limited (see stage 2). The actual inventory on hand is smaller than the safety stock derived from our model (see the average and the model for inventory and stock out probability in Table 3). It is logical because production is not able to replenish its stock in one period of time. As a matter of fact, it is quite difficult to obtain the ‘true’ inventory level for capacitated systems. However, our model gives a good approximation of the upper bound of the safety stock to achieve a certain service level. Futhermore, it makes use the base stock policy for a capacitated system with a simple modification. The stock out probability for each stage in our
simulation is near to 0.01 when we use $z_\alpha = 2.33$.

6 Conclusion

Although supply chain managers try to reduce inventory costs, a certain level of safety stock is always needed to ensure a predefined service level if demand is subject to variation. If well-located and well-sized, the safety stock is able to protect supply chains against demand variability at an acceptable cost. Modelling the safety stock placement in capacitated supply chains is, however, rather difficult. Simulation experiments indicate that the safety stock in a capacitated stage has to be increased by a correction factor to obtain the same stock out probability as in the uncapacitated case. When the safety stock is increased (and hence the base stock), the modified base stock policy tries to produce more than under the normal policy (i.e. without taking into account capacity). If extra production is possible, it can be used as a buffer to meet future demand.

In a capacitated stage, there are two cases regarding the net replenishment time, i.e. positive and non positive net replenishment time. The safety stock is intended to cover variation of demand during this net replenishment time. Even when the net replenishment time is zero or negative, a certain level of safety stock is still needed, in particular when capacity is small. If capacity is large or equivalently unlimited (i.e. $\rho_j \geq z_\alpha$) and the net replenishment time is non positive then the safety stock is not needed because production is able to replenish stock $100(1-\alpha)$% of the time. When capacity is limited, we need to increase the safety stock by a correction factor which depends on the parameter $\rho_j$, i.e. excess capacity divided by the standard deviation of demand during the net replenishment time. The decision to guarantee services to customers or downstream stage depends on the capacity and the service time guaranteed by the upstream stage. The outbound service time of one stage is influencing the inbound service time of the next stage, hence the net replenishment time. Thus, we can say that the model optimizes the size and the location of safety stock simultaneously while maintaining a certain stock out probability.

Although our findings are based on a limited set of simulation experiments, they can offer guidance to practitioners in locating and sizing safety stock in a capacitated serial supply chain. We make use the base stock policy with a simple modification when capacity is limited. Further research will be directed towards a theoretical derivation of the correction factor.

REFERENCES


Federgruen, A., and Zipkin, P., An Inventory Model with Limited Production Capacity and Uncertain
REFERENCES


