Finished goods management for JIT production: new models for analysis

By: Richard Ehrhardt


Made available courtesy of Taylor and Francis: [http://www.taylorandfrancis.com/](http://www.taylorandfrancis.com/)

***Reprinted with permission. No further reproduction is authorized without written permission from Taylor and Francis. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document.***

Abstract:
A firm is considered that manages its internal manufacturing operations according to a just-in-time system, but maintains an inventory of finished goods as a buffer against random demands from external customers. The finished goods inventory may be analysed by the methods of classical inventory theory in order to characterize the trade-off between inventory costs and schedule stability. A model is formulated in which the supply of finished goods is replenished by a small fixed quantity each time period. The size of the replenishment quantity may be revised only at pre-specified intervals. The single-interval problem is analysed, the cost-minimizing value of the replenishment quantity for a given revision interval length is computed, and the optimal cost is characterized as a function of the revision interval length. The dynamic problem is shown to be convex, with relatively easily computed optima. Finally, alternative formulations of the problem are described and suggestions made for further research.

Article:

1. Introduction

Just-in-time (JIT), developed in Japan, has become an effective approach to gaining competitive advantage in manufacturing. This approach, which integrates and manages manufacturing activities, obviously leads to cost efficiency and to consistently high product quality. Its lean inventories also allow rapid response to the changing needs of the marketplace. Many firms have been successful in using JIT to integrate and manage their internal operations and even to coordinate their operations with those of their suppliers, but difficulties often arise at the interface with the marketplace. That is due to the demand for finished goods, which is often unmanageable and random. For this reason, many firms still maintain a buffer inventory of finished goods. These types of inventory systems are not well managed by the traditional models of stochastic inventory theory because JIT production processes assume stable production schedules. This paper proposes management models that are designed for balancing the conflicting needs of a stable JIT production schedule and the unpredictable demands of the marketplace. We consider two general approaches for restraining changes to the production rate.

One approach requires the replenishment quantities (or number of Kanbans) to remain constant for a fixed interval of time, and to then allow revision to a series of larger or smaller quantities which remain fixed for the same time interval. One can investigate then how the optimal inventory policy and resultant costs depend upon the length of the interval. It is then left to the manager to make an informed trade-off decision between inventory costs and schedule stability. This environment may arise when a capacity constraint forces management to set the values of the replenishment quantities of all products simultaneously (see Section 6). The trade-off decision would necessarily be highly dependent upon the details of the manufacturing environment and possibly upon the firm’s marketing strategy as well. For example, schedule stability may be more highly valued in a very complex manufacturing system where bottlenecks may develop in a number of different areas when production schedules change. Ideally, the models of this paper would be linked to detailed models of the manufacturing system’s internal operations (a true CIM application). Then the specific internal impacts of schedule instability can be estimated and weighed against the inventory and service impacts predicted by the models of this paper.
A second approach requires the replenishment quantity to remain fixed unless the inventory of finished goods strays outside specified bounds, at which time it is increased or decreased. One can then investigate the cost and stability effects of various policies for setting the bounds and replenishment quantity change-amounts. As in the first approach, the objective is to provide managers with tools for making informed trade-off decisions between inventory costs and schedule stability.

Ultimately, the results of this research could form a software module that could be integrated with other models that seek to improve the internal operation of JIT. Such models utilize the methods of queuing networks to investigate appropriate batch sizes and buffer stocks at multiple work centres within the manufacturing organization. If successfully developed, the models of this paper could form an important vital link between the technology of managing internal production logistics and the random forces of the marketplace, thus helping manufacturing managers to more effectively employ JIT to make their firms economically competitive and to gain strategic advantage.

We survey the existing modelling literature in Section 2, and then proceed to study the model in which the replenishment quantity stays constant for fixed intervals of time. We formulate and analyse the single-decision case in Section 3, and then formulate and discuss the dynamic version of the problem in Section 4 (we present a sample of computational results as well). Section 5 discusses models that revise the replenishment quantity based on the level of inventory, and Section 6 formulates a multi-item version of the fixed-revision-interval model. Finally we discuss general research strategies in Section 7, and summarize our study in Section 8.

2. The models

JIT production systems attempt to reduce inventories of raw materials, work-in-process and finished goods. The general approach is simple: produce (or deliver) small batches of items in the precise amounts needed by subsequent production processes (or customers) at exactly the time needed. In general, JIT is a pull system, in which a production process pulls material from a prior process (which doesn’t operate unless there is need for its output) in support of the final assembly schedule, which is closely coordinated with customer demand (Krajewski and Ritzman 1987).

There is general agreement that for JIT to work effectively, production schedules must be level and stable (Hall 1983, Monden 1983, Vollmann et al. 1984). While this condition is agreed upon, research has tended to focus upon translating a sequence of known future demands for products into a final assembly schedule that demands parts as uniformly as possible over time (Miltonburg 1989, Steiner and Yeomans 1993). A parallel body of research investigates methods for devising effective schedules when the production system is not organized as a JIT pull process (Jackson et al. 1988, Maxwell and Muckstadt 1985, Roundy 1985). In both cases, the assumption of deterministic demand allows research to address the complex issues of detailed scheduling of multiple products and multiple workcentres.

Research is less copious when demand is stochastic. Most papers model such production planning problems using stochastic inventory theory, i.e. they regard the production process as the supplier of goods and manage the finished goods warehouse in a manner that minimizes setup and inventory costs. Although many studies have been published, the complications of stochastic demand lead to single-item models with few details of the production process included (Gavish and Graves 1980, Doshi et al. 1978, De Kok et al. 1984, Altio 1989). Much of the JIT literature, on the other hand, has been, until recently, qualitative. This should not be surprising, as the JIT approach is inherently organizational. Much of its appeal lies in its simplicity of operation and the subtle synergistic benefits that result. While promoters of JIT emphasize the advantages of reducing inventories to zero, some theorists and many practitioners recognize that some inventories are desirable (Zangwill 1987). This is especially likely when certain aspects of the system are unavoidably random.

We formulate mathematical models aimed at one of the gaps between the mathematical analysis of inventories and the practical, non-mathematical realities of JIT. A stream of research already exists for the interface between suppliers of input materials and the JIT process (Yano and Gerchak 1989, Pan and Liao 1989,
Other studies have modelled both the supply and delivery ends of the process, but with deterministic demand (Goldhar and Sarker 1992, Sarker and Parija 1994). We seek instead to model the interface between the JIT process and the random demands of customers.

Many firms operate a JIT manufacturing process in-house, but produce to stock, that is, they maintain an inventory of finished goods as a buffer against random customer demands. The modelling approach we discuss here is to minimize the expected value of costs associated with finished goods inventories (or to effectively achieve customer service targets) while restraining changes in the replenishment quantity so as to preserve the stability of the production schedule that is so important to JIT production.

One way in which the trade-off between the competing forces of schedule stability and inventory costs appears in the professional literature is production smoothing models. These models take the traditional inventory theory approach and add a cost for changing the replenishment rate to the usual costs of replenishment, holding inventory, and backlogging or losing unmet demand. It has been shown (Sobel 1969) that the form of an optimal policy is rather complicated, and considerably more difficult to compute than what is usually encountered in a pure inventory management model. We choose instead to ignore the cost of changing the production rate and leave the trade-off decision to management. It may be that other models could ultimately be linked to ours, allowing details of the production process to be factored into the trade-off decision.

We consider two general approaches for restraining changes to the production rate:

1. **Fixed revision intervals.** One approach is to require the replenishment quantities (or number of Kanbans) to remain constant for a fixed interval of time, and to then allow revision to a series of larger or smaller quantities which remains fixed for the same interval of time, and so on. One can then investigate how the optimal inventory policy and its costs depend upon the length of the interval. It is then left to the manager to make an informed trade-off decision between inventory costs and schedule stability. This environment may arise when a capacity constraint forces management to set the values of the replenishment quantities of all products simultaneously.

2. **Inventory-triggered revision.** A second approach is to require the replenishment quantity to remain fixed unless finished goods inventory strays outside of management specified bounds, at which time it is increased or decreased as appropriate. One can then investigate the cost and stability effects of various policies for setting the bounds and replenishment quantity change-amounts. As in the fixed revision interval case, the objective is to provide managers with tools for making informed trade-off decisions between inventory costs and schedule stability.

The next section investigates in detail the management of one interval of $n$ time periods during which the replenishment quantity must remain constant. The reader may wish to temporarily skip forward to the following section (Section 4) which formulates a dynamic model, embedding the details of this model in a macro-sequence of revision intervals. Research on the dynamic model is ongoing.

### 3. Fixed revision interval static model: analysis of a single decision

We consider a single decision to fix the replenishment quantity, followed by an eventual salvage of any excess stocks or satisfaction of any backlogged demand. Consider the problem of selecting a fixed replenishment quantity $z$ to be delivered in each of $n$ consecutive periods. Let $x_t$ represent the inventory level at the end of period $t$, after having experienced demand of size $D_t$. We assume that $D^o \{D_t, t \in \mathbb{N}\}$ is an iid sequence with common cdf $F$ and mean $l$. Letting $x_0$ represent the beginning inventory level, we have

$$x_t = x_{t-1} + z - D_t = x_0 + tz - \sum_{i=1}^{t} D_i$$

$$= x_0 + tz - S_t \quad t = 1, \ldots, n$$
where the $S_i = \sum_{j=1}^{i} D_i$ are the partial sums of $D$. Let $F_t(u) \equiv P(S_i \leq u)$ denote the $t$-fold convolution of $F()$.

Suppose the relevant costs consist of a linear replenishment cost and a holding or shortage cost in each period, and a salvage cost for any inventory remaining at the end of period $n$. Specifically, let $c$ represent the unit replacement cost, $g(y, d)$ be the holding or shortage cost in any period if $y$ is the inventory available to meet demand and $d$ is the demand in that period, and $s(x)$ be the salvage value of $x$ items of inventory remaining on hand at the end of period $n$. Letting $C(z, D, x_0)$ denote the total $n$-period cost if the initial inventory is $x_0$, we have

$$C(z, D, x_0) = \sum_{i=1}^{n} \left[ cz + g_i(x_{i-1} + z, D_i) \right] - s(x_n)$$

$$= nz + \sum_{i=1}^{n} g_i(x_0 + tz - S_{i-1}, D_i) - s(x_0 + nz - S_n)$$

where $S_0 \equiv 0$. Suppose that $s(\cdot)$ is convex, and $g(\cdot, d)$ is convex for each $d$. Then $C(z; x_0) \equiv \varepsilon C(z; D, x_0)$ is a convex function, and first-order conditions establish optimality.

A special case of general interest is when the salvage value, and holding/shortage cost functions are linear:

$$g(u, d) = h \cdot (u - d)^+ + p \cdot (d - u)^+$$

$$s(u) = s \cdot (u)^+ - c \cdot (-u)^+ \quad 0 \leq s \leq c$$

Then

$$C(z; x_0) = nz + \varepsilon \sum_{i=1}^{n} \left[ h \cdot (x_0 + tz - S_i)^+ + p \cdot (S_i - x_0 - tz)^+ \right] - s \varepsilon (x_0 + nz - S_n)^+ + c \varepsilon (S_n - x_0 - nz)^+$$

which, after some manipulation, takes the form

$$C(z; x_0) =$$

$$\left[ c(nz - x_0) + (c - s) \int_{0}^{x_0 + nz} (x_0 + nz - u) d F_s(u) \right] + p(nz - x_0) + (h + p) \int_{0}^{x_0 + nz} (x_0 + tz - u) d F_t(u)$$

Notice that the bracketed terms represent the costs of producing and salvaging goods, while the remaining terms represent the costs of inventories and backlogs. Taking the derivative, we obtain

$$C'(z; x_0) = (c - s)n F_s(x_0 + nz) - \frac{pn(n+1)}{2} + (h + p) \sum_{i=1}^{n} F_t(x_0 + iz)$$

Setting $C'(z; x_0) \geq 0$ and rearranging terms, we find that the optimum value of $z$ is the smallest non-negative $z^*$ satisfying

$$A(z^*; x_0) + \frac{2(c - s)}{(n + 1)} F_s(x_0 + nz^*) \geq \frac{p}{(p + h)}$$

(2)

where

$$A(z; x_0) \equiv \frac{2}{n(n+1)} \sum_{i=1}^{n} F_t(x_0 + iz)$$

Notice that $A(z; x_0)$ has all the properties of a cdf, i.e. it is
non-decreasing and has limits of 0 and 1. When \( n = 1 \), \( A \) is the demand cdf, \( F \). One can show that as \( n \) increases, the mean of \( A \) remains fixed, while its variance increases. This fact is illustrated in Figure 1, and leads one to suspect that optimal expected costs per period would also increase with \( n \). This result is similar to the concept of equivalent demand distributions (Naddor 1966).

As the second term of equation (2) is concerned with salvage costs, it is clear that if \( c = s \) the optimal order size is governed by a generalized news vendor expression, namely the smallest non-negative \( z^* \) satisfying

\[
A(z^* ; x_0) \geq \frac{p}{(p+h)} \tag{3}
\]

A significant consequence of this result is that, unlike the traditional news vendor and \((s,S)\) models, minimizing expected costs does not result in a target service level (probability of satisfying all demand) which is uniquely determined by the value of \( p/h \). The optimal service level will generally vary from one time period to another, and the average service level

\[
\frac{1}{n} \sum_{t=1}^{n} F_t(x_0 + tz)
\]

may be greater or less than the familiar ratio \( p/(p+h) \). For example, when \( n = 8 \), demand follows the Poisson distribution with mean 4, \( p = 9, h = 1 \) and \( c = s = 0 \), the optimal average service level is 0.869, which violates the news vendor criterion (service greater than or equal to 0.9).

Figure 1. \( A(z;0) \) versus \( z \) for \( n = 1, 3 \) and 8; Poisson demand with mean 4.

The expected cost per period \( C(z;0)/n \) is graphed in Figure 2 assuming Poisson demand with mean 4, \( p = 9, h = 1 \). We assume also that \( c = s = 0 \), as this is perhaps most indicative of the behaviour of the long-run dynamic model. For this case, we find that \( z^* \) decreases from 7 to 6 to 5 as \( n \) increases from 1 to 3 to 8.

Figure 2 illustrates how the penalty for choosing a non-optimal value of \( z \) increases with \( n \). For example, when \( n = 8 \) the expected cost per period is nearly double the optimal value if \( z \) is set at 4 rather than 5. Here we see another significant managerial implication. The desire for a stable production schedule argues for larger values of \( n \), with the unfortunate consequence of higher penalties for misspecified values of \( z \). Many managers do not have precise demand information, and must rely upon statistical estimates of demand parameters, so high penalties will likely be paid for stable production schedules.

4. Fixed revision interval dynamic model: sequential decision making

Consider a sequence of \( N \) decisions like the one described above. We shall refer to each sequence of \( n \) periods with constant replenishment quantities as a \textit{revision interval}. Let \( z_i \) be the replenishment quantity chosen for the
ith revision interval, and let $x_{i,t}$ and $D_{i,t}$ be the ending inventory and demand for the $t$th period within the $i$th revision interval, respectively. We assume that

$$\mathcal{D} = \{D_{i,t}; t = 1, \ldots, n, i = 1, \ldots, N\}$$

is an iid sequence with common c.d.f. $F$. Letting $Z \equiv \{z_{i,i} = 1, \ldots, N\}$ be the sequence of replenishment quantities chosen, and $T(Z; x)$ be the total cost for all revision intervals when the initial inventory is $x$, we have

$$T(Z; x_{1,0}) = \sum_{i=1}^{N} \sum_{t=1}^{n} [c_{i,t} + g(x_{i,t-1} + z_{i,t}, D_{i,t})] - s(x_{N,n})$$

$$= \sum_{i=1}^{N} \left[ nc_{i,t} + \sum_{t=1}^{n} g(x_{i,0} + tz_{i,t} - S_{i,t-1}, D_{i,t}) \right] - s(x_{N,n})$$

if the replenishment quantity can be adjusted without cost at the start of each revision interval. To find the replenishment policy that minimizes the expected total cost, it is convenient to formulate the problem as a dynamic program. Let

$$G(z, x) = nc_{x} + \mathbb{E} \sum_{t=1}^{n} g(x + tz - S_{t-1}, D_{t}) \quad (4)$$

Notice that if the holding and shortage costs are linear, then $G(z, x)$ has the same form as the single decision cost function $C(z; x)$ (1) if $s = c = 0$. Defining $f_i(x)$ as the minimum expected cost from revision interval $i$ through the end of the horizon, if we start revision interval $i$ with $x$ units of inventory, we have

$$f_i(x) = \min_{z \geq 0} \left\{ G(z, x) + \mathbb{E}_{i+1} \right\} \quad (5)$$

$$f_{N+1}(x) = -s(x)$$

The recursion is similar to those governing standard inventory problems with linear purchase cost functions. Optimal policies can be characterized by studying the properties of the function $G(z, x)$. Of particular interest is how the optimal expected cost per period varies with $n$. This information would be of managerial interest.
because it characterizes the trade-off between the stability of the production schedule and finished goods inventory costs. The recursion is different from the standard inventory problem’s because the function $G$ is more complicated. It cannot be simplified by conversion to a function of $y = x+z$ because of the sum over periods within each revision interval. Because of this, it almost certainly does not admit myopic optimal policies (Heyman and Sobel 1984).

We illustrate the behaviour of dynamic optima, using recursion (5) to compute optimal dynamic policies for the assumptions of Section 3, setting $n = 1, 3$ and 8, and $N=1, 2, 3$ and 4. We compare optimal policy functions in Figures 3 and 4, where first-period optima ($N$ periods remaining) are graphed.

When the revision interval length $n$ is only 3 periods, Figure 3 displays near-myopic optima, as optimal policies are identical for $N=2$, 3 and 4, and the single-interval policy ($N=1$) is also the same when beginning inventory levels are at least -2. The situation is quite different for the longer revision interval of 8 periods. Figure 4 displays significantly different ordering policies as $N$ is increased from 1 to 4.

Although additional research is needed to clearly understand this model, especially for the important infinite horizon problem, the following observations can be made at this point. First, the recursion formulated above defines a Markov decision process so we can conclude that the infinite horizon version of the problem has a stationary optimal policy. Even though the problem is more complicated than the standard stochastic inventory problem, the cost functions are all convex so that relatively simple rules should govern the computation of optimal policies. It seems reasonable to speculate that the infinite horizon optimal policy for a given value of $n$ would be given by a function $z^*(x)$, a non-decreasing function of the starting inventory level $x$ of the revision interval. Results from the single-decision model suggest that the expected cost per period would be increasingly sensitive to misspecified replenishment quantities as $n$ increases, a property with potentially serious implications for managerial decision making.

![Figure 3](image.png)

**Figure 3.** Optimal replenishment quantities for $n=3$; Poisson demand, $\mu = 4$, $p/h = 9$.

### 5. Inventory-triggered revision of replenishment quantities

The inventory-triggered revision models allow the replenishment quantity to remain fixed unless finished goods inventory strays outside of management specified bounds, at which time it is increased or decreased as appropriate. We shall use the term target range to indicate those inventory levels between the upper and lower bound for which the replenishment quantity remains fixed. When inventory strays outside the target range, the replenishment quantity maybe changed in a variety of ways. We discuss two possible ways. The first is to increase or decrease $z$ by a fixed amount $a$, so that replenishments stay this new size until inventory is again outside the target range. Defining $zt$ to mean the replenishment quantity in time period $t$, the inventory level at the end of period $t$ would be given by
The second way to adjust the replenishment quantity when inventory strays outside the target range is to order a double batch or no batch at all for the next period only, that is, a one-time added or missed lot. The first approach would appear to be more effective when demand distributions are likely to vary with time, while the second approach might be more appropriate when demand distributions are time homogeneous. Random walk theory could be used to better understand the dynamics of this model. Of special interest is the relationship between the expected revision frequency and the model’s control parameters $U$, $L$ and $a$. Numerical computations could focus on values of the control parameters that match expected revision frequencies with those investigated in the fixed revision interval study to make comparisons between models easier.

6. A multi-item model

The multi-item model can be conceptualized in the following manner. A fixed-revision-interval approach is assumed because the capacity constraint would make the inventory-triggered approach unmanageable. Consider a family of $m$ products that are managed on the same fixed revision interval, and let $x$ and $z$ be $m$-vectors. A production capacity constraint is imposed each time the products’ replenishment quantities are revised. New versions of (4) and (5) are

$$G(z, x) \equiv \sum_{j=1}^{m} [cz_j + 2 \sum_{i=1}^{n} g_i(x_j + tz_j - S_{i-1}, D_i)]$$

and

$$f_i(x) = \min_{z \geq 0} \left\{ G(z, x) + \mathcal{D}_{i+1}(x + nz - S_n) \right\}$$

$$\sum_{j=1}^{m} c_j(z_j) \leq C \right\} = 1, \ldots, N$$

$$f_N(x) = -s(x)$$

This recursion, and its infinite horizon version could be analysed to better understand how to manage a capacity constrained manufacturing facility where schedule stability is valued. As the recursion suffers from dynamic programming’s infamous curse of dimensionality, obtaining computational insights may be challenging.
7. Research themes
The objectives of the models described above are both to shed light on possible new methods of managing manufacturing systems and to contribute to the modelling literature of industrial engineering and management science. It is imperative that research aimed at applying theoretical results to real management systems recognizes the limiting aspects of information availability and computing complexity in applied settings. Therefore, a three-fold procedure is proposed in examining these models:

1. Theoretical analysis is employed to understand the mathematical properties of optimal policies under the most general assumptions that yield analytic solutions.

2. The properties of optimal policies are exploited to derive analytic approximations (heuristic policies) that are numerically accurate and more amenable to use in applied settings. Issues of computational simplicity and limited statistical information regarding parameter values are of central interest here.

3. The approximations are functionally adapted to be effective in the context of applying statistical estimates as values for the model parameters. The approximations may be explicitly modified to accommodate the statistical context, or the research may examine the effects of using standard statistical estimates in place of actual parameter values.

An example of this general approach is found in the context of the classical stochastic dynamic inventory model with independent and identically distributed demands, linear holding and shortage costs, and a replenishment setup cost. It was first shown (Scarf 1959) that the optimal policy is of the $(s,S)$ form, and then an iterative algorithm was devised (Veinott and Wagner 1965) to compute optimal policies. Asymptotic renewal theory was then used (Roberts 1962) to find limiting functional forms that governed policy behaviour as parameters grow large, and those functional forms were used (Ehrhardt 1979) to devise approximately optimal closed-form expressions for the policy. These formulas are not iterative and therefore extremely fast to compute. Further, these analytic approximations are immediately amenable to the context of statistically estimated demand parameters since they involve only the mean and variance of the demand distribution (and the lead-time distribution, when the lead-time is stochastic). It was also shown (Ehrhardt 1979) how the formulas performed when exact values are used for the demand parameters, and also how performance is affected when the parameter values are replaced with periodically revised statistical estimates. This entire approach was later extended (Ehrhardt 1984) to a generalization of the model that includes stochastic delivery lead times.

8. Summary and conclusions
We have formulated a variety of models for studying the trade-off between finished goods inventory costs and JIT manufacturing schedule stability. First we studied a single decision to fix replenishment quantities for a sequence of $n$ periods, and found a generalization of the classical news vendor result. Numerical illustrations suggest that as $n$ grows larger, expected costs per period become more sensitive to setting the replenishment quantity suboptimally. This could have serious implications for the manager who wishes to have a very stable production schedule and therefore desires a large value of $n$.

We also formulated a dynamic version of the model and showed how it is more complicated than the classical stochastic inventory model. The infinite horizon model should have a stationary optimal policy with a reasonably simple form.

Finally, we indicated other types of model formulations which could apply to the general finished goods management problem. These models would be more responsive to the inventory level, but could lead to frequent schedule adjustments when multiple products are managed in a single facility.

References


