Statistical inventory control in theory and practice

Curtis P. McLaughlin, Gyula Vastag, D. Clay Whybark*

Business School at Chapel Hill, Campus Box 3490, Carroll Hall, University of North Carolina, Chapel Hill, NC 27599-3490, USA

Abstract

In this paper the authors report on three areas where statistical inventory control (SIC) expectations diverge from reality. First, actual inventory performance seems immune to the use of modern techniques like material requirements planning (MRP) or just-in-time (JIT). Second, simulation studies seem to provide higher than expected customer service levels. Finally, dynamic organizational actions appear to change both the rules of the game and the way that it is scored. These observations suggest that the lack of effectiveness of SIC models in practice cannot be blamed exclusively on the scientists or the practitioners. The paper suggests that practitioners have not done well in applying the models that are available to them. It also points out that theory and practice are still far apart and suggests some research to remedy this.

1. Introduction

Even though scientific papers continue to accumulate in the area of statistical inventory control (SIC), SIC has not enjoyed the attention or impact in recent years that it has in the past. This is partly due to other initiatives in industry and academia like material requirements planning (MRP), just-in-time (JIT), total quality management (TQM), and so forth, and somewhat due to the “bad name” that SIC has picked up in some circles. The SIC problem, even for a single item at a single location, can be extremely complicated as defined by the scientific community. Nevertheless, the problem is “solved” in industry everyday for multiple items in multiple locations.

This paper will explore three broad areas where the scientific expectations for SIC and reality differ.

* Corresponding author.

The first area is concerned with empirical observations of industrial performance. They indicate that actual inventory and order lateness results do not seem to be affected by the new initiatives mentioned above among the small percentage of firms that have implemented them. In the second area, scientific research, many research projects have been based on simulation models which provide levels of customer service that are much higher than the parameters would suggest. This suggests a technical problem in the simulation models. In the area of actual practice, levels of service provided to customers are often substantially higher than would be predicted from models of the inventory control system used, implying that the models do not capture reality.

2. Actual inventory and lateness performance

The data used to evaluate the impact of various management techniques on inventory levels and
lateness performance comes from firms around the world in the small machine tool and non-fashion textile industries. Over the past few years the Global Manufacturing Research Group has gathered data on manufacturing practices in areas of the world as diverse as Australia, China, Hungary, Japan, Korea, North America and Western Europe in a project described by Whybark and Vastag [1]. Among other information, the data base includes finished goods, work-in-process (WIP) and raw material inventory levels, finished goods safety stock and delivery lateness.

In an earlier paper, Vastag and Whybark [2] used some of this data to explore the impact of JIT on manufacturing lead-time, inventory and delivery performance. One of their findings was that JIT techniques are not widely practiced, despite the length of time they have been around and the publicity JIT has gotten. In evaluating JIT methods to help manage inventories, they found that the benefits were limited to reductions in raw materials, something that suppliers have suspected for a long time.

The data base on manufacturing practices does not have any direct information on the use of SIC techniques. Consequently, in this paper, other methods of inventory management are considered. Specifically the use of computers in forecasting and in production planning and scheduling; and the use of the economic order quantity (EOQ) technique, MRP systems and JIT approaches were investigated. These methods were analyzed as to their effectiveness for reducing inventories and decreasing lateness.

The evaluation will be based on a comparison of the inventory levels, safety stock levels and lateness performance of the firms using the techniques or computers with those that do not. The data used comes from more than 400 respondents around the world, with no less than 236 respondents for any single question. All data are normalized. The inventory data are expressed in days of supply of raw material, WIP, and finished goods inventory, while safety stock is expressed in weeks of protection. The lateness performance is expressed as the average number of days late in delivery of an order to the customer.

The statistical significance of the comparisons are computed using the Kruskal–Wallis non-parametric test. This test determines whether the factors (use of computers or techniques) are significant in accounting for differences in inventory and lateness performance. It is based on the ranks of the individual responses for each of the factors.

Table 1 presents the data for the question of whether a firm uses the computer in developing forecasts. The response was simply yes or no. The first observation is that only 20% of the respondents said that they used computers in forecasting. There is an improvement in inventory levels (accept for safety stock) and lateness for those firms using computers although the significance values are not very high. While there is some evidence that using the computers for forecasting could be helpful, their use is not widespread.

The responses to the question on use of computers in production planning and scheduling is presented in Table 2. A higher percentage of the respondents use computers for these tasks than for forecasting. Overall, however, there are still fewer than half that do. The results are not highly significant for these data, except for WIP, and the pattern
is not consistent. Any reductions in inventory gained from the use of the computer seems to be offset by increased delivery lateness.

Table 3 presents the data for analyzing the use of the EOQ. The question asked simply whether the EOQ was used for determining production or purchasing quantities. As with the computer usage, an overwhelming majority, 73%, make no use of the technique. The significance values are not particularly strong, although there is some evidence that the use of the EOQ helps in reducing WIP and finished goods inventories. There seems to be a slight increase in lateness occasioned by the reduction in finished goods inventory, however. When the users of MRP and JIT were dropped from the “no use” category, the differences were still not significant although more than 50% of the respondents use none of the three techniques.

The MRP question asked about use and attempted to get at the intensity of use and success MRP as well. Although more than half had heard of MRP fewer than half were trying it, some without benefits, as seen in Table 4. Some of the responses represent very small sample sizes and the significance values are not very strong. Even when the data were divided into non-users, new or poor users and benefitting users, the results did not change. There is little evidence of benefit from the use of the technique.
A slightly higher percentage of the respondents have heard of JIT than MRP, as can be seen in Table 5. Again, however, fewer than 50% are using it. There is a little more evidence that there is value in implementing JIT, both for inventories and for lateness. There are statistically significant reductions in raw material and decreases in lateness for those benefitting from JIT. This also holds when the data are divided into three groups as they were for MRP. In this evaluation of the application of computers and techniques for managing inventories around the world, only one conclusion can be strongly drawn. There are many fewer firms using these resources than are not. The data on the benefits from their use are very mixed. There is no consistent pattern of inventory reductions or lateness decreases shown. Thus, there is not a compelling argument for increasing their use. By extension, the same can probably be said for statistical inventory control procedures as well.

3. Statistical inventory control research

There is a long history of research in statistical inventory control. Even recently, as shown in a study by Meredith and Amoako-Gyampah [3], 20% of a sample of doctoral dissertations were on inventory control and that increases to more than 40% when scheduling and forecasting are included. The most frequent topic (nearly 25%) of articles published from 1982–1987 was inventory control. The topic is literature driven, however, not industrially defined. This has lead to SIC research getting a bad reputation from many practitioners, consultants and some academics.

The lack of application of SIC techniques, the studied indifference of business managers to the ever expanding literature of SIC, and the lack of testimonials of benefits from SIC models have contributed to the depreciation of the field. On the other hand, the real problem of trying to manage thousands of individual items, facing uncertain independent demand, in several locations still exists. The inability of MRP or JIT to address some of the key issues associated with this problem still leaves the field wide open for appropriate research. How to direct that research should be a key priority for the scientific community.

This section will look at two technical issues arising in past SIC research. The first of these has to do with the definitions of customer service that are used in defining safety stock levels for models of independent inventory systems. The second has to do with the generation of random numbers for simulation experiments, a popular way to conduct research on inventory systems. Both of these issues stem from experience with and observations of simulation research over several years. Many of the simulation results can be characterized as having much lower stockout levels than would be expected from the parameters used. This leads to much higher fill rates (customer service levels) than might be anticipated.
3.1. The customer service criterion

Technical discussions of customer service levels almost always start out with a description of demand uncertainty and the use of "safety stock" to protect against stockouts except for the rare occasions when demand is very high during the lead time (i.e. the time during which a replenishment order is outstanding). Some portion (often \( k \)) of the standard deviation of the demand during lead time (\( \text{SDDLT} \)) is used to calculate the safety stock. For continuous review policies, the reorder point (ROP) is then calculated by adding the safety stock to the average demand during lead time (\( \text{DDLT} \)): 

\[
\text{ROP} = \text{DDLT} + (k \times \text{SDDLT})
\]

Assuming a well-understood distribution, like the normal, appropriately selecting \( k \) can provide a specified level of protection, e.g., a 95% coverage. That means there will be a stockout during only 5% of the reorder cycles. Unfortunately, that is not the same as having 5% stockout levels or 95% fill rates. Moreover, several other service level criteria might be used for SIC as pointed out by Schneider [4].

Vollmann et al. [5] provide the theory and the formulas for calculating safety stock values and fill rates. In order to determine the fill rate, the expected number of stockouts per reorder cycle associated with the \( k \) value must be found. This requires calculating the expected number of stockouts for each time demand exceeds the safety stock coverage and weighting that by the probability of having a stockout. The expected number of stockouts per reorder cycle divided by the demand per reorder cycle will give the expected shortage rate and the complement is the expected fill rate. Clearly, large order quantities that mean long reorder cycles will increase the fill rates for a given uncertainty distribution and \( k \) value.

Using a \( k \) value to establish a probability of stockout and equating that to the shortage rate is a possible explanation for the high levels of customer service that get generated in simulations of SIC systems. Uniformly high service levels means it is hard for researchers to discriminate between the policies that they are evaluating in their studies. As a result they use low or negative safety stock values to reduce service levels in order to discern the differences between them (e.g. [6]).

<table>
<thead>
<tr>
<th>Number of standard deviations above the mean</th>
<th>Probability of a value exceeding ( Z )</th>
<th>Difference in probability values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1587</td>
<td>0.1607</td>
</tr>
<tr>
<td>2</td>
<td>0.0228</td>
<td>0.0223</td>
</tr>
<tr>
<td>3</td>
<td>0.0013</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

3.2. Generating random numbers

Another possible explanation for the high levels of fill rates observed in simulation models of SIC systems has to do with random number generation. Some approaches base the generation of other distributions on the uniform distribution (although different technologies do exist). In early simulations, for example, the generation of normally distributed random numbers was based on the sum of 12 uniformly distributed numbers. The estimation of the normal this way can lead to differences in the probabilities in the tail of the distribution. This has been calculated by Yang [8] and is shown in Table 6.

Table 6 shows that demands in the range of one standard deviation are overestimated, while those in the "tail" are underestimated. The differences are not great, but it takes demands from the tail to have shortages in any reasonable inventory system. Since these high demands are low probability events anyway, any underestimation of demand in the tails would overstate the fill rates relative to the theoretical value. In a variation on the theme, Silver and Rahnama [7] indicate that when statistical estimates of demand parameters are used to calculate reorder points, the reorder points should be biased upwards. In other words, the estimates of the demand distribution parameters underestimate the negative consequences.

4. Inventory control in industrial practice

There are a number of organizational and situational responses that can explain some of the
differences between actual fill rates and those predicted by theoretical and simulation models. The customer-oriented organization has a number of ways, some subtle and some not, to respond to the demands and due dates initially expressed by customers.

These responses can occur at all stages of the process of fulfilling a customer's request: quotation, order entry, stock status checking, order acknowledgement, production planning, production scheduling and dispatching, sales service and expediting, and shipping. At each of these stages there are steps that tend to increase the experienced fill rate. Each stage is discussed here.

4.1. Quotations

Quotations, whether formal or informal, include both price and delivery. If the delivery date (sometimes immediate) is unsatisfactorily based on stock status checks or lead times, the customer will not place the order, biasing the orders actually placed toward higher fill rates. On the other hand, sometimes salespersons will be dishonest in estimating lead times to stimulate sales. This will bias the system toward lower-than-anticipated fill rates.

4.2. Order entry

A savvy sales service or order entry group reviews orders as they are received to identify potential outliers in terms of large quantities, unusual delivery lead time, special tooling requirements, batching requirements for infrequently cycled product groups, and in interpretation of ambiguous instructions such as ASAP (as soon as possible). Properly handled, such interventions can lead to significantly improved fill rates by providing special treatment for probable outliers.

4.3. Stock status

Once the order is entered into the information system, the next step is a stock status check. The specifics of this check depend on the specific inventory system used, but the intent is the same. Orders are netted out against available inventory either on the shelf or in existing production schedules. The stock status check produces information, however, beyond whether or not the item is in stock or on order. Information about the quantity of shortages or displays of run-out times can lead to changes in batch sizes already scheduled and to modifications of the cycling of families of products through the plant. The information from the stock status check can also lead to major adjustments in the stages that follow: order acknowledgement and production planning.

4.4. Order acknowledgement

Most simulation studies assume demands occur according to some specified distribution independent of inventory or production status. This is equivalent to due dates that are set by customers and demands occurring against inventory on the due date without any company knowledge. In actuality, the manufacturer knows that the customer-specified due date may or may not conform to the quoted delivery times and the quoted delivery times usually reflect aggregated backlogs. Thus, the manufacturer seldom accepts the customer delivery request directly, but respecifies the delivery based on the stock status results. Furthermore, the company typically computes its fill rates on this later promised date, not on the customer-specified due date. This adjustment of due dates virtually assures a high fill rate, if the information in the inventory system is accurate.

The customer either accepts the revised date or lodges a complaint on receipt of the acknowledgement. If the customer complains the order is moved out of the routine system that the models represent and into an exception handling mode.

4.5. Production planning

Most inventory theory assumes that a company operates with one planning system or another (e.g. SIC, MRP or JIT). But that is seldom the case. Assembled items may be assembled to order on
a JIT basis out of components that are planned by an MRP system. Original equipment manufacturer (OEM) customers may be on a JIT system, while spare parts are managed by SIC. Major customers may have ordering patterns or special requirements that move their orders from one type of system into a different system.

An example of mixed systems is provided by manufacturer of replacement recording charts for electronic instruments which are normally shipped from stock managed by SIC. However, its circular charts are printed in green and two of its major customers produce photographic film under red lighting. Green cannot be read under red lights. Therefore, the charts for these two customers are printed quarterly in special colors and shipped in separate lots. Obviously, the fill rates for these special items are 100%.

The same manufacturer has some large government customers for chart paper who order their annual requirements on a single bid which calls for a single shipment against a relatively long scheduled lead time. Again, these orders would have unusually high fill rates, because they were not processed through the SIC system through which a high percentage of the orders, but a lower percentage of the unit volume proceeded.

4.6. Production scheduling and dispatching

Once production planning is done, production scheduling and dispatching respond in a number of ways to out-of-stock situations and to the special needs of key customers. For example, one of the situations where an automated SIC system often has difficulty is the new stock item, since there is little sales history. However, given the potential importance of such items, production planning often overrides the SIC system to make sure that the new item is available to fill pipelines and to sample new customers. Similarly, production planning meetings emphasize any shortages of high volume or key items and these are pushed ahead in the dispatching process. Where products are batched in families, inventory status and internal politics can determine which items go first within the family sequence.

4.7. Sales service and expediting

A continual set of negotiations and responses goes on between the customer, customer service, production planning, the shop floor, and shipping to maximize fill rates by mechanisms seldom duplicated in models or simulations. Delivery dates are moved forward and backward in response to the needs of key (or noisy) customers. The lagging items are expedited, or in some cases stolen from other orders having longer lead times calling for the same items.

As an example, consider a firm that receives an order for an item on March 14 for shipment on June 1. The item is in stock, but this order drops available inventory below the reorder point triggering production of a batch of the item to be delivered on May 15. A second order is received on March 21 for delivery April 30 which calls for a quantity in excess of the available inventory. The system would treat the second order as out-of-stock and ship it on May 15 when the new batch is physically available. However, an expediter could note that the material for the June 1 delivery could be taken from the batch arriving on May 15 and the order due out on April 30 could be filled from the existing stock, avoiding any loss of fill rate. Alternatively, if the customer uses the item continuously, the sales service organization could call the customer and arrange to split the order, shipping part now from stock and the rest on May 15. Then, instead of the failure to fill initially reported, there are two orders, both of which are filled on time.

4.8. Shipping

For the customer, the operative fill rate concerns the arrival of the material at their point of use, not the time of departure from the warehouse or plant. Therefore, customer-observed fill rates can be affected by choices of modes of transportation and by enhanced coordination with transportation firms. The addition of electronic data interchange (EDI) systems for the rapid transfer of order data, when combined with periodic review and shipment systems, also serves to speed up cycle times and provide better coverage in the field.
Most academic studies treat each order as if it were only for a single item. More often, however, orders are received for multiple items. This is especially true in the distribution systems that have truly independent demand from wholesalers and retailers. With multiple-item orders one has to deal with discrepancies between line item and total order fill rates. In logistics systems, due dates are often as much determined by the shipping schedule as by the customer’s needs or the completion of the order. Increasingly economies of scale in transportation have pushed distribution into periodic review systems, which theoretically call for larger safety stocks than continuous review systems. However, periodic shipment systems are ideal for the shipment of partial orders, since the incremental transportation cost of splitting the order is nil.

The combination of periodic shipments, EDI, order splitting, multiple items per order and customer needs means that the fill rate must be defined to meet the customers’ needs, not the producer’s system. For example, if split line item orders mean high-order fill rates, but low item fill rates, the real question must be how well the customer was served. All this takes negotiation — something that inventory research does not currently encompass.

4.9. Summary

Many real situations involve tactics by inventory managers that are not well modeled by our current scientific approaches. Managers may change demand by influencing customers’ orders when stocks are low (i.e. asking a customer to postpone an order or to order less). In some instances the measures are changed in order to make the reports look better (i.e. counting a partial order as “meeting the customer needs”). In many instances there is extra safety stock in the drawer of a salesperson’s desk or in an inventory bin known only to a few people. In numerous cases the managers override parameters in an effort to improve their service measures.

The models and simulations of inventory theory are intended to be simplified representations of the real world. However, the real world is itself both complex and dynamic. Given the existence and use of these organizational and situational responses, it is not surprising that what is reported empirically seems to work out somewhat better than what would be expected theoretically.

5. Conclusions

This paper accepts that there is a need for statistical inventory control (SIC) in industry and notes three areas where expectations diverge from reality. First, actual inventory performance seems immune to the use of modern techniques like MRP or JIT. Second, simulation studies seem to provide higher than expected fill rates. Finally, dynamic organizational actions appear to change both the rules of the game and the way that it is scored. These observations suggest that the lack of effectiveness of SIC models in practice cannot be blamed exclusively on the scientists or the practitioners.

The implications for research are clear. The models must permit the balancing of supply and demand on a short term, negotiated basis. This might call for research using open simulations with interventions by decision makers in two “organizations” that use various information support systems to negotiate with each other. More field research may be required in order to establish the nature of demand, due date setting and the other parameters that feed the models of SIC situations. In any event there is certainly no evidence that SIC is dead (or should be) as a field of study.

References


