

## Make-to-Stock under Drum-Buffer-Rope and Buffer Management Methodology

### WHY MAKE-TO-STOCK?

At least from the theory of constraints (TOC) perspective this is a valid question. The traditional drum-buffer-rope (DBR) planning approach emphasizes make-to-order. There is something sensible about producing only to order: there is no waste of materials or capacity in making things that will not sell. As market demand is always a major constraint, make-to-order is a straightforward subordination to the market: make what is needed and refrain from making what is not needed.

There is only one reason that will make it necessary to consider producing for stock and that is when the time it takes to produce an order is longer than the time customers are willing to wait. In this situation there is a need to start production before the customer asks for the product. A reduction in lead time is one of the targets of several methodologies, including lean and TOC. But lead time reduction may be not enough, especially when the customer does not want to wait at all. Same day shipment can be described as “off the shelf” for all practical purposes, and in such a case there is a need to have the products in stock, even in a very fast response manufacturing environment.

Supply chains make the need for make-to-stock even more universal. Even if every link in the chain has succeeded in drastically shrinking the lead time, the overall lead time throughout the chain coupled with the fast-changing demands would make it imperative to hold some stocks between the links.

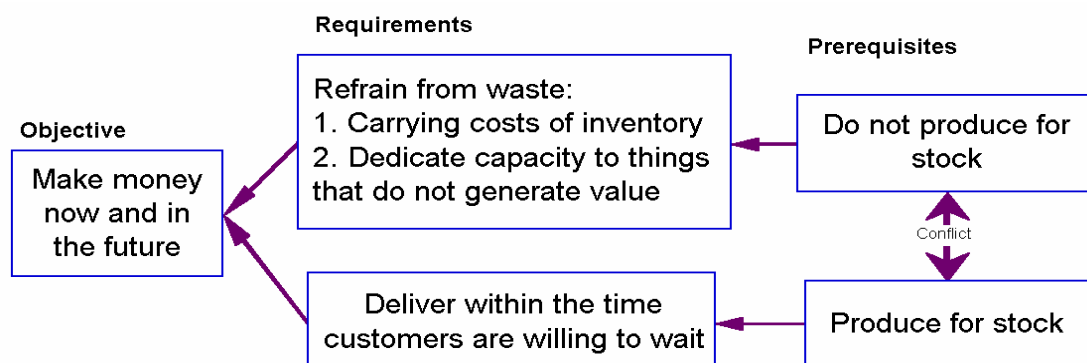
The objective of finished goods stock is to protect sales, but maintaining too much stock causes two types of waste. The first is the direct

financial cost of carrying stock, including the possibility of goods being scrapped or dumped. The second is the dedication of capacity that does not generate value, and even worse, this captured capacity results in goods required by the market not being produced. TOC and JIT/lean see the second waste as the more severe of the two. The lost capacity does not only mean waste of costs, but also waste of opportunity.

So there is a generic conflict here, make-to-stock or not make-to-stock. TOC presents such a problem as in **Figure 1**.

Each prerequisite is necessary in order to achieve the requirement it points to, and each requirement is necessary to achieve the objective. These two prerequisites are in direct conflict. So, on the one hand we certainly need, in certain situations, to make-to-stock in order to deliver in acceptable response time to our customers, and satisfying the customers (by delivering in acceptable time) is absolutely necessary in order to make money. On the other hand, maintaining stocks of finished goods generates significant waste that is counter to the objective. Hence, we should not produce stocks.

The direction of the solution in such a case is to challenge the necessity of one of the prerequisites. In this particular case, if we succeed in holding a very low stock of finished goods, then the waste practically vanishes. Carrying costs won't be more than normal in a make-to-order environment, and practically everything that is produced is going to be sold pretty soon. The emphasis is on very low stocks! That means that within a short time all the stock will have gone. In order to continue to satisfy the customer the replenishment of those low stocks needs to be VERY FAST.



**Figure 1. A Generic Conflict**

## **FAST REPLENISHMENT AND THE CLASH WITH CURRENT PRACTICES**

The replenishment time is very significant in evaluating the stock levels. Two common practices directly clash with the idea of shrinking the stock replenishment time to the extent that make-to-stock won't generate the two types of waste discussed above.

The first practice that clashes with fast replenishment is the norm of batching. TOC and JIT/lean are in total agreement regarding the huge damage caused by batching. Figure 1 shows two necessary requirements for achieving the common objective. Note that *both* requirements are tempered by the use of batches. The above requirement is straightforward, but through the use of batches more inventory is produced than is considered absolutely necessary. The rest is pure waste in terms of carrying both costs and capacity. The end requirement is the expectation of immediate and reliable delivery. The use of batches weakens the ability to respond quickly. Hence, both requirements are at odds with the objective of making money.

TOC, however, recognizes that in some cases batching cannot be avoided. Thus, there is a need to clearly define when batching is absolutely required and to ascertain the minimum size of a batch that will overcome a particular obstacle.

A case in point can occur when very small batches might turn a specific non-constraint resource into a bottleneck, due to the accumulation of setup time. It is important to verbalize it that way in order to come up with a clear idea of what is the minimum batch that will not turn the specific resource into a bottleneck. Certainly we do not need to increase batches just in order to improve "efficiencies." The only place where efficiency is important is at the capacity-constraint resource (CCR).

Stock levels are frequently managed according to min-max levels. The batching policy is hidden under the size of the max minus the min. The stock that should clearly protect sales is the "min" or order point. Anything above that is surplus inventory that generates the above two types of waste.

The second clash is the use of forecasts. Any production for stock involves some sort of forecasting. The "order point" level is a crude forecast regarding possible sales during the replenishment time. However, a sophisticated make-to-forecast instead of make-to-stock can result in a waste of effort and could turn out to be disastrous. Eventually even with all the sophistication and high stock levels, shortages occur regularly.

Forecasting tries to answer the following question: How many units shall we sell during the next period? However, the answer given by any type of forecasting is not very reliable. That

is not to say that without formal or computerized forecasting the answer will be more accurate; nevertheless, we're well aware that we don't really know the answer and thus prepare for different scenarios. When we realize we don't know, we try not to look too far ahead and keep to a reasonably short lead time. It seems that in forecasting some organizations try to look too far ahead, creating huge batches, and then chase the market trends using frequent reforecasting to correct the actions taken before the last forecast. One of the characteristics of any forecast is that its quality deteriorates pretty fast from the horizon. If the forecast for next week can be pretty good, the forecast for next month is much more exposed to variance and the forecast for next year is probably meaningless.

An important point is that what really counts most of the time is not the average demand, but how much it might be. Hence, the concept of safety stock comes to mind. In many cases the "noise" in the actual sales is the same as, or even more, than the official forecast, which tried to predict the average demand.

Maintaining stocks for protecting sales needs to consider not only the fluctuations in demand, but also the fluctuations in replenishment time. After all, the protection is active when a request for replenishment is issued. Until then the sales depend upon the availability of stock. So, the safety stock serves not only to cover deviations from the forecast, but also the deviations from the lead time/replenishment time.

## **FAST REPLENISHMENT IN THE TOC WAY**

The TOC approach to make-to-stock might be considered similar to lean/JIT in several parameters. But, the TOC control part (buffer management) is quite different.

Instead of trying to be very precise in very uncertain situations (like using sophisticated forecasting techniques), TOC strives to build a robust design that is good enough. The initial parameters are based on crude forecasting that is complemented by crude assessment of the variability of both market demand and replenishment time. What complements good-enough planning is a very flexible and priority-driven execution control system that is capable of taking care of the exceptions.

The key parameter is the replenishment level that has to be defined per product. This is the equivalent of the shipping buffer used in the DBR methodology for make-to-order. The behavior of the replenishment buffer is similar to the min-max method, but the min equals the max. Hence, every day when some units of the product are sold, a work order for the same amount is generated for the next day. The default thinking is that no minimum batch is used. As we'll see later, the system regulates itself, so that when the load is

severe, perhaps because of too many setups, then larger batches would naturally be released. But, TOC is not an extremist approach. In some cases there should be a need to dictate a minimum batch, otherwise the load on the CCR would be too large. So, minimum batch is an option, but it does not “hide” behind the difference between two parameters.

The size of the replenishment level should cover the average demand plus the safety stock to properly protect possible peaks of demand and possible delays of replenishment orders. If we like to maintain very low stocks, the replenishment time should be as short as possible.

The regular DBR approach is able to maintain very fast replenishment, based on the load and capacity of the CCR—the weakest link in the production chain. In establishing very short lead times, the role of the shop-floor control becomes extremely important.

### **BUFFER MANAGEMENT IN A MAKE-TO-STOCK ENVIRONMENT**

Buffer management is a control mechanism. The idea behind it can be summarized as identifying situations where the planned protection is almost exhausted. Once such a local case is identified, a warning is issued, resulting in high priority being given to the problematic orders and then utilizing the rest of the protection to remedy the local disruption.

In the make-to-order DBR, the buffers are all time-buffers. Exhausting the buffers means coming too close to the critical date. In the make-to-stock environment, the main buffer is real finished goods stock. By almost exhausting the protection means, in this case, very low on-hand stock—low to the extent that it could be exhausted before any reinforcement would arrive. This emergency level, sometimes called the “red-line level” or “zone 1,” represents an exception that should trigger action. When the on-hand stock goes down below the emergency level—efforts must be made to expedite completion of the work-order for that item as soon as possible.

The default for the emergency level is one-third (33 percent) of the replenishment level. The emergency level size should conform to two criteria:

1. The remaining stock still leaves time to expedite the work-in-process and reach the finished-goods stock on time. If this is not the case the emergency level should be higher, which might have a similar impact on the replenishment level as well.
2. Crossing the emergency level is neither too frequent nor too rare. If this is not true the replenishment level should be changed.

Our experience shows that 33 percent is a good initial guess, and the two criteria above should help to “tune” the relationship between the replenishment and the emergency levels.

The buffer status of a product is the ratio between the quantity missing to the replenishment level divided by the replenishment level. A buffer status of 70 percent means that the on-hand stock is 30 percent of the replenishment level. Considering the default value of the emergency level and the buffer status of 70 percent, the emergency level has been penetrated. The buffer status for each product dictates the priorities on the shop floor. It seems common sense that when two replenishment orders show up at a work center, one belonging to a product with a buffer status of 36 percent and the other a buffer status of 50 percent, then it is clear that the latter has a higher priority.

### **BUFFER MANAGEMENT AS FEEDBACK TO THE PLANNING PARAMETERS**

As already mentioned, buffer management identifies exceptions where the protection is almost exhausted. This identification serves to keep the planning intact by setting the priorities and expediting them when needed. But the occurrence of an exception might have a longer-term impact as well. The number and intensity of occurrences in a period of time testifies to the effectiveness of the replenishment and emergency levels.

Suppose product P10 had been below the emergency level for 10 days (“in the red”) last month. Assuming 22 working days in a month, then 10 out of 22 being in emergency is quite substantial. The reaction in such a case should be to increase the buffer (the replenishment level). The buffer should not be increased by a small percentage that is well within the “noise” (the regular fluctuation level) in the system; rather it should be something like a 25 percent increase.

An opposite situation also holds true. Suppose product P15 did not penetrate even once the “red-line” (the emergency level), and also did not cross the 2/3 line of the replenishment buffer—this would enable a reduction of the buffer by 25 percent.

This kind of analysis is a crude forecast, but it takes into account all the parameters that impact the stock: the average market demand, the average replenishment time, and the level of fluctuations of both. Hence, it is the ultimate sensor for the validity of the planning parameters.

### **THE REPLENISHMENT ORDER RELEASE PROCEDURE UNDER DBR METHODOLOGY**

Every day a check must be made for all products that are make-to-stock. When the on-hand stock is lower than the replenishment level, a further check is needed to verify how much work-in-process (WIP) exists for that product. If the summation of the on-hand plus the WIP is still lower than the replenishment level, then a new replenishment order has to be released. However,

before we do that, the total load on the capacity-constraint resource (CCR), needs to be checked. Each replenishment order does have a priority for being released at that day. The priority is, the replenishment order size divided by the replenishment level. This might be different than the buffer status for the product because the buffer status takes into account only the on-hand finished goods stock, while the priority for the new replenishment orders takes the WIP for the product into account.

Under DBR methodology the CCR is protected by the CCR time-buffer. Every day replenishment orders are launched up to the limit of the CCR time-buffer. When the CCR is loaded more than that, then the scheduling algorithm would delay the material release until CCR-time-buffer before the scheduled start of the CCR. As more sales are done every day, the replenishment orders that were not released on a particular day, due to relatively low priority and high load on the CCR, would be larger the next day (more to replenish due to additional sales) with higher priority to be released.

This procedure self-adjusts the average batch according to the load on the CCR. In off-peak periods the load on the CCR is not high, hence most replenishment orders would be released on the first day, meaning relatively small batches. When the load on the CCR is high, the replenishment orders would compete on being released—hence the average batch would grow.

Thus the system ensures self-adjusting to the load. Both the replenishment levels and the average load grow when the load grows, and go down when the load goes down. Thus the system is stable, as long as the CCR has still enough capacity to deal with the market demand.

### **FUNCTIONING UNDER MIX-MODEL—MAKE-TO-ORDER AND MAKE-TO-STOCK**

In many environments it makes sense to use a mix-model of make-to-stock and make-to-order. The more common products are better handled as make-to-stock, thus providing immediate delivery while keeping low-enough stocks that will soon be sold. Less common products, sometimes called “slow products,” while important to the market, are more problematic to handle as make-to-stock. In many cases customers for the slow products are ready to wait a short time for delivery. With a short time response, these products can be handled as make-to-order.

How does the above procedure match both types of orders? The DBR scheduling algorithm can load the CCR according to the due dates and the shipping-time-buffer as dictated by the due dates for the slow product, then complement the CCR schedule with the replenishment orders that do not have any specific due date.

Replenishment orders should NOT be given due dates. The material release is done according to the above procedure. Once the materials have been released, the urgency of a replenishment order does not depend on the time that has passed but on the product buffer-status. If, after the release of materials, sales for that product are high, then that order should become *very* urgent. If sales are lower than average, then there is no pressure.

The author considers that it is a mistake to attach an artificial due date for a replenishment order. It would be stupid to expedite such an order if the finished goods stock is not low. And it would be stupid *not* to push the replenishment order when the due date is not close, *but* the stock level is dangerously low.

Buffer management maintains the right priorities for both make-to-order and make-to-stock. The buffer status of a customer order (make-to-order) is based on how much time passed since material release divided by the time-buffer size. For replenishment orders, the buffer status depends on the missing quantity for replenishment level divided by the replenishment level. Even though the calculation of the buffer status is different, they are fully comparable.

An example:

Suppose on March 1, three orders are at the CCR:

- A large customer order (make-to-order) for 900 units of P3, due at March 28. The shipping buffer is 2 weeks.
- Replenishment order (make-to-stock) for 500 units of product P1. The replenishment buffer is 1,000 units. On-hand stock is 430 units.
- Replenishment order (make-to-stock) for 100 units of P2. The replenishment buffer is 300 units. On-hand stock is 106 units.

The emergency zones for both make-to-stock and make-to-order are 30 percent of the buffer, or 70 percent penetration into the buffer.

What should the CCR work on first?

Certainly *not* the customer order for P3. That order has four weeks until its due date, and the shipping buffer is only two weeks. The shipping buffer status, which is equal to the shipping buffer minus the time left until the due date, then divided by the shipping buffer, is -100 percent penetration.

How could an order be in negative penetration into the buffer?

Several possible explanations:

- Someone has released the order too early—against the DBR schedule.
- Noting that we look here only at the shipping buffer, while that order is still behind the CCR. This opens the way to a further explanation of the amount of negative shipping buffer:
  - The CCR buffer is relatively long, but that order reached the CCR early.

- The original CCR schedule has placed the order very early due to scheduling considerations on the CCR.

The buffer status for P1:  $(1000-430=570)$  units to be replenished. The ratio to the replenishment level is  $570/1000 = 0.57$ , meaning 57 percent penetration.

The buffer status for P2:  $(300-106)/300 = 0.65$ , meaning 65 percent penetration into the buffer.

If we have the flexibility to reassess the schedule on the CCR, then the current priorities should be P2 first, then P1, then P3.

Suppose now there were more orders at the CCR and somehow the same three orders remained a further 10 days at the CCR. What should be the priorities then?

We don't need more information about the customer order for P3. Its buffer status is still negative.

We should need updated information about the finished goods stock for P1 and P2. Suppose we find out that 200 units of P1 are left, and 102 units of P2.

The buffer status for P1 is  $(1000-200)/1000 = 0.8$ , or 80 percent.

The buffer status for P2 is  $(300-102)/300 = 0.66$ , or 66 percent.

So, the priorities have now changed due to an alteration in the market demand for both products. Now P1 is the most urgent. As a matter of fact it crossed the red-line level and is in an emergency situation. Hence, it may be necessary to take urgent measures.

Note the difference between make-to-stock order, whose priorities lie with the state of the finished goods stock, and make-to-order for which each order is evaluated according to its own due date.

## CONCLUSIONS

I hope one message has been clarified through the discussion about make-to-stock under DBR and buffer management. Make-to-stock implies a somewhat different approach than make-to-order, because the priorities are constantly changing based on the immediate market demand and the finished goods stock.

Converting replenishment orders to pseudo-customer orders by sticking an artificial due date to them is, to my mind, wrong. At most, it reflects the right priorities when the planning is carried out, but, the fictional due dates might lose their relevancy pretty soon.

Buffer management, using the same concept of buffer status, and the percentage of the consumption of the planned protection, succeeds in bringing together both types of orders and noting the current priorities.

Being able to hold very low finished goods stock and being very agile in replenishing them, provides a very good stable solution to immediate delivery while keeping the excess capacity for what is truly needed—providing flexibility to grab more opportunities.

## ABOUT THE AUTHOR

Eli Schragenheim is an international consultant and educator. He is the author of *Management Dilemmas*, co-author of *Manufacturing at Warp Speed* (with William H. Dettmer) and of Dr. Goldratt's new business novel *Necessary But Not Sufficient* (also with Carol A. Ptak). He is the developer of several educational simulators that are used to experience the problems in running manufacturing and projects and the TOC solutions.