생산-재고시스템에서 수요의 변동이 미치는 영향

김준석†

세종대학교 경영학과

The Impact of Demand Variability in Production-Inventory Systems

Joon-Seok Kim†

Department of Business Administration, Sejong University

Key Words : Production-Inventory Systems, Queueing Systems, Demand Variability, Safety Stock

Abstract

본 연구에서는 공급리드타임이 한정된 용량을 가진 생산시스템에 의해서 내부적으로 결정되어지는 다품종 재고시스템에서의 제품의 수요변동이 미치는 영향을 조사하였다. 공급리드타임이 시스템의 외부에서 결정되어 주어지는 경우와는 달리 리드타임 수요의 변동이 제품 수요의 변동이 증가할 때 실제 감소하는 결과를 보였다. 또한, 수요의 변동이 증가할 때 전체 재고 중 안전재고의 비율이 감소되는 것도 보였다. 더 나아가서 수요 변동의 크기가 어느 정도 이상으로 커지면 안전재고가 완전히 제거되는 사실도 입증하였다. 이 같은 결과는 제품 수요의 변동이 매우 클 때 리드타임 수요의 변동을 줄이기 위한 전략은 별로 효용이 없음을 보여준다.

1. Introduction

In this paper, we examine the effect of demand variability in a capacity-constrained production-inventory system with multiple items. We consider a system where inventory is continuously reviewed and controlled by a base-stock policy. Demand occurs one unit at a time according to a renewal process. Demand can be for one of K items produced by the system. Inventory for different items are kept in separate buffers. Different holding and backordering costs (or service levels) may be associated with different items. If available, an order is satisfied from stock, otherwise it is backlogged with the production system. The production system has a finite production rate and stochastic production times. Therefore, order lead times are load-dependent and affected by the current size of the order queue with the production system.

In contrast to systems with exogenous lead-times, we show that variability in lead-time demand actually decreases with demand variability. We also show that higher demand variability leads to a smaller fraction of total stock being devoted to safety stock. More significantly, we show that a sufficiently large increase in demand variability can lead to the elimination of safety stocks altogether. Our results suggest that strategies used to reduce lead-time demand variability would be less valuable when demand variability is high.

The effect of demand variability has been studied in the context of continuous review inventory systems with exogenous and stochastic lead times by Song (1994a, 1994b). She showed that for a particular definition of variability, the optimal cost and base-stock levels increase with lead-time demand

† 교신저자 jskim@sejong.ac.kr

* 이 논문은 2007년도 세종대학교 교내연구비 지원에 의한 논문임

variability. In the context of a single period newsvendor problem, Gerchak and Mossman (1992) showed that the optimal cost is increasing in demand variability when variability is varied according to a mean preserving transformation, although the optimal cost may or may not increase with higher variability. Gerchak and He (2003) analyzed the relation between the benefits of inventory pooling and the variability demand. They showed that the more random the individual demands, in the sense of the mean preserving transformation, the larger the benefits from consolidating them. Using an example, they show that such may not be the case for other forms of increased variability.

Jemai and Karaesmen (2005) studied the effect of demand variability on the performance of a make-to-stock queue. They show that when order inter-arrival times are ordered in a stochastic sense, an increase in inter-arrival time variability leads to higher base-stock levels and higher costs (in section 3, we generalize this result to systems with multiple items). The effect of variability in queueing systems is widely studied. Higher variability is generally associated with deterioration in performance (Wolff, 1989). Various stochastic orderings involving queueing systems are discussed in Stoyan (1983), Shaked and Shanthikumar (1994) and Benjaafar et al. (2005).

The rest of the paper is organized as follows. In section 2, we describe our model and characterize the optimal base stock levels and the optimal cost. In section 3, we examine the effect of demand variability on the distribution of lead-time demand. In section 4, we consider the relation between safety stocks and demand variability. In section 5, we offer concluding comments.

2. Model Description

We consider a production-inventory system where demand occurs one unit at a time according to a renewal process with rate λ. The inter-arrival time between orders is a random variable denoted by X, with $E(X)=1/\lambda$. Demand can be for one of K items produced by the system. An order is for an item of type i with probability p_i, where $i=1,...,K$. Separate inventory buffers are kept for each item. If available, an order is always satisfied from buffer stock. If not, the demand is backordered. The system incurs a holding cost h_i per unit of inventory of type i per unit time and a backordering cost b_i per unit of type i backordered per unit time. Alternatively, a service level may be specified for each item. The inventory buffer of each item is managed according to a continuous review base-stock policy with base-stock level s_i for item i. This means that the arrival of each new order triggers the placement of a replenishment order with the production facility. Replenishment orders at the production facility are processed on a first come-first served basis. The production facility can process only one order at a time and orders that arrive when the facility is busy must wait in a queue. We assume that production times are *i.i.d.* exponentially distributed random variables with mean $1/\mu$. In order to ensure stability, we assume that $\rho\equiv\lambda/\mu<1$ where ρ is the utilization of the production facility. When viewed in isolation, the number of orders at the production facility forms a GI/M/1 queue. In combination with the inventory buffers, our system is an example of a make-to-stock queue (Buzacott and Shanthikumar, 1993).

In our model items are not differentiated by processing times, although they may differ in features, functionality, or components. This assumption is satisfied in industries where different grades of the same component are used to differentiate items (e.g. the electronics and computer industries) and manufacturing/assembly times are not significantly affected by component. It is also satisfied in industries where products have been redesigned so that the customization step is carried out late in the production process and represents only a small fraction of total production time. For our analysis, the assumption of homogenous production times allows us to isolate more readily the effect of the demand variability.

We assume that a base-stock policy consisting of a vector $s=\{s_i: i=1,\dots,K\}$ is chosen so that the long run expected total cost per unit time is minimized. We denote this expected total cost by

$$z(\mathbf{s}) = \sum_{i=1}^{K} z_i(s_i) = \sum_{i=1}^{K} E[h_i I_i + b_i B_i] \tag{1}$$

where I_i and B_i are random variables equal in distribution to, respectively, the steady-state inventory and backorder levels for each of the K items and $z_i(s_i)=E[h_i I_i+b_i B_i]$ when base stock level s_i is used for item i. Alternatively, a base-stock policy may be chosen so that only the holding cost component of expected total expected is minimized with a requirement that a specified service level for each item is met. A service level can be specified in several ways, including setting an upper bound on the probability of a stock-out or choosing a target fill rate (the fraction of orders filled from stock). In our case, since demand occurs one unit at a time, the requirement on either the stock-out probability or the fill rate can be specified as a constraint on the probability that an arriving order of type-i finds no inventory on hand. This is also equivalent to finding si units already on-order. If we let denote the number of units on-order of type-i as seen by an arrival, then the service level constraint can be stated as :

$$\Pr(\tilde{N}_i \ge s_i) \le \alpha_i, \tag{2}$$

where $\alpha_i \in [0, 1]$ is the specified service level for item i.

In Lemma 1, we show how an optimal vector of base-stock levels can be calculated for both cost-based and a service level-based models.

Lemma 1: *In a multi-item production-inventory system that fits the above description, the base stock level for item i, i = 1, ..., K that minimizes expected total cost per unit time is given by* $s_i^* = \lfloor \tilde{s}_i \rfloor$, *where*

$$\tilde{s}_i = \frac{\ln\left[\gamma_i \dfrac{\beta}{\rho}\right]}{\ln[r_i]}, \tag{3}$$

$$\gamma_i = h_i/(h_i + b_i) \tag{4}$$

$$r_i = \frac{p_i \beta}{1-\beta(1-p_i)} \tag{5}$$

and β *is the solution of the characteristic equation*

$$\beta = \int_0^{\infty} e^{-\mu t(1-\beta)} dG(t) \tag{6}$$

where G is the distribution of order inter-arrival times. When a service level is enforced, $s_i^* = \lceil \tilde{s}_i \rceil$, *where*

$$\tilde{s}_i = \frac{\ln[\alpha_i]}{\ln[r_i]}. \tag{7}$$

In order to examine the effect of demand variability, we shall make use of stochastic comparisons between different random variables. Since we are interested in comparing distributions with the same mean but with different higher moments, we shall use the following definition of variability, which is consistent with the increasing convex ordering of random variables (Shaked and Shanthikumar, 1994).

Definition 1: *We say that a random variable X is more variable than a random variable Y, and we write* $X \ge_v Y$, *if* $E[f(X)] \ge E[f(Y)]$ *for all increasing convex functions f.*

Using the above definition, it is not difficult to show that the following lemmas hold. Lemma 2 follows from our definition of variability. Lemma 3 is borrowed from Wolff (1989).

Lemma 2: *Let X and Y be two non-negative random variables such that E(X)=E(Y), then* $X \ge_v Y$ *implies* $E(X^n) \ge E(Y^n)$. *Hence, we have* $X \ge_v Y \Rightarrow C(X) \ge C(Y)$, *where C(·) denotes the coefficient of variation*

Lemma 3: *Consider two GI/M/1 queues with service rate* μ *and arrival rate* λ. *The arrival process is renewal in both queues. Let* X_j *(j = 1, 2) denote a ran-*

dom variable that describes the inter-arrival times associated with the arrival process to queue j. Then $X_1 \geq_v X_2$, implies $\beta_1 \leq \beta_2$, where β_j is the solution of the characteristic equation $\beta_j = \int_0^{\infty} e^{-\mu(1-\beta_j)} dG_j(t)$ and G_j is the distribution of X_j

We now use lemma 3 to show that higher demand variability always leads to higher base-stock levels. This generalizes the result obtained by Jemai and Karaesmen (2005) to systems with multiple items and service level constraints, as well as offer an alternative proof for the single item case.

Lemma 4: *Consider two GI/M/1 make-to-stock queues with service rate μ and arrival rate λ. The arrival process is renewal in both queues, with the probability that an arriving customer is of type i being p_i. Let X_j (j = 1, 2) denote a random variable that describes the inter-arrival times associated with the arrival process to queue j. Then $X_1 \leq_v X_2$ implies $s_i^{*(1)} \leq s_i^{*(2)}$ and $z_i(s_i^{*(1)}) \leq z_i(s_i^{*(2)})$ for i=1,..., K, where $s_i^{*(j)}$ refers to the optimal base stock level of item i in queue j.*

3. The Distribution of Lead-time Demand

In this section, we examine the impact of demand variability on the variability of lead-time demand. The lead-time demand for an item of type-*i* is the amount of type-*i* demand that arrives between the time a type-*i* order is placed and when that order is delivered to the inventory buffer. This is equal in distribution to the number of items of type-*i* seen upon completion at the production system by an order of type-*i*. Let D_i denote the lead-time demand for item-*i*, then it can be shown that:

$$\Pr(D_i = x_i) = (1 - r_i) r_i^{x_i} \tag{8}$$

Theorem 1: *Consider two GI/M/1 make-to-stock queues with service rate μ and arrival rate λ. The arrival process is renewal in both queues, with the probability that an arriving customer is of type i being p_i. Let X_j (j = 1, 2) denote a random variable that describes the inter-arrival times associated with the arrival process to queue j. Then $X_1 \leq_v X_2$ implies $C(D_i^{(1)}) \geq C(D_i^{(2)})$ where $C(D_i^{(j)})$ refers to the coefficient of variation of lead-time demand $D_i^{(j)}$ of item i in queue j for i = 1,..., K.*

Proof: From (8), we have

$$E(D_i^{(j)}) = \sum_{x_i=1}^{\infty} r_i^{(j)} (1 - r_i^{(j)})^{x_i} = \frac{r_i^{(j)}}{1 - r_i^{(j)}} \quad \text{and}$$

$$Var(D_i^{(j)}) = \frac{r_i^{(j)}}{[1 - r_i^{(j)}]^2}$$

From which, we obtain

$$C^2(D_i^{(j)}) = \frac{1}{r_i^{(j)}} = \frac{1 - \beta_j (1 - p_i)}{p_i \beta_j} \tag{9}$$

Since, by virtue of lemma 3, $\beta_1 \leq \beta_2$, we have $C^2(D_i^{(1)}) \geq C^2(D_i^{(2)})$. Hence proved. □

With theorem 1, we have shown the surprising result that higher demand variability induces smaller variability in lead-time demand. This occurs despite the fact that both the mean and variance of lead-time demand increase with variability (both are increasing in *β*). Note that the coefficient of variation is strictly decreasing in p_i. This means that items with a smaller demand rate experience higher lead time demand variability. However, in the limit case, we have $\lim_{\beta \to 1} C^2(D_i) = 1$ for all values of *i*.

Although higher demand variability has the effect of reducing the variability in lead-time demand for every item type, it does not the have the same effect on the variability of the number of items on order (that is, the number in queue + in-process at the production system). As we show in the following theorem, the effect of demand variability in this case depends on the relative magnitude of each item.

Theorem 2: *Consider two GI/M/1 queues with serv-*

ice rate μ and arrival rate λ. The arrival process is renewal in both queues, with the probability that an arriving customer is of type i being p_i. Let X_j (j = 1, 2) denote a random variable that describes the inter-arrival times associated with the arrival process to queue j. Then $X_1 \leq_v X_2$ implies $C(N_i^{(1)}) > C(N_i^{(2)})$ if $p_i < 0.5$, $C(N_i^{(1)}) = C(N_i^{(2)})$ if $p_i = 0.5$ and $C(N_i^{(1)}) < C(N_i^{(2)})$ if $p_i > 0.5$ for i = 1,..., K, where $C(N_i^{(j)})$ refers to the coefficient of variation of the number of items of type i in system j (in queue + in process).

Proof: We have

$$E(N_i^{(j)}) = \sum_{n=1}^{\infty} n \frac{\rho}{\beta_j} \{ [r_i^{(j)}]^n (1 - r_i^{(j)}) \} = \frac{\rho}{\beta_j} (\frac{r_i^{(j)}}{1 - r_i^{(j)}})$$

and

$$E([N_i^{(j)}]^2) = \frac{\rho}{\beta_j (1 - r_i^{(j)})^2} r_i^{(j)} \{1 + r_i^{(j)}\}$$

Using the above, we can now write $C^2(N_i^{(j)})$ as:

$$C^2(N_i^{(j)}) = \frac{1 - \beta_j + 2 p_i \beta_j - p_i \rho}{p_i \rho} \tag{10}$$

Taking the difference leads to

$$C^2(N_i^{(1)}) - C^2(N_i^{(2)}) = \frac{(\beta_1 - \beta_2)(2 p_i - 1)}{p_i \rho}$$

Since $\beta_1 \leq \beta_2$ we have:

$$p_i < 0.5 \Rightarrow C(N_i^1) \geq C(N_i^2)$$
$$p_i = 0.5 \Rightarrow C(N_i^1) = C(N_i^2)$$
$$p_i > 0.5 \Rightarrow C(N_i^1) < C(N_i^2)$$

which completes our proof. ▯

Theorem 2 shows that an increase in demand variability can either increase, decrease or leave unchanged the variability in the number of items on order for a particular type. From (10), we can also show that $C^2(N_i) > (2-\rho)/\rho$ for $p_i > 0.5$, $C^2(N_i) = (2-\rho)/\rho$ for $p_i = 0.5$ and $C^2(N_i) > (2-\rho)/\rho$ for $p_i > 0.5$. In the limit case of very high variability, we have $\lim_{\beta \to 1} C^2(N_i) = (2-\rho)/\rho > 1$.

Although they have important implications to the analysis of production-inventory systems, Theorems 1 and 2 are queueing results that are of independent interest. They describe how variability in the arrival process to a GI/M/1 queue affects the variability in the number of customers observed either at an arrival instant (theorem 1) or at a random instant (theorem 2). In our setting, theorem 1 is of particular importance since it implies that with higher demand variability there is less uncertainty regarding lead-time demand. In turn, this suggests that strategies, such as holding or aggregating demand, used to cope with variability are less useful when demand variability is high. In section 4, we show that this intuition is indeed true.

It is not difficult to find instances of commonly used demand distributions where the above results apply. Consider, for example, the family of Gamma distributions with parameter $1/t$ and λ/t where $t>0$. If we let X_t be a random variable that describes the inter-arrival time between orders, then $E(X_t)=1/\lambda$, and $Var(X_t)=t/\lambda^2$ so that the mean is constant but the variance is increasing in t. Hence, the random variable X_t becomes more variable when t increases. More specifically, it can be shown that if $t_1 \leq t_2$ then $X_{t_1} \leq_v X_{t_2}$ for any t_1, $t_2>0$ (Stoyan 1989). In order to examine the effect of demand variability on the variability of lead-time demand, we can vary t and observe $C(D_i)$ and $C(N_i)$. Equivalently, we can vary the coefficient of variation of order inter-arrival times $C_A \equiv C(X_t) = \sqrt{t}$ and observe its effect on $C(D_i)$ and $C(N_i)$.

A special case of a Gamma distribution is the Erlang. A transformation that varies the number of stages k while maintaining the same arrival rate λ can be shown to be consistent with the variability ordering, with $C_A = \sqrt{1/k}$. In this case, varying k allows us to vary C_A in the interval (0, 1].

It is also possible to consider more general classes of distributions by using a mean preserving transformation of the form

$$X_t = tX + (1-t)\lambda, \quad 0 \le t \le 1 \tag{11}$$

where X is a random variable with mean $E(X)=1/\lambda$. In this case, $E(X_t)=E(X)$ and $Var(X_t)=t^2 Var(X)$ where X_t becomes again more variable as t increases. In particular, we have $C_A \equiv C(X_t) = tC(X)$. The mean preserving transformation can be shown to be consistent with the variability ordering (Gerchak and He, 2003).

In a GI/M/1 queue, a relationship between the parameter β and the coefficient of variation in inter-arrival times, C_A, can be obtained by considering the following well known upper bound on the expected number of customers in the system (Wolff 1989):

$$E(N) = \rho/(1-\beta) \le [(C_A^2 + \rho^2)/2(1-\rho)] - (1-\rho)C_A^2/2 \tag{12}$$

The bound is asymptotically sharp in heavy traffic ($\rho \to 1$) and is exact for the M/M/1 and D/D/1 cases. Inequality (12) can be rewritten as

$$[2(1-\rho)/\rho(1-\beta)] - 1 \le C_A^2 \tag{13}$$

where the left hand side can be interpreted as a lower bound on C_A^2. We can see that this lower bound is increasing in β and approaches ∞ as $\beta \to 1$. Thus, β can be viewed, for fixed ρ, as an indicator of demand variability.

4. Safety Stock

In this section, we examine the effect of demand variability on safety stock. We initially limit our discussion to a cost-based model. Throughout, we shall use the following definition of safety stock.

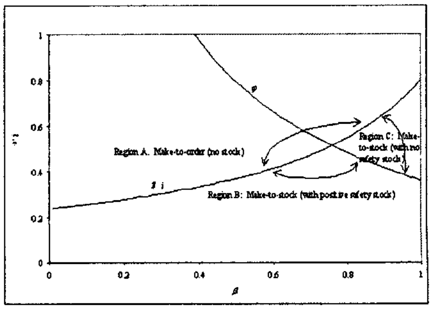
Definition 2: *An item i is said to have positive safety stock if the optimal base-stock level s_i^* for that item satisfies the inequality: $s_i^* > E(D_i)$, where $E(D_i)$ is the expected lead-time demand for item i. We let safety stock refer to the difference $ss_i^* = s_i^* - E(D_i)$.*

The above statement is in line with common definitions of safety stock. For example, Silver et al. (1998) define safety stock as "*as the average level of the net stock just before a replenishment arrives.*" That is, safety stock is the amount left (on average) after fulfilling lead-time demand.

In lemma 5, we show that the parameters Υ_i, β, ρ and p_i define three regions: region A where we produce to order, B where we produce to stock with a positive safety stock, and C where we produce to stock but do not carry any safety stock.

Lemma 5: *Consider a multi-item GI/M/1 make-to-stock queue with service rate μ and arrival rate λ. The probability that an arriving customer is of type-i is p_i. Then, $s_i^* \ge 1$, if and only if $\gamma_i \le \chi_i \equiv (\rho/\beta) r_i$ and $ss_i^* > 0$ if and only if $\gamma_i \le \phi_i \equiv (\rho/\beta)/r_i^{(r_i/(1-r_i))}$*

The proof follows from lemma 1 and definition 2. Region A is defined by cases where the tuple (Υ_i, β, ρ, p_i) satisfies $\Upsilon_i > \chi_i$, B satisfies $\Upsilon_i \le \chi_i$ and $\Upsilon_i \le \phi_i$, and C satisfies $\Upsilon_i \le \chi_i$ and $\Upsilon_i > \phi_i$. These three regions are illustrated in Figure 1. Figure 1 suggests that an increase in β could lead to a shift from a make-to-order to a make-to-stock mode of production (a shift from A to B or to C). It also suggests that an increase in β could lead to a shift from the region where we hold safety stock (B) to the one where we hold no safety stock (C). In theorems 3 and 4, we show that this is indeed the case. More significantly, we show that in regions where we hold safety stock, the fraction of total stock due to safety stock decreases with variability.



<Figure 1> Υ_i and β on safety stock inventory

(p_i=0.3, ρ=0.8)

Theorem 3: *Consider two GI/M/1 make-to-stock queues with service rate* μ *and arrival rate* λ. *The arrival process is renewal in both queues, with the probability that an arriving customer is of type i being* p_i. *Let* X_j $(j = 1, 2)$ *denote a random variable that describes the inter-arrival times associated with the arrival process to queue j. Then* $X_1 \leq_v X_2$ *implies* $\phi_i^{(1)} \geq \phi_i^{(2)}$ *for* $i=1,\ldots,K$ *whenever* $s_i^{*(j)}$ *and* $ss_i^{*(j)}>0$, *where* $\phi_i^{(j)} = ss_i^{*(j)}/s_i^{*(j)}$.

An important implication of theorem 3 is that, when safety stock is held (region *B*), the fraction of the base stock due to safety stock is smaller in systems with higher demand variability. This effect is illustrated in Figure 2 for a system with Gamma distributed demand, where C_a is varied according to the scheme described in section 3.

Before, we present our next result, we need the following lemma.

Lemma 6: *For fixed* ρ, χ_i *is increasing and convex in* β, *with* $\lim_{\beta\to 0}\chi_i = p_i\rho$ *and* $\lim_{\beta\to 1}\chi_i = \rho$, *and* ϕ_i *is decreasing in* β *with* $\lim_{\beta\to 0}\phi_i = +\infty$ *and* $\lim_{\beta\to 1}\phi_i = e^{-\rho}$.

Proof: Noting that

$$\chi_i = \frac{p_i \rho}{1-\beta(1-p_i)},$$

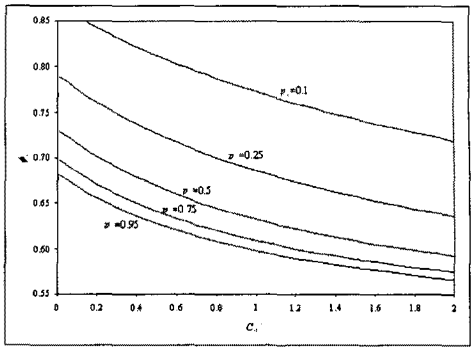
it is straightforward to show that χ_i is increasing and convex in β and the above limits hold. For ϕ_i, we first show that $\ln[\phi_i]$ is decreasing in β. This can be verified by noting that

$$f(\beta) \equiv \ln[\varphi_i] = \frac{r_i \ln[r_i]}{1-r_i} - \ln\left(\frac{\beta}{\rho}\right)$$

Since

$$\frac{df(\beta)}{d\beta} = \frac{1-r_i+\ln(r_i)}{(1-r_i)^2}\left(\frac{p_i^2}{(1-\beta(1-p_i))^2}\right) - \frac{1}{\beta} \leq 0,$$

a result that follows from $1-r_i+\ln[r_i]\leq 0$ for $0\leq r_i \leq 1$, $f(\beta)$ is non-decreasing in β. In turn, this means that ϕ_i is decreasing in β. The limits follow upon application of l'Hopital's rule.



<Figure 2> The effect of demand variability on the relative size of safety stock (ρ=0.8, Υ=0.1)

Theorem 4: *Consider two GI/M/1 queues with service rate* μ *and arrival rate* λ. *The arrival process is renewal in both queues, with the probability that an arriving customer is of type i being* p_i. *Let* X_j $(j=1, 2)$ *denote a random variable that describes the inter-arrival times associated with the arrival process to queue j such that*

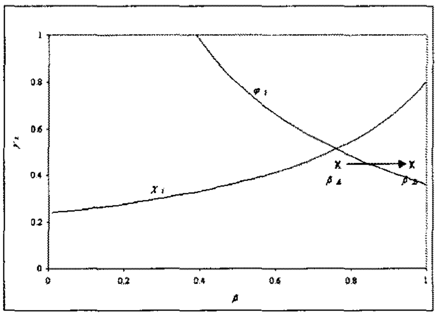
(1) $X_1 \leq_v X_2$,

(2) $e^{-\rho} < \gamma_i < (\rho/\beta_1) r_i^{(1)}$ *and* $\gamma_i < (\rho/\beta_1) r_i^{\frac{r_i^{(1)}}{1-r_i^{(1)}}}$ *for an item i,*

(3) $e^{-\rho} < \gamma_i < (\rho/\beta_2) r_i^{(2)}$ *and* $\gamma_i > (\rho/\beta_2) r_i^{\frac{r_i^{(2)}}{1-r_i^{(2)}}}$ *for the same item I,*

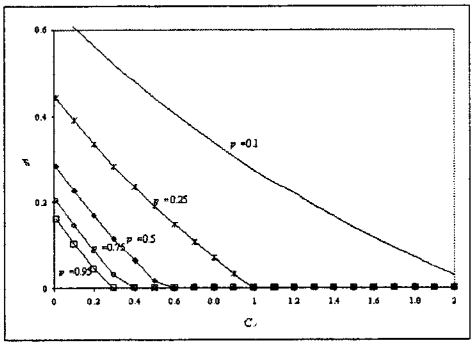
then $ss_i^{*(1)} > ss_i^{*(2)}$ *with* $ss_i^{*(2)}=0$.

Theorem 4 means that a sufficiently large increase in demand variability can lead to the elimination of safety stock. A graphical illustration of this effect is shown in Figure 3 where an increase from β_A to β_B leads to the elimination of safety stock. Numerical results from a system with Gamma distributed demand are shown in Figure 4.



<Figure 3> Increasing β from β_A to β_B eliminates safety stock

The above analysis can be extended to systems with service level constraints. With a constraint on the probability of backordering, the optimal base-stock level is always greater than one and we always produce to stock. Therefore, the parameters Υ_i, β, ρ and p_i define two regions, both make-to-stock, one in which we hold safety stock and one in which we do not. Within the region where we hold safety stock, we can show that the fraction of total stock due to safety stock is smaller when demand variability is higher. We can also show that a sufficiently large increase in demand variability can lead to the elimination of safety stock.



<Figure 4> Increasing demand variability eliminates safety stocks (ρ=0.8, Υ_i=0.5)

5. Conclusion

Using a simple model of a multi-item production-inventory system, we explored the effect of demand variability on various system characteristics. Counter to immediate intuition, we found that higher demand variability leads to a lower coefficient of variation in lead-time demand. Since lead-time demand variability is what determines the amount of inventory buffering that is needed, we found that the fraction of safety stock due to total stock decreases with demand variability, and in some cases reduces to zero.

Although an intuitive explanation for some of these results is not easy, they appear to be related to the endogenous nature of lead-time. Both demand variability and capacity loading affect congestion in the production system and therefore affect supply lead-time, which in turn affects lead-time demand. Although both the mean and variance of lead-time demand increase with higher demand variability and higher loading, their ratio decreases in these same parameters. This appears to play the crucial role in many of the observed effects.

References

[1] Benjaafar, S., W. L. Cooper and J. S. Kim (2005), Benefits of Pooling in Production-Inventory Systems, *Management Science*, 51(4), 548-565

[2] Buzacott, J. A. and J. G. Shanthikumar(1993), *Stochastic Models of Manufacturing Systems, Prentice-Hall*, Englewood Cliffs, N.

[3] Gerchak, Y. and Q. M. He(2003), On the Relation between the Benefits of Risk Pooling and the Variability of Demand, *IIE Transactions*, 35(11), 1027-1031.

[4] Gerchak, Y. and D. Mossman(1992), On the Effect of Demand Randomness in Inventories and Costs, *Operations Research*, 40, 804-807.

[5] Jemai, Z. and F. Karaesmen(2005), The Influence of Demand Variability on the Performance of a Make-To-Stock Queue, *European Journal of Operational Research*, 164, 195-205.

[6] Shaked, M. and J. G. Shanthikumar(1994), *Stochastic Orders and Their Applications*, Academic Press, London.

[7] Song, J. S.(1994a), Understanding the Lead- Time Effects in Stochastic Inventory Systems with Discounted Costs, *Operations Research Letters*, 15, 85-93.

[8] Song, J.S.(1994b), The Effect of Lead-Time Uncertainty in a Simple Stochastic Inventory Model, *Management Science*, 40, 603-613.

[9] Stoyan, D.(1983), *Comparison Methods for Queues and Other Stochastic Models*, John Wiley, Berlin.

[10] Wolff, R. W.(1989), *Stochastic Modelling and the Theory of Queues*, Prentice-Hall, New Jersey.

[11] Zipkin, P. H.(2000), *Foundation of Inventory Management*, McGraw-Hill, New York.

〈Appendix〉

Proof of Lemma 1. The conditional probability $\Pr(N_i = n_i | N = n)$, where N_i is the number of units of item i that are on order and $N = N_1 + N_2 + \ldots + N_K$ has a binomial distribution with probability p_i. Hence, we have:

$$\Pr(N_i = n_i \mid N = n) = \frac{n!}{n_i!(n-n_i)!}(p_i)^{n_i}(1-p_i)^{n-n_i} \qquad \forall n \geq n_i$$

Noting that the distribution of the total numbers of customers in a GI/M/1 queue is given by (Wolff, 1989):

$$\Pr(N = n) = \begin{cases} 1-\rho & n = 0 \\ \rho(1-\beta)\beta^{n-1}, & n = 1, \ldots, \infty \end{cases}$$

where β is the solution of $\beta = \int_0^\infty e^{-\mu t(1-\beta)} dG(t)$, we obtain

$$\Pr(N_i = n_i) = \sum_{n=n_i}^{\infty} \Pr(N_i = n_i \mid N = n)\Pr(N = n)$$
$$= \sum_{n=n_i}^{\infty} \frac{n!}{n_i!(n-n_i)!} p_i^{n_i}(1-p_i)^{n-n_i}\rho(1-\beta)\beta^{n-1}, \quad \text{for } n_i \geq 1,$$

which can be rewritten as

$$\Pr(N_i = n_i) = \frac{\rho(1-\beta)}{\beta}\left(\frac{p_i}{1-p_i}\right)^{n_i} \sum_{n=n_i}^{\infty} \frac{n!}{n_i!(n-n_i)!}(\beta(1-p_i))^n.$$

Using the fact that

$$\sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} a^n = \frac{a^k}{(1-a)^{k+1}}$$

leads to

$$\Pr(N_i = n_i) = \frac{\rho(1-\beta)}{\beta}\left(\frac{p_i}{1-p_i}\right)^{n_i} \frac{(\beta(1-p_i))^{n_i}}{(1-\beta(1-p_i))^{n_i+1}}$$

or equivalently as:

$$\Pr(N_i = n_i) = \frac{\rho}{\beta}\left(\frac{1-\beta}{1-\beta(1-p_i)}\right)\left(\frac{\beta p_i}{1-\beta(1-p_i)}\right)^{n_i}.$$

Letting

$$r_i = \frac{\beta p_i}{1-\beta(1-p_i)}$$

leads to

$$\Pr(N_i = n_i) = \frac{\rho}{\beta}(1-r_i)r_i^{n_i}. \qquad \text{(A.1)}$$

For $N_i = 0$, we have

$$\Pr(N_i = 0) = 1 - \sum_{n_i=1}^{\infty} \Pr(N_i = n_i) = 1 - \left(\frac{\rho}{\beta}\right) r_i. \qquad \text{(A.2)}$$

Using (8) and (9), we can now obtain the expected inventory for each item as

$$E(I_i) = \sum_{n_i=0}^{s_i} (s_i - n_i)\Pr(N_i = n_i) = s_i - \frac{\rho}{\beta}\left(\frac{r_i}{1-r_i}\right)(1-r_i^{s_i}) \qquad \text{(A.3)}$$

and the expected number of backorders as

$$E(B_i) = \sum_{n_i=s_i}^{\infty} (n_i - s_i)\Pr(N_i = n_i) = \frac{\rho}{\beta}\left(\frac{r_i^{s_i+1}}{1-r_i}\right). \qquad \text{(A.4)}$$

The expected inventory cost can thus be rewritten as:

$$z = \sum_{i=1}^{N} h_i\left[s_i - \frac{\rho}{\beta}\left(\frac{r_i}{1-r_i}\right)(1-r_i^{s_i})\right] + b_i\left[\frac{\rho}{\beta}\left(\frac{r_i^{s_i+1}}{1-r_i}\right)\right], \qquad \text{(A.5)}$$

which is clearly convex and separable in the s_i's. Hence, the optimal base stock level for each item i can be obtained by finding s_i that solves $z(s_i) - z(s_i - 1) = 0$. This yields

$$s_i = \tilde{s}_i = \frac{\ln[\gamma_i \beta / \rho]}{\ln[r_i]}.$$

Because the base stock level must be an integer, we have $s_i^* = \lfloor \tilde{s}_i \rfloor$.

In the case where a service level is enforced, we minimize $z = \sum_{i=1}^{N} h_i E(I_i)$ subject to $\Pr(\tilde{N}_i \geq s_i) \leq \alpha_i$ for all $i = 1, \ldots, K$. Noting that the distribution of number of customers in the queue at an arrival instant is given by:

$$p(\tilde{n}) = (1-\beta)\beta^{\tilde{n}} \quad \text{for } \tilde{n} = 1, \ldots, \infty, \qquad \text{(A.6)}$$

it is not difficult to show that

$$\Pr(\tilde{N}_i \geq s_i) = 1 - \sum_{n_i=0}^{s_i-1} (1-r_i)r_i^{n_i} = r_i^{s_i}. \qquad \text{(A.7)}$$

Since z is strictly increasing in the s_i's, the constraints are always binding. Therefore, the optimal base-stock levels can be found by solving the equality

$$r_i^{s_i} = \alpha_i \quad \text{for } i=1,\dots,K.$$

This leads to

$$s_i = \breve{s}_i = \frac{\ln(\alpha_i)}{\ln(r_i)}.$$

Since the base stock level must be an integer that meets the service level constraint, we have $s_i^* = \lceil \breve{s}_i \rceil$.

Proof of Lemma 4. From Lemma 1, $s_i^{*(j)}$ is the smallest integer that satisfies the inequality

$$\frac{\rho}{\beta_j}[r_i^{(j)}]^{s_i^{*(j)}} \le \gamma_i,$$

which can be rewritten as

$$\frac{\rho p_i}{1-\beta_j(1-p_i)}[r_i^{(j)}]^{s_i^{*(j)-1}} \le \gamma_i.$$

Note that both the quantities $\frac{\rho p_i}{1-\beta_j(1-p_i)}$ and $r_i^{(j)}$ are increasing in β_j. Since $X_1 \le_v X_2$ implies $\beta_1 \le \beta_2$, we have $s_i^{*(1)} \le s_i^{*(2)}$. Now note that $z_i^{*(j)}$ can be written as

$$z_i^{*(j)} = h_i s_i^{*(j)} + (h_i + b_i)\frac{\rho}{\beta_j}\left(\frac{r_i^{(j)}}{1-r_i^{(j)}}\right)\left([r_i^{(j)}]^{s_i^{*(j)}} - \frac{h_i}{h_i+b_i}\right).$$

Suppose $s_i^{*(2)}=0$, then we must also have $s_i^{*(1)}=0$. From A.5, we have

$$z_i^{*(2)} = b_i \frac{\rho}{\beta_2}\left(\frac{r_i^{(2)}}{1-r_i^{(2)}}\right) \ge b_i \frac{\rho}{\beta_1}\left(\frac{r_i^{(1)}}{1-r_i^{(1)}}\right) = z_i^{*(1)}.$$

On the other hand, if $s_i^{*(2)}>0$, then

$$z_i^{*(2)} - z_i^{*(1)} \ge (h_i + b_i)\frac{\rho}{\beta_1}\left(\frac{r_i^{(1)}}{1-r_i^{(1)}}\right)\left([r_i^{(2)}]^{s_i^{*(2)}} - [r_i^{(1)}]^{s_i^{*(2)}}\right) \ge 0,$$

which completes the proof.

Proof of Theorem 3. Rewriting ϕ_i as $\phi_i = 1 - E(D_i)/s_i^*$, and using Lemma 3, it is sufficient to show that the ratio $E(D_i)/s_i^*$ is increasing in β. From the proof of lemma 4, we know that $\tilde{s}_i$ is increasing in β. Since $s_i^* = \lfloor \tilde{s}_i \rfloor$, it is sufficient to show that $\tilde{\phi}_i \equiv E(D_i)/\tilde{s}_i$ is increasing in β. The ratio $E(D_i)/\tilde{s}_i$ can be written as follows:

$$\tilde{\phi}_i = \frac{r_i \ln(r_i)}{(1-r_i)\ln(\frac{\beta}{\rho}\gamma_i)} = \frac{r_i \ln(r_i)}{(1-r_i)[\ln(\gamma_i) + \ln(\beta) - \ln(\rho)]}.$$

Let

$$f(\beta) = r_i \text{ and } g(\beta) = [\ln\left[\frac{\beta}{\rho}\gamma_i\right]]^{-1}$$

Then

$$\tilde{\phi}_i = \frac{f(\beta)\ln[f(\beta)]}{1-f(\beta)}g(\beta).$$

Differentiating with respect to β, we obtain:

$$\frac{d\tilde{\phi}_i}{d\beta} = \frac{[1-f(\beta)+\ln[f(\beta)]]f'(\beta)}{(1-f(\beta))^2}g(\beta) + \frac{f(\beta)\ln[f(\beta)]}{1-f(\beta)}g'(\beta)$$

Since $0 \le f(\beta) < 1$, we have $1-f(\beta)+\ln[f(\beta)]<0$. Since $g(\beta)<0$, the first term on the left hand-side of the inequality is positive. Since $\frac{f(\beta)\ln[f(\beta)]}{1-(\beta)} < 0$ and $g'(\beta) = \frac{-1}{\beta[\ln[\beta r_i/\rho]]^2} < 0$, the second term is also strictly positive. Therefore, $\frac{d\tilde{\phi}_i}{d\beta} > 0$, which completes the proof.