Manufacturing & Service Operations Management

On the Benefits of Inventory-Pooling in Production-Inventory Systems

Joon-Seok Kim, Saif Benjaafar,

To cite this article:
Joon-Seok Kim, Saif Benjaafar, (2002) On the Benefits of Inventory-Pooling in Production-Inventory Systems. Manufacturing & Service Operations Management 4(1):12-16. http://dx.doi.org/10.1287/msom.4.1.12.284

On the Benefits of Inventory-Pooling in Production-Inventory Systems

Joon-Seok Kim
Department of Mechanical Engineering, University of Minnesota, Minneapolis, Minnesota 55455
kimxx074@umn.edu

Faculty Mentor: Saif Benjaafar
Department of Mechanical Engineering, University of Minnesota, Minneapolis, Minnesota 55455

Inventory pooling is the consolidation of multiple inventory locations into a single one. Inventory locations may be associated with different geographical sites, different products, or different customers. The need to assess the benefits of inventory pooling arises in a variety of contexts. For example, manufacturing firms need to determine whether to have few large centralized distribution warehouses or several smaller local ones. In doing so, they must trade off the inventory economies realized from a centralized warehouse against the value of proximity to customers that comes from having several local ones (Eppen 1979, Schwarz 1989). Similarly, firms need to determine how much variety and differentiation to offer in their product portfolios. Offering few standardized products allows a firm to pool its finished goods inventories and minimize its risk exposure when demand is variable. On the other hand, increasing variety lets it offer more customized products, which could lead to greater market share or higher prices (Desai et al. 2001).

Firms that choose to offer high product variety must determine whether it is worthwhile to redesign their products and processes so that the product customization steps are delayed as much as possible. Delayed product differentiation is another form of inventory pooling. Here firms must balance the pooling benefits of postponed customization against any associated product and process redesign costs (Lee and Tang 1997). The need to assess the value of pooling also arises when a firm must decide how much component commonality its products should have. An increase in the number of shared components among products leads to the consolidation of component inventories and to scale economies in purchasing and manufacturing (Thonemann and Brandeau 2000, Baker et al. 1986, Gerchak et al. 1988). However, because a shared component is usually more expensive than some of the specialized ones it replaces, the inventory benefits of component sharing must be traded off against higher unit costs. Inventory pooling also arises when considering transshipments between different retail locations (Axsäter 1990, Robinson 1990). Lately, the economies from inventory pooling have been used to explain the competitive advantage of online retailing over traditional retailing (Eisenmann and Brown 2000).

In this paper, we examine the benefits of inventory pooling in systems where supply leadtimes are endogenously generated by a production system with finite capacity. We analyze the cost advantage derived from consolidating inventory from multiple locations into a single one. We show that this cost advantage is sensitive to the utilization of the production system, with the relative advantage of pooling disappearing when utilization is very high. Surprisingly, we find that the effect of utilization is not monotonic. There is a utilization level (determined by the holding costs, backordering costs, and number of inventoried items) for which the value of pooling is maximum. In addition to loading, we show that the cost advantage of pooling is affected by the variability in demand and in production times. Counter to intuition, we show that the value of pooling is generally decreasing in either demand or production-time variability, with pooling having little relative advantage when variability is high. For systems with asymmetric costs, we show that pooling is not always beneficial and could indeed lead to higher total costs.

We use our results to highlight differences with systems, where leadtimes are exogenous and load independent. In particular, we bring attention to the important role congestion and variability play in determining the value of pooling. We draw several managerial insights that can be used to define guidelines as to when inventory pooling is particularly valuable. The models we present are generic enough to be useful to managers in evaluating important trade-offs that are associated with pooling and in understanding the role various parameters play in these trade-offs. The following is a brief description of the model and the main results. Additional details can be found in Benjaafar et al. (2001).

1. Model Description and Summary of Results

We consider a production-inventory system with N types of items, where each item is managed according to its own base-stock policy. Demand for each item occurs one unit at a time according to independent Poisson processes, and all stock-outs are backordered. Supply leadtimes are endogenous and determined by the production and queueing times at the facility. Whenever a unit of type i inventory is demanded, an order is triggered at the common production facility (for all inventory types). The production facility can process one order at a time, and production times are assumed to be i.i.d. exponentially distributed random variables. Orders are processed on a first-come, first-served basis, and orders that arrive at a busy production facility must wait in a queue. The overall demand process is Poisson, and each demand causes an order at the production facility; therefore, when viewed in isolation, the number of orders at the production facility forms an $M/M/1$ queue—we later discuss extensions to systems with general distributions for demand and production times. To ensure stability, we assume $\rho \equiv \lambda/\mu < 1$, where ρ is the utilization of the production facility, λ is the aggregate demand rate, and μ is the production rate.

For each system with multiple and distinct inventory locations, we consider an equivalent *pooled* system where the inventory locations are consolidated into a single one and from which all the demand streams are satisfied. Note that because the individual demand streams are Poisson, the aggregated demand stream in the pooled system is also Poisson. In practice, this kind of pooling arises when the "original" items correspond to warehouses that stock the same product; in the pooled system, the individual warehouses are replaced with a centralized warehouse from which all demand is served. This can also represent the situation where N different components used in the manufacturing of N different products are replaced by a single common component that can be used in any of the N original items. Alternatively, this may correspond to a system where the original items are products that are redesigned, so that differentiation is postponed until demand is realized and, once needed, takes relatively little time to carry out.

First, we examine the effect of pooling on the distribution of leadtime demand, where leadtime demand refers to the amount of demand that arrives between the time an order is placed and when that order arrives back at the inventory buffer. For a system with symmetric costs and demand rates (i.e., $\lambda_i = \lambda/N$), we show that if the items are pooled into a single location, then the coefficient of variation in leadtime demand is reduced by a factor of $\theta(\rho, N) \equiv \sqrt{N(1-\rho)+\rho}$. This factor is strictly decreasing in the utilization ρ of the production facility, with $\theta(\rho, N) \to \sqrt{N}$ when $\rho \to 0$, and $\theta(\rho, N) \to 1$ as $\rho \to 1$. It is of interest to note that these effects are absent from inventory systems where leadtimes are treated as exogenous and i.i.d. In that case, it can be shown that the factor by which leadtime demand is reduced is simply $\sqrt{N}$, independent of the leadtime distribution. This leads to the following observation.

Observation 1. The reduction in demand variability due to pooling is smaller in a production-inventory system than it is in an inventory system with exogenous leadtimes. The variability reduction is most significant when the production facility is lightly loaded and is insignificant when the loading of the facility is high.

These effects are due to the somewhat surprising result that variability of leadtime demand is actually decreasing in the utilization of the production facility regardless of whether we pool or not, with the coefficient of variation approaching 1 as $\rho \to 1$. These results suggest that the desirability of pooling is affected by the utilization of the production system, and that the benefit from pooling may be less significant when utilization is high. As discussed below, analysis of the optimal costs under distributed and pooled configurations confirms this intuition.

To examine the effect of number of items on cost, we assume that each unit of stock in an inventory buffer accrues cost at rate h per unit time and that each unit of backordered demand accrues cost at rate b per unit time. We show that for a symmetric system, the optimal cost is given by

$$\begin{aligned} TC^*(\rho, N) &\equiv TC(s^*, \rho, N) \\ &= N\left[h\left(s^* - \frac{r(1-r^{s^*})}{1-r}\right) + b\frac{r^{s^*+1}}{(1-r)}\right], \end{aligned}$$

where s^* is the optimal base-stock level for each item and is given by

$$s^* = s^*(\rho, N) = \lfloor \tilde{s} \rfloor \quad \text{where}$$

$$\tilde{s} = \tilde{s}(\rho, N) = \frac{\ln(h/(h+b))}{\ln(r)}$$

and $r = \rho/(N(1-\rho)+\rho)$. We show that $TC^*(\rho, 1) \le TC^*(\rho, N)$; hence, a pooled system is always cost superior to a distributed one.

Because of the integrality of s^*, direct analysis of the optimal cost function is difficult. However, several important observations can be made. First, we note that $TC^*(\rho, N)$ is not always increasing in N. In fact, for each ρ there is a value of N (which depends upon ρ) beyond which the optimal cost remains constant. This value of N, denoted $N_{\max}$, can be obtained by noting that $\tilde{s}$ is a decreasing function in N and is less than one when $N > b\rho/\{h(1-\rho)\}$. This yields $s^* = 0$ and $TC^*(\rho, N) = b\rho/(1-\rho)$ whenever $N \ge N_{\max} = 1 + b\rho/\{h(1-\rho)\}$. The value $N_{\max}$ marks the number of items beyond which it becomes optimal not to hold any inventory and to produce to order (note that this number is increasing in ρ).

For a fixed N, there is similarly a value of ρ, below which the optimal cost is not affected by N. The existence of this value, given by $\rho_{\min} = Nh/(Nh+b)$, follows from the fact that $s^* = 0$ when $\rho \le \rho_{\min}$. This means that for sufficiently small utilization, it is optimal not to hold any inventory and to simply produce to order. In this case, the value of $TC^*(\rho, N)$ is also given by $b\rho/(1-\rho)$. These insights are summarized in the following.

Observation 2. In a production-inventory system, there is a maximum value of N, beyond which we switch from producing to stock to producing to order. For $N \ge N_{\max}$, total cost is no longer affected by changes in N. Similarly, there is a value of ρ, below which we always produce to order. For $\rho < \rho_{\min}$, total cost is not affected by N.

If we ignore the integrality of s^*, then the approximate optimal cost has the simpler form:

$$TC(\tilde{s}, \rho, N) = hN\tilde{s} = hN\frac{\ln[h/(h+b)]}{\ln[\rho] - \ln[N(1-\rho)+\rho]}.$$

Although this is an upper bound on the optimal cost, it provides a reasonable approximation of the optimal cost when $N < N_{\max}$ (i.e., in the region where $TC^*(\rho, \cdot)$ is a function of N). In this region, total cost is increasing in N at a rate approximately proportional to $N/\ln(N)$.

To assess the relative advantage of pooling, we examine the ratio of the optimal total cost for the distributed system to the optimal total cost for the pooled one; $\delta(\rho, N) = TC^*(\rho, N)/TC^*(\rho, 1)$. Although this ratio is always greater than or equal to one, it is highly sensitive to ρ, the utilization of the production facility. Although it can be significant when ρ is in the midrange, it equals one when ρ is sufficiently small; i.e., $\rho \le h/(h+b)$. In fact, we show that $\lim_{\rho\to 1}\delta(\rho, N) = 1$, which means that the relative advantage of pooling disappears when utilization is high. This result should not be surprising given that the coefficient of variation in leadtime demand tends to one when ρ approaches one. However, it is surprising to see that the effect of ρ is not monotonic. While initial increases beyond $\rho_{\min}$ tend to increase the ratio $\delta(\rho, N)$, additional increases cause $\delta(N, \rho)$ to decrease and, eventually, to converge to one. Thus, there is a value of ρ, where $h/(h+b) \le \rho < 1$, for which the relative advantage of pooling is maximum.

Because of the integrality of s^* it is difficult to exactly characterize the value of ρ for which δ is maximum. However, we empirically observed that this maximum occurs when ρ is in the neighborhood of $hN/(hN + b)$. Of course, this is $\rho_{\min}$, the utilization below which the distributed system does not hold any inventory. An explanation of why this also defines the neighborhood where δ is maximum is as follows. When $h/(h + b) \leq \rho < hN/(hN + b)$, the pooled system is able to hold stock and, therefore, is able to counter increases in backordering costs as ρ increases. In contrast, the distributed system must continue to produce in a make-to-order fashion when ρ is in the same region. This leads the distributed system to incur increasingly higher backordering costs as ρ increases. These costs are at their highest when ρ approaches but does not exceed $\rho_{\min}$. For $\rho \geq \rho_{\min}$, both the distributed and pooled systems can hold stock and the pooled system loses some of its advantage. In fact, for $\rho \geq \rho_{\min}$, the variability effect takes over, and δ becomes mostly decreasing in ρ. We summarize our key result in the following observation.

Observation 3. In a production-inventory system, the relative advantage of pooling is determined by the utilization of the production facility. Although the benefits of pooling can be significant when utilization is moderate, they are insignificant when utilization is either very high or sufficiently small.

In addition to examining the effect of utilization, we also study the impact of demand and production-time variability on the relative advantage of pooling. To this effect, we relax the assumption of Poisson demand and exponential processing times. Intuitively, one might expect that as either demand or process variability increases, the advantage of pooling would also increase. We show that this is not true and that, in fact, an increase in variability could diminish the benefits of pooling. More significantly, we show that in the limit cases of very high variability in either demand or production time, there is no advantage to pooling at all. Furthermore, we find that the effect of variability on the distribution of leadtime demand is similar to that of utilization. In particular, we show that an increase in either demand or production-time variability generally leads to a decrease in leadtime demand variability. Consequently, the reduction in leadtime demand variability because of pooling is also decreasing in both demand and production-time variability. Additional discussion of the effect of variability can be found in (Benjaafar and Kim 2001).

Finally, we consider various extensions to our basic model, including systems with service-level constraints, joint capacity and inventory pooling, and asymmetric demand and cost parameters. It is particularly worth mentioning that in systems with asymmetric demand and asymmetric backordering costs, pooling is not always beneficial. In a distributed system, inventory is dedicated to each item, which allows us to set different base-stock levels for items, based on their backordering costs. This offers protection against running out of the item with the more expensive backordering cost while limiting the amount of inventory held for items with low costs. In a pooled system, inventory is shared between multiple products. To offer the same level of protection for the more expensive items, a significantly larger amount of inventory needs to be held. This effect is more pronounced when the demand for the more expensive items is significantly lower than that of the cheaper ones. Surprisingly, the maximum benefit from pooling is not realized when the products have perfectly symmetric costs and demands. This maximum appears to take place when there is still a small measure of asymmetry.

2. Concluding Comments

In this paper, we examine the benefits of inventory pooling in systems where supply leadtimes are endogenously generated by a production system with finite capacity. We show that the realized value from pooling depends on several system parameters, including system loading, variability, service levels, and backordering and holding costs. We identify conditions where pooling carries little or no advantage relative to a system with multiple items. This includes systems with high variability, high loading, or asymmetric costs. These results suggest that managers should take greater caution in evaluating the impact of pooling in systems where supply is driven by a capacitated production system. Managers also need to be aware of the important interaction between congestion in the production facility and the performance of the inventory system.

In this paper, we focus on quantifying inventory-related cost savings that might be realized from pooling. Clearly, these savings need to be traded off against possible cost increases in other areas. For example, maintaining a centralized warehouse could lead to increased transportation costs that would be necessary to maintain the same customer response time. If inventory consolidation is achieved via delayed product differentiation, a significant and costly redesign of the product or the process could be necessary. Similarly, if pooling is realized by standardizing product features and offering less product variety, there could be loss of market share or profits. On the other hand, there might be additional benefits from pooling in the form of lower overheads, simpler inventory control, and streamlined production processes. Managers need to be aware of these additional costs and benefits when deciding when to pool inventory and by how much.

References

Axsäter, S., 1990. Modeling emergency lateral transshipments in inventory systems. *Management Sci.* **36** 1329–1338.

Baker, K. R., M. J. Magazine, H. L. Nuttle. 1986. The effect of commonality on safety stock in a simple inventory model. *Management Sci.* **32** 982–988.

Benjaafar, S., J. S. Kim. 2001. When does higher demand variability lead to lower safety stocks. Working paper, Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN.

——, W. L. Cooper, J. S. Kim. 2001. On the benefits of inventory pooling in production. Working paper, Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN.

Desai, P., S. Kekre, S. Radhakrishnan, K. Srinivasan. 2001. Product differentiation and commonality in design: Balancing revenue and cost drivers. *Management Sci.* **47** 37–51.

Eisenmann, T., A. Brown. 2000. Online retailers. Technical report 9-801-306, Harvard Business School, Harvard University, Cambridge, MA.

Eppen, G. 1979. Effects of centralization on expected costs in multilocation newsboy problems. *Management Sci.* **25** 498–501.

Gerchak, Y., M. J. Magazine, A. B. Gamble. 1988. Component commonality with service level requirements. *Management Sci.* **34** 753–760.

Lee, H. L., C. S. Tang. 1997. Modeling the costs and benefits of delayed product differentiation. *Management Sci.* **43** 40–53.

Robinson, L. W. 1990. Optimal and approximate policies in multiperiod, multilocation inventory models with transshipments. *Oper. Res.* **38** 278–295.

Schwarz, L. 1989. A model for assessing the value of warehouse risk pooling over outside-supplier leadtime. *Management Sci.* **35** 828–842.

Thonemann, U. W., M. L. Brandeau. 2000. Optimal commonality in component design. *Oper. Res.* **48** 1–19.