Inventory reservation and nested allocation as tools to improve performance

Behzad Samii – Vlerick Leuven Gent Management School

Companies often wish to provide different levels of service to different classes of customers. Customer-differentiated service levels may be motivated by differences in the perceived customer lifetime value or by specific contractual agreements that include service level guarantees. One way to provide differentiated service levels is to reserve a certain portion of the available inventory exclusively for certain classes of customers.

In practice, companies commonly base inventory decisions on service level and fill rate measure targets.

Companies often wish to provide different levels of service to different classes of customers. Customer-differentiated service levels may be motivated by differences in the perceived customer lifetime value or by specific contractual agreements that include service level guarantees. One way to provide differentiated service levels is to reserve a certain portion of the available inventory exclusively for certain classes of customers. For example, while waiting for a planned stock replenishment, a major electronics manufacturer rations its limited stock of ADSL kits among several different telecom clients to ensure that the contractually agreed service levels for this high-priority class of customers can be satisfied.

Current approaches to inventory reservation are typically based on cost and revenue measures. An organisation assumes that it can assign specific revenue or penalty cost to a customer for any order or unit of demand fulfilled or unfulfilled. In practice, however, it is very difficult to accurately estimate the (especially the long-term) financial implications of not meeting customer demand and corresponding service level requirements on an individual order basis.

For instance, in the case of a pandemic disease, it is obvious that critical personnel who are needed in healthcare institutions to assist those who are sick should be given priority in obtaining the medications. On the other hand, the general population should also have sufficient access to the medications, and so some amount of the drugs should be made available to them, even if on a first-come-first-served basis. Consequently, public health officials have to develop protocols for rationing the available medications based on multiple, non-financial competing objectives.

In practice, companies commonly base inventory decisions on service level and fill rate measure targets. However, if a decision-maker chooses to reserve inventory to ensure an agreed service level for certain customer classes, he or she can be confronted with a conflict: if the product is perishable, reserving inventory to ensure the fulfilment of higher-priority

Insert 1

We can split the fill rate of high-priority demand into its two building blocks: the ‘reservation effect’ and the ‘nesting effect’. The reservation effect represents all instances in which the fulfilled high-priority demand consumes only from the reserved portion of the inventory. Nesting can affect the high-priority demand fill rate only after the reserved inventory is fully consumed. Consequently, the nesting effect includes instances where competing with low-priority demand for the unreserved stock on a first-come-first-served basis leads to the fulfilment of high-priority demand. Figure 1 illustrates how the sum of the reservation effect and the nesting effect determines the expected fill rate of high-priority demand for any given quantity of reserved inventory.

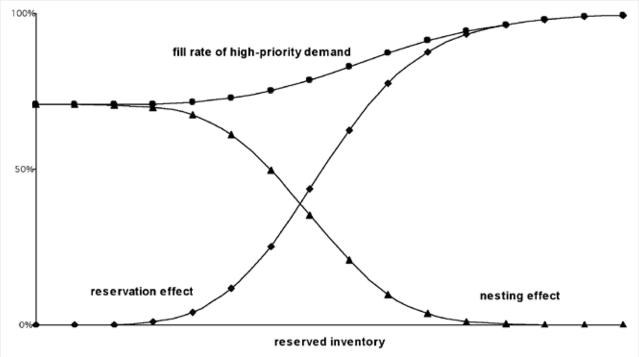

Figure 1. Fill rate of high-priority demand as the sum of the reservation and nesting effects

If no inventory is reserved, the reservation effect remains at zero, and the expected fill rate of high-priority demand is determined entirely by the nesting effect. As more and more inventory is reserved for high-priority demand, the reservation effect increases while the nesting effect decreases simultaneously. In other words, when the reserved inventory has been consumed by arriving high-priority demand (or spoils waiting for this demand), less unreserved inventory remains available for possible equal competition between future high-priority and low-priority demands. However, the fill rate of high-priority demand (as the sum of the two effects) continues to increase with the increased reserved inventory.

At the extreme of full inventory reservation, no inventory remains available for equal competition between high- and low-priority demands. Hence, the nesting effect is zero, and the resulting expected fill rate of high-priority demand is determined entirely by the reservation effect. This represents the highest achievable fill rate of high-priority demand.

demand may have a detrimental impact on the performance of the overall system due to spoilage. That is, overall system performance is negatively impacted when reserving inventory for prioritised customers leaves too little stock to fulfil the demand of lower-priority customers; and, furthermore, if a portion of the reserved inventory remains unused, it spoils and is lost. As a consequence, the decision-maker must evaluate the non-linear trade-off between the benefits of ensuring service levels for prioritised customer classes and the potential decrease in overall performance by limiting unreserved stock quantities.

Optimising the trade-off

Our research attempts to lend insight into this important trade-off. We mathematically model a single period inventory reservation problem with two classes of customers – high-priority and low-priority. At the beginning of the planning horizon, the decision-maker decides how much inventory to reserve exclusively for the high-priority customers. Inventory stock will not be replenished during the planning horizon. We use two non-financial performance measures to analyse the decision-maker's trade-off when determining the reserved quantity:

- The (expected) fill rate of the high-priority demand
- The (expected) decrease in the overall system fill rate due to inventory spoilage

In our model, *nesting* refers to the allocation policy in which demand from high-priority customers competes equally with demand from low-priority customers for any remaining unreserved inventory once (and if) the high-priority demand has exhausted its reserved inventory. To capture the 'reservation effect' and the 'nesting effect' separately (***see insert 1***), we have to consider the demand arrival processes for both customer classes over time.

Conclusion

We can develop exact expressions that enable us to compute, for any given reserved quantity, the fill rate of the high-priority demand and the corresponding decrease in the fill rate of the overall system. This can be used to support a decision-maker in choosing her 'optimal' reserved inventory quantity. In other words, the decision-maker can assess whether or not it is prudent to 'pay the price' of losing some overall system performance to achieve a certain high-priority demand fill rate. For example, public health planners can now calculate achievable immunization rates for all target population groups *prior* to reserving a portion of the available vaccines for the exclusive use of higher-priority medical professionals.

The same information provides valuable support to a decision-maker, not only when setting specific reserved quantities, but also prior to offering performance-based agreements to a certain class of customers. For instance, based on total available inventory, the obtainable high-priority demand fill rates of all telecom clients can be determined *prior* to committing a portion of the available ADSL kits for these more important customers (***see insert 2***). This enables the decision-maker to offer sustainable fill rate-based contracts to the telecom clients.

Insert 2

Increasing the reserved quantity for the exclusive use of high-priority demand has a positive impact on the fill rate of high-priority demand and a disproportionately negative impact on the fill rate of overall demand. Figure 2 shows the expected fill rate of high-priority demand and the corresponding (expected) decrease in the fill rate of overall demand depending on the quantity of reserved inventory. At low reservation levels, the chance that the reserved inventory will spoil (before it is consumed by high-priority demand) is low, leading to a negligible decrease in the fill rate of overall demand. At higher levels of reservation, the chance that some of the reserved inventory will spoil rises correspondingly, resulting in a proportionately large decrease in the fill rate of overall demand.

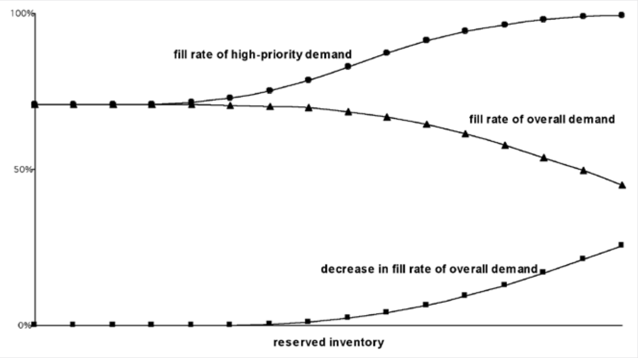

Figure 2. The decrease in the fill rate of overall demand rises with higher reserved inventory

To ensure a high fill rate of high-priority demand, the decision-maker faces a significant decrease in the fill rate of overall demand. Not surprisingly, the best obtainable fill rate of high-priority demand corresponds to the greatest decrease in the fill rate of overall demand.

This clearly illustrates the multiple competing objectives the decision-maker faces. Given her preferences, the decision-maker has to determine the optimal quantity of reserved inventory based on the illustrated trade-off.

We also provide indications for how the set of potential reserved quantities can be restricted when partial information about the decision-maker's preferences is available (e.g. if a health planner requires a minimum vaccination coverage rate for all target groups). Because we are able to formally characterise the trade-off between the two key performance measures, our results can also be used to support the decision-maker in determining such reserved quantity restrictions based on the partial preference information. Simple decision rules corresponding to these inventory threshold values can be used to pre-configure an automated inventory reservation engine.

In summary, our analysis generates a number of insights for managers regarding the impact of inventory reservation on overall system performance using a nesting allocation policy under various system parameters. Although we limit our analysis to a single period inventory reservation problem, we expect that our results also apply to a wide range of supply chains in which a decision-maker has to ration a perishable resource among different classes of customers.

For more information

MIT ZARAGOZA
International Logistics Program

This paper synthesises some of the results of Behzad Samii's doctoral dissertation, written under the supervision of Professor Richard Pibernik in the MIT-Zaragoza International Logistics Program. For more information, please contact Professor Behzad Samii: (behzad.samii@vlerick.com).

Research carried out by the Vlerick Centre for Supply Chain Excellence

Vlerick Leuven Gent Management School
The Autonomous Management School of Ghent University and Katholieke Universiteit Leuven

Goodman

Goodman is the leading Partner in the Vlerick Centre for Supply Chain Excellence through the Goodman Chair in Logistics Management

April 2010