THE PRODUCT CYCLING PROBLEM IN SYSTEMS WITH UNCERTAIN PRODUCTION YIELDS

Candace Arai Yano
Department of Industrial and
Operations Engineering
The University of Michigan
Ann Arbor, MI 48109-2117

THE PRODUCT CYCLING PROBLEM IN SYSTEMS WITH
UNCERTAIN PRODUCTION YIELDS

Candace Arai Yano

Department of Industrial and

Operations Engineering

University of Michigan

Ann Arbor, MI 48109-2117

December 1987

Revised March 1990

THE PRODUCT CYCLING PROBLEM IN SYSTEMS WITH UNCERTAIN PRODUCTION YIELDS

ABSTRACT

This paper addresses a production planning problem for a single machine producing multiple parts, where the yield rate (fraction acceptable) for each part may be uncertain and the production policy is to cycle through the set of parts. We analyze a simplified version of the problem which, while still quite complex, becomes solvable as a result of several properties that we prove in the paper. We also provide numerical results to illustrate the solution procedure.

THE PRODUCT CYCLING PROBLEM IN SYSTEMS
WITH UNCERTAIN PRODUCTION YIELDS

1. INTRODUCTION

The single machine economic lot scheduling problem (ELSP) is to find a repeating production schedule for several products, each with a constant demand, which are produced on the same machine. Such problems arise often in the manufacture of glass, rolled steel, and paper, where one machine (or one system of machines) provides adequate capacity to produce a variety of products. Similar problems are found in systems where, for simplicity of scheduling or other technical considerations (e.g., tooling), a machine is dedicated to a family of parts. The increasing adoption of group technology concepts in the design of manufacturing systems has led to the widespread occurrence of systems in which a machine is dedicated to a part family. The primary tradeoff in these problems is the time and/or cost involved in changing over from one product to another, and the inventory that accumulates if changeovers are too infrequent.

In many of these processes, the yield rates are random, and are not observed until after the product has completed its production run (i.e., after inspection). In the examples above, even with in-process sensors and feedback, we would have to wait for the glass to cool or the paper to dry before all product defects could be detected. In other systems, sensing and/or feedback may be infeasible or uneconomical, but the problem remains the same. Our research was motivated principally by applications in which (i) process technology is changing rapidly enough that the technology often becomes obsolete before the process is well-understood and well-controlled, or (ii) factors other than the process itself have an impact on the yield, but are difficult to control closely. These are situations in which the "yield problem" cannot be solved even with state-of-the-art technology, and methods are needed to cope with anticipated ongoing yield uncertainty. The procedures that we develop also can be used to assess the benefits of installing improved equipment or in-process controls.

Research to date on economic lot scheduling (see Elmaghraby 1978 for a review; also Hodgson and Nuttle 1987 and Dobson 1988 for more recent references) has assumed that production yields are deterministic, if not perfect. It is already known that

the general deterministic, discrete time, single-machine ELSP is NP-hard (Hsu 1983). Consequently, many practitioners resort to using good or optimal "pure rotation" policies in which products are produced using a fixed permutation of the products, and the permutation is repeated again and again (e.g., ABCDABCD). This is commonly referred to as the *product cycling problem*. More recently, "powers of two" policies, in which each product is produced at an interval which is a power of two, have been analyzed (e.g., Maxwell and Singh 1983).

Almost all of the research on lot-sizing for processes with random yields has dealt with single-product situations and has not considered capacity constraints. The reader is referred to papers by Karlin (1958), Silver (1976), Mazzola et al. (1987), Sepehri et al. (1986), and Gerchak et al. (1985) for a representative set of results on the single-product problem, as well as surveys of related work. Research on lot sizing with random yields is surveyed by Yano and Lee (1989). We are not aware of any published research on lot-sizing issues in multi-product environments with capacity constraints and random yields. For the convenience of the reader, in Appendix 1 we provide an explanation of how random yield problems differ from those with random demand, and why it is not always possible to convert a random yield problem into an equivalent problem with random demand.

We investigate the problem of finding a pure rotation schedule for a single machine with uncertain production yields. We consider a simplified, static (planning) version of the problem for two reasons. First, even this problem is difficult, as we will explain in the next section. Second, Jones and Inman (1989) have found that for the deterministic version of the problem, the pure rotation policy is not far from optimal for many realistic problems.

We view the resulting plan as a basis for specifying targets that provide guidelines for controlling the system while considering long-term aggregate tradeoffs. At the control level (not addressed here), shorter term decisions are made to keep the system near the target. Without a hierarchical decomposition, the problem is intractable, as we explain later in the paper. We focus on the planning problem for three reasons. First, this problem has not been addressed in the literature, and its solution is necessary before it is possible to develop appropriate control policies. Second, both the goals, and the types and degree of flexibility involved in short-term decisions often depend on the specific application. On the other hand, once it has been decided what aspects of the plan constitute the "target," it is usually not difficult to formulate an appropriate problem and to find techniques to solve it. Third, many practitioners will agree that a good (realistic and cost-effective) plan is useful in itself,

even if a simple (manual) control policy is used to keep the system near the target. Indeed, a very common mode of operation in manufacturing operations is for "management" to specify a set of targets and for production workers or their supervisors to find a good way to meet those goals.

Section 2 contains a detailed problem statement and assumptions. A solution approach for the problem is developed in section 3. In section 4, we describe a particular manufacturing facility which fits most of the characteristics in our problem description and discuss how our results might be used to help that facility. Section 5 contains a study of the effects of different yield rate distributions on the long term problem using an adaptation of data obtained from this facility. Finally, section 6 contains a summary and conclusions.

2. PROBLEM STATEMENT AND ASSUMPTIONS

We address the problem of determining pure rotation policies for a single machine with uncertain yields. Annual demand for each product is assumed to be known and to occur constantly over time, although demand rates for the various products may differ. We assume that the yield rates (i.e., the ratio of usable output to input) have distributions which may differ among products but are stationary over time, and do not vary with the batch size. This is a reasonable assumption when external factors (such as characteristics of the raw materials, temperature or humidity) affect the entire batch in much the same way, when the effect of yield loss due to startup is minimal, and when equipment does not deteriorate with the length of the production run. This particular means of modeling yield uncertainty has become increasingly popular because it is a fairly accurate representation for many applications (see for example, Bitran and Dasu 1989, Lee and Yano 1988, Bassok and Akella 1988) and can be handled both analytically and computationally.

We assume that the facility operates continuously (except for maintenance), making overtime production impossible. It would be possible to relax this assumption, but the modeling (already very complicated) would be difficult to handle except with very simple assumptions about the overtime policy (e.g., equal amount of time every day.)

We assume that the output is 100% inspected and that the inspection process is perfect. It is also assumed that all defective output is treated in the same way: either disposed at no additional cost or reworked. In the former case, the variable cost per unit is the entire variable production cost plus the inspection cost. In the latter case, the

variable production cost is only the cost of processing on the machine in question plus the cost of inspection, since the input material can be reused. We assume, however, that all excess demand is backordered, so in the long run, the yield-adjusted demand of each part must be produced each year. Consequently, variable production costs are not included in the formulation.

We assume that there are non-negative setup costs and setup times for each item which may be sequence dependent. Capacity is limited and setups decrease the available productive time. One of the difficulties in determining optimal pure rotation policies is sequence dependence. However, if (i) the same rotation policy minimizes both setup costs and setup time, (ii) the setup times are sequence independent, or (iii) the setup costs are sequence independent, the problem is simplified considerably. In the second case, it is sufficient to minimize setup costs, and in the third case, it is sufficient to minimize setup time. In the first case, either can be done. Each appropriate problem can be solved using any optimal or heuristic traveling salesman procedure (see Lawler et al. 1985 for further details). For simplicity, we will assume that one of the conditions above is true. Other costs to be included are inventory holding costs and shortage costs. We later explain how these costs are charged.

Even if we restrict our solution to pure rotation policies, this problem is difficult to solve optimally because of the size of the state space. The state would be defined by:

(i) the product for which the machine is set up or being set up; and

(i i) the inventory position of each product.

If the machine is being set up, the optimal decision is easy because the solution is restricted to pure rotation policies, so the setup must continue. On the other hand, if the machine is already set up for a product, an exact policy must specify *for each possible inventory position vector* whether to produce another unit of the current product or to set up the next product in the rotation. Even the problem of finding the infinite-horizon optimal solution for a particular inventory position vector is difficult.

This problem bears some resemblance to a variety of other problems, including "restless" multi-armed bandit problems (Weber and Weiss 1989) and polling systems in the queueing literature (Takagi 1989). However, our problem has features that have yet to be incorporated into existing approaches to these related problems. For example, approaches to restless multi-armed bandit problems cannot yet handle changeover costs or times, and models of polling systems cannot efficiently handle non-exhaustive service (which roughly corresponds to shortages in our model), general arrival distributions (deterministic in our case), general service distributions (which are fairly complicated

in our problem), and changeover times. Consequently, it is necessary to use heuristic procedures to solve this problem.

As mentioned earlier, we view the planning problem as the first level in a hierarchical approach to the problem with two levels in the hierarchy. The goal of the planning problem is to determine a fixed cycle duration, T, during which all products will be produced once. Our reason for focusing on the cycle duration is that it represents a simple yet logical target. One of the primary reasons why capacity-constrained systems with uncertainty in the yields (or demands) become unstable is too frequent setups resulting from short-term decisions being overly responsive to current inventory levels--that is, attempting to avoid shortages at all costs. When setups consume productive capacity, adherence to the planned cycle duration helps to ensure that there is sufficient remaining capacity to meet demand over the long term. It also helps in planning for procurement of raw materials.

The control problem would take this value of T as given and determine input quantities in view of demand data over the short term and the current inventory position. In similar settings with random demand, it has been suggested (Leachman and Gascon 1989) that when the system is falling behind schedule, lengthening the cycle to spread out the setups, thereby freeing up some production time, might be desirable. While it might be possible (although not easy) to adapt their control policy to the case of random yields, their approach to controlling the problem also requires a target cycle length. In this paper, we develop one way to determine a good target cycle.

Note that since the yield rates are random, we must distinguish between input and output quantities. We can make decisions about the input quantities but the inventory and shortage costs are determined by the random output quantities.

2.1 Initial Formulation

The objective in this problem is to find a production cycle which minimizes the expected cost per unit time subject to a capacity constraint. In doing so, we also need to decide what fraction of the cycle, on average, to allocate to production of each product. In this subsection, we first discuss some modeling issues, and then propose two different formulations of the problem. Throughout this section we use the following notation:

S_i = setup cost for product i

π_i = shortage cost per unit per unit time for product i

D_i = annual demand for product i

p_i = actual yield rate for product i (random)

$\bar{p}_i$ = average yield rate for product i (for notational simplicity)

K_i = production rate for product i

τ_i = setup time for product i

Q_i = input quantity for product i

T = cycle length

$f_i(\cdot)$ = density of yield rate for product i

$F_i(\cdot)$ = distribution of yield rate for product i

The primary difficulty in formulating this problem is that, regardless of the particular planning or control policy, the system regenerates (returns to exactly the same state) very infrequently. Thus, it is difficult to write an analytically tractable, exact expression for the expected cost per unit time. Even if we ignore the fact that there may be many production cycles between regenerations and treat the production cycles as if they were statistically identical, the analysis is complicated by the interaction between the production policy and safety stock levels (i.e., the average on-hand inventory immediately before production of a product begins). In the following, we develop an approximation of the safety stock level under the assumptions that (i) the starting times of a product's production runs are equally spaced in time, (ii) net demand for each product during a cycle is always non-negative (otherwise a pure rotation cycle is unrealistic), and (iii) the production policy has a particular form. The first condition is never true, but since we are only interested in average inventory levels (and not their distributions), this should not create a serious error in the approximation. We later show that (i) under mild conditions, the second condition is satisfied with a high probability, and (ii) the structure of the optimal production policy (under the assumptions of our model) has the assumed form of the production policy.

We assume that the production policy is

$$Q_i = a_i (DT - I_i) \tag{1}$$

where a_i is a constant and I_i is the inventory of product i when its production run begins. In the following, we will treat a single item, and will therefore drop the product index subscript. Instead, the subscript will denote the cycle index.

Let X_n = inventory at end of cycle n,

p_n = observed yield rate in cycle n, and

Q_n = production input quantity in cycle n

= $a (DT - X_{n-1})$,

and assume that $X_0 = 0$. Note that we have expressed Q_n as $a(DT - X_{n-1})$, not $a(DT - X_{n-1})^+$ because of our assumption that net demand is non-negative.

Lemma 1: $X_n = (1 - ap_n)(X_{n-1} - DT)$.

Proof: Since $X_0 = 0$, we have that

$Q_1 = a (DT - 0) = aDT$, and

$X_1 = X_0 + Q_1 p_1 - DT = (1 - ap_1)(X_0 - DT)$.

So X_1 has the indicated form. Let us assume that this holds for $X_2,...,X_{n-1}$ (inductive hypothesis). We then have

$Q_n = a (DT - X_{n-1})$ and

$X_n = X_{n-1} + ap_n(DT - X_{n-1}) - DT$

$= (1 - ap_n) (X_{n-1} - DT)$.　　　　　　　　Q.E.D.

Let us now expand X_n by recursive substitution. This gives

$X_n = -[(1-ap_n)DT + (1-ap_n)(1-ap_{n-1})DT + ...$

$+ (1-ap_n)(1-ap_{n-1})\cdots(1-ap_1)DT]$

$+ (1-ap_n)(1-ap_{n-1})\cdots(1-ap_1)X_0$.

Noting that the last term is zero because $X_0 = 0$ and that the p_ns are i.i.d., we have

$$E(X_n) = -DT \sum_{k=1}^{n} (1-a\bar{p})^k.$$

Taking the limit as n -> ∞, we have

$$\lim_{n \to \infty} E(X_n) = DT (a\bar{p}-1)/(2-a\bar{p}) \tag{2}$$

if $|1 - a\bar{p}| < 1$. Note that $a = \bar{p}^{-1}$ means that the production input quantity is adjusted to consider the average yield rate. Thus, since $a > 0$ and $\bar{p} > 0$, the condition $|1 - a\bar{p}| < 1$ is equivalent to $a < 2\bar{p}^{-1}$. This means that the production input quantity is less than *twice* the yield adjusted amount. Although this imposes another condition on the approximation, this would not be constraining in most applications. However, this condition, in conjunction with the assumption of non-negative demand ($X_n < DT$), places an additional constraint on $a\bar{p}$. Note that for $a\bar{p} << 1.5$, $(a\bar{p}-1)/(2-a\bar{p}) << 1$, so it is unlikely that net demand would be negative. On the other hand, for $a\bar{p} \geq 1.5$, $\lim_{n \to \infty} E(X_n) > DT$, so there is a significant chance of negative net demand. For the moment, we will assume that $a\bar{p}$ is sufficiently small that this is not problematic.

Using the production policy defined by (1) and safety stock quantities as estimated by (2), we have the following formulation:

$$
\begin{aligned}
\text{(P1)} \quad \text{Minimize} \ \sum_i \{ S_i &+ [0.5 h_i D_i T^2/(2-a_i\bar{p}_i)^2] \int_0^{a_i^{-1}} [a_i\bar{p}_i - 1 + a_i p_i(3-2a\bar{p}_i)]^2 f_i(p_i) dp_i \\
&+ [h_i D_i T^2/(2-a_i\bar{p}_i)] \int_{a_i^{-1}}^{1} [1.5 a_i\bar{p}_i - 2 + a_i p_i(3-2a_i\bar{p}_i)] f_i(p_i) dp_i \\
&+ [0.5 \pi_i D_i T^2/(2-a_i\bar{p}_i)^2] \int_0^{a_i^{-1}} [(1-a_i p_i)(3-2a_i\bar{p}_i)]^2 f_i(p_i) dp_i \} / T
\end{aligned}
$$

(3)

$$
\text{subject to} \quad \sum_i [a_i D_i T(3-2a_i\bar{p}_i)/K_i(2-a_i\bar{p}_i)] \leq T
$$

$$
a_i \geq 0, \quad \forall i
$$

$$
T \geq 0
$$

Within the brackets in the objective function is the cost per cycle for product i. The terms represent (i) setup cost, (ii) expected inventory holding costs when a shortage situation occurs, (iii) expected inventory holding costs when a shortage situation does not occur, and (iv) expected shortage costs. The first constraint ensures

that capacity is not exceeded, while the others ensure non-negativity of the decision variables. Derivations of the integral expressions in the objective function and the first constraint appear in Appendix 2.

For simplicity, cycle stock is represented as if production occurs instantaneously. We could have incorporated finite production rates into the cycle stock calculations, but because the yield rates are random, so are the production rates (i.e., the rate of output of good units). The resulting model (not presented here) is extremely complex. Incorporating the effect of average yield rates and finite production rates is straightforward, but unnecessarily complicates the algebra, so we chose not to do so for clarity of presentation. Even so, the infinite production rate approximation is fairly accurate when production of each part occurs only during a small portion of the year.

It is useful to point out that in this formulation, shortage costs are charged on average backorders outstanding to reflect the fact that longer cycles lead to longer stockout durations. Because variable production costs are not included in the model (as discussed above), these shortage costs represent the cost of delayed revenue and loss of customer goodwill per unit time. In addition, inventory holding costs are charged on average inventory in order to reflect the fact that the cycle duration does influence time-weighted inventory levels.

Note that the Q_is do not appear at all in this formulation. Thus, one only needs to find the a_i values, proving our conjecture about the form of the production policy under the assumptions discussed above.

Problem P1 is a non-convex optimization problem with nonlinear constraints. It is also algebraically very complex. Although it might be possible to solve such problems numerically, it is unlikely that we would be able to characterize optimal solutions or develop insights into good policies. For this reason, we analyze an even simpler, but significantly more tractable version of the problem which lends itself to obtaining useful insights. We later explain how the simplified problem can be adapted, in an iterative way, to incorporate some of the effects discussed above.

2.2 Simplified Formulation

The simplified version of the problem is based on the assumption that the initial inventory of each product at the beginning of its production run is zero. Because shortage penalties normally exceed inventory holding costs, we would expect the optimal production quantities to exceed the yield adjusted quantities (i.e., $Q_i^* > D_i T / \bar{p}_i$). Thus, on the average, we would expect to have positive initial inventory. As a consequence,

less capacity would be needed than would be indicated by our solution. It would be possible to have a shorter production cycle (T), or alternatively more flexibility (idle time) within the given cycle. Thus, the solutions for the simplified formulation can be viewed as conservative. Since we are treating a static version of the problem here, this additional flexibility may be useful, or even necessary, in the true dynamic environment. Consequently, it is likely that the solutions to P2 are more realistic than the solutions to P1. Nevertheless, for the sake of consistency, we later explain how the average initial inventory levels can be incorporated into the solution procedure.

With this simplification, the problem becomes:

$$(P2) \quad \text{Minimize} \sum_i [S_i + (h_i/2D_i) \int_0^{D_iT/Q_i} (Q_ip_i)^2 f_i(p_i) dp_i$$

$$+ h_iT \int_{D_iT/Q_i}^{1} (Q_ip_i - 0.5D_iT) f_i(p_i) dp_i$$

$$+ (\pi_i/2D_i) \int_0^{D_iT/Q_i} (D_iT - Q_ip_i)^2 f_i(p_i) dp_i]/T \quad (4)$$

$$\text{subject to} \quad \sum_i (Q_i/K_i + \tau_i)/T \leq 1$$

$$Q_i \geq 0, \forall i$$

$$T \geq 0$$

Within the brackets in the objective function is the cost per cycle for product i. The terms represent (i) setup cost, (ii) expected inventory holding costs when a shortage situation occurs, (iii) expected inventory holding costs when a shortage situation does not occur, and (iv) expected shortage costs. The first constraint ensures that capacity is not exceeded, while the others ensure non-negativity of the decision variables. Details of the derivation of the objective function appear in Appendix 2. Note that the objective function is not jointly convex in T and the Q_is, and therefore cannot be solved by standard techniques.

2.3 Solution Procedure

In this subsection, we propose an iterative procedure based on Lagrangian relaxation to solve this problem. We first establish a series of results on which the procedure is based.

Relaxing the capacity constraint and rewriting (4) we obtain:

$$\text{Minimize } L(Q,T,\lambda) = \{\sum_i [S_i + (h_i T \bar{p}_i) Q_i - 0.5 h_i D_i T^2$$

$$+ 0.5(\pi_i + h_i)(Q_i^2/D_i) \int_0^{D_i T/Q_i} p_i^2 f_i(p_i) dp_i$$

$$- (\pi_i + h_i) T Q_i \int_0^{D_i T/Q_i} p_i f_i(p_i) dp_i$$

$$+ 0.5(\pi_i + h_i) D_i T^2 F_i(D_i T/Q_i)]\} / T$$

$$+ \lambda [\sum_i (Q_i/K_i + \tau_i)/T - 1] \tag{5}$$

We can now obtain the Kuhn-Tucker conditions:

$$\partial L(Q,T,\lambda)/\partial Q_i = [h_i T \bar{p}_i + (\pi_i + h_i)(Q_i/D_i) \int_0^{D_i T/Q_i} p_i^2 f_i(p_i) dp_i$$

$$- (\pi_i + h_i) T \int_0^{D_i T/Q_i} p_i f_i(p_i) dp_i + \lambda/K_i]/T = 0 \tag{6}$$

$$\partial L(Q,T,\lambda)/\partial T = \{ \sum_i [h_i \bar{p}_i Q_i - h_i D_i T - (\pi_i + h_i) Q_i \int_0^{D_i T/Q_i} p_i f_i(p_i) dp_i$$

$$+ (\pi_i + h_i) D_i T/Q_i] - (L+\lambda)\}/T = 0 \tag{7}$$

$$\partial L(Q,T,\lambda)/\partial \lambda = \sum_i (Q_i/K_i + \tau_i)/T - 1 = 0 \tag{8}$$

If T and λ were given, the Q_i's could be obtained directly from (6). This is relatively easy to do if the two integral expressions are tabulated in advance for a range of values of the upper limit (which is constrained to be less than or equal to 1). Also, as we show later in this section, $\partial^2 L(\mathbf{Q},T,\lambda)/\partial Q_i^2 > 0$ (i.e., the first derivative is increasing) for all non-negative λ, so even simple search procedures can be used. Furthermore, with T and λ fixed, the problem is separable by product. Unfortunately, T and λ are fairly difficult to find from (7) and (8). What typically is done for problems of this type is to solve the optimization problem for several values of T and to choose the T giving the minimum cost. It is not obvious, however, that the resulting minimum cost function is convex in T. We discuss this point in more detail later in this section.

To simplify the solution process, we can rewrite $\partial L(\mathbf{Q},T,\lambda)/\partial Q_i = 0$ as

$$h_i \bar{p}_i + \lambda/K_i T + (\pi_i + h_i)\ [Q_i/(D_i T)]\ \int_0^{D_i T/Q_i} p_i^2 f_i(p_i) dp_i$$

$$-\ (\pi_i + h_i)\ \int_0^{D_i T/Q_i} p_i f_i(p_i) dp_i = 0 \tag{9}$$

Observe that for each T and λ we only need to find a value of ratio D_i/Q_i, which can then be used to specify a Q_i for any possible D_i value. Before doing so, however, we must ensure that L is convex in Q_i. We have

$$\partial^2 L(\mathbf{Q},T,\lambda)/\partial Q_i^2 = [(\pi_i + h_i)/D_i T]\ \int_0^{D_i T/Q_i} p_i^2 f_i(p_i) dp_i$$

which is non-negative for any finite Q_i. Thus $L(\mathbf{Q},T,\lambda)$ is convex in Q_i.

Now $\partial^2 L(\mathbf{Q},T,\lambda)/\partial Q_i \partial \lambda = 1/K_i T > 0$, so as λ increases, the Q_is must decrease, as expected. In other words, $\partial L(\mathbf{Q},T,\lambda)/\partial Q_i$ is increasing with λ, and since we need to have $\partial L(\mathbf{Q},T,\lambda)/\partial Q_i$ equal to zero for all i, we must modify the Q_is as λ increases. Since $\partial^2 L(\mathbf{Q},T,\lambda)/\partial Q_i^2 \geq 0$, this is accomplished by decreasing the Q_is. Because of these monotonicity properties, for fixed T, it should be relatively easy to find the appropriate value of λ so that the capacity constraint (if binding) is satisfied at equality. Finding the optimal value of T is somewhat more difficult.

It is well known that the total cost is convex in T for deterministic problem. Thus, if the unconstrained T^* is not capacity-feasible, then the optimal decision is to

make T* as small as possible. In the deterministic problem, the relationship between Q_i and T is linear (i.e., $Q_i = D_iT$). However, in the problem with random yields, the relationship between Q_i and T is not necessarily linear. In fact,

$$\partial^2 L(\mathbf{Q},T,\lambda)/\partial Q_i\partial T = - \{(\pi_i + h_i)(Q_i/D_i) \int_0^{D_iT/Q_i} p_i^2 f_i(p_i)dp_i + \lambda/K_i\}/T^2 < 0.$$

Thus, while Q_i increases with T, the relationship is not linear, and clearly depends upon both the yield rate distribution and the production rate for item i.

In Appendix 3, we show that for fixed λ, L is unimodal in T at the optimal values of Q_i which are consistent with the values of T. Thus, one can find the optimal value of T given λ. Using such an approach would require a hierarchical search procedure in which one searches for λ at the highest level, and searches for the optimal value of T (and associated Q_i) for a given λ. The values of Q_i are needed to check for feasibility. An alternate method, which is much more efficient, is described below.

The procedure uses a form of $\partial L(\mathbf{Q},T,\lambda)/\partial Q_i$ in which the T in the denominator of (6) is retained and we define $\beta_i = D_iT/Q_i$. This first order condition becomes

$$h_i\overline{p}_i + (\pi_i + h_i)\beta_i^{-1} \int_0^{\beta_i} p_i^2 f_i(p_i)dp_i$$

$$-(\pi_i + h_i) \int_0^{\beta_i} p_i f_i(p_i)dp_i + \lambda/K_iT = 0. \tag{10}$$

which is obtained by direct substitution of β_i for D_iT/Q_i. Here β_i is the ratio of the demand during the cycle to the production quantity. Note that using β_i does not constrain the solution in any way, since one is essentially using it to choose Q_i for a given T.

Suppose that for a moment we fix the ratio λ/T, and let $g_i(\beta_i)$ be the first three terms of (10). It is easily shown that $g_i(\beta_i)$ is monotonically non-increasing with β_i. Thus, it is easy to find the appropriate β_i values for each λ/T ratio. The question now is what the optimal value of T should be.

Let $G(\beta,\lambda,T)$ represent L constrained to the values of $Q_i^*(T)$, where $Q_i^*(T)$ is the optimal value of Q_i given T (i.e., $Q_i^* = D_iT/\beta_i$). Then

$$G(\beta,T,\lambda) = \sum_i [S_i - h_i D_i T^2/2 - (\pi_i + h_i)(D_i T^2/2)\beta_i^{-2} \int_0^{\beta_i} p_i^2 f_i(p_i)dp_i$$

$$+ (\pi_i + h_i) D_i T^2 F_i(\beta_i)/2]/T$$

$$+ \lambda(\sum_i \tau_i/T - 1)$$

which is obtained by substituting $Q_i \cdot (\partial L(Q,T,\lambda)/\partial Q_i) = 0$ for all i (from (4)) into (3a), substituting β_i where appropriate, and simplifying. Now

$$\partial G(\beta,T,\lambda)/\partial T = \sum_i \{ - (S_i + \lambda\tau_i)/T^2$$

$$+ D_i [- h_i + (\pi_i + h_i)\beta_i^{-2} \int_0^{\beta_i} p_i^2 f_i(p_i)dp_i$$

$$+ (\pi_i + h_i) F_i(\beta_i)]/2\}.$$

Let $g_i(\beta_i)$ represent the expression in brackets. Then setting this first derivative to zero and solving for λ, we get

$$\lambda = [T^2\sum_i D_i g_i(\beta_i) - 2 \sum_i S_i]/[2 \sum_i \tau_i]. \tag{11}$$

We propose an iterative procedure in which one starts with a feasible λ/T ratio (e.g., zero). One then sequentially uses equation (10) to find the β_i values, equation (8) to find a new value of T, and equation (11) to find a new value of λ. With the new λ/T ratio, the process repeats until convergence is attained. Although this procedure bears some resemblance to standard fixed point methods, there are two limitations in implementing this iterative procedure exactly. In computing T and λ, it is possible to obtain negative values, which are infeasible. Thus, in our actual implementation, we take the following precautionary steps.

1. If T is negative, it is the result of capacity utilization exceeding 100%. Thus, the values of β_i must be increased, which can be accomplished by increasing λ/T. We arbitrarily double the λ/T ratio until a feasible solution is found. This provides a new starting point in the iterative procedure.

2. After finding a positive T in step 1 above, if λ is negative (making λ/T negative), we simply set λ/T equal to its absolute value to give a new, feasible starting point for the iterative procedure.

This procedure may appear to be *ad hoc*, but it simply represents one possible procedure to search for λ/T. Another possible procedure is a bisection search, but it may be difficult to identify an appropriate upper bound *a priori*. On the other hand, the bisection search is guaranteed to converge. It is important to point out that we only need to obtain values of λ/T, T, and the β_is to satisfy (8), (10, and (11) simultaneously, or if the capacity constraint is not binding, find T and the β_is to satisfy (10) and (11) with $\lambda = 0$. Any reasonable way to do this will find the optimal solution. The main point is that searching for λ/T rather than λ is much easier and much more efficient.

Recall that this procedure is based on the assumption of zero initial inventory. In reality, we would expect initial inventory, on the average, to be positive. Given any set of β_is, we can determine an *adjusted* net demand rate for product i by dividing the actual demand rate by $\overline{p_i}/\beta_i$. We could then re-solve the problem, find a new set of β_is, and iterate until the β_is stabilize. In this way, we could heuristically account for the effects of safety stock induced by the production policy. Although it is not obvious that such a procedure would always converge, intuition suggests that it should. If the initial β_is are too small (i.e., too much safety stock), the resulting net demand will be small, reducing the amount of "risk" in the system, since much of the demand will be satisfied by on-hand inventory. The next iteration will adjust to the decreased uncertainty by increasing the β_is. Thus, the feedback appears to drive subsequent solutions in the right direction.

4.0 A POSSIBLE APPLICATION

In this section, we describe one facility, among many, that provides motivation for this research. The facility is a painting operation which tries to use a pure rotation policy. The production process within this facility includes three major steps. The first step produces parts in a batch fabrication process. These parts are then painted by grouping parts which need to be painted the same color. Average yield rates vary considerably by color, and there are approximately one dozen colors. The painted parts are sent to an assembly area where work proceeds at a fairly constant pace. Thus, the aggregate "demand" on the paint facility is also fairly constant, although the mix of

colors fluctuates from day to day. The weekly demand for each color remains fairly stable.

The painting facility is highly automated. Parts are placed on carriers which ride on a conveyor through the system at a constant rate. Large parts require more spacing between carriers than small parts. Changeovers (setups) between colors take only a few minutes. However, since the production rate of the system is high, many units can be painted in the time required for a setup. Moreover, the paint is expensive and the entire system must be purged of paint and cleaned with solvent each time the color is changed.

The facility operates continuously except for preventive maintenance and repairs. Thus, there is no opportunity for additional production on overtime.

The system has very good average yield rates, but many different factors (e.g., temperature, humidity, as well as equipment calibrations) affect the acceptability of the finished product. Thus, yield rates (i.e., the ratio of acceptable output to input) are highly variable and this complicates the scheduling problem.

The parts are 100% inspected upon completion of the painting process, Unacceptable items are reworked when possible. The rework process usually involves sanding and complete repainting, so the input material is saved, but the cost and time associated with painting is otherwise the same as for an unpainted part.

There are some fairly tight storage capacity constraints which essentially require that production track demand fairly closely. Unfortunately, since painted and unpainted parts share the same storage area, it is difficult to specify an exact storage capacity constraint for the painted parts. (Setups in the fabrication stage take approximately one hour, so it is customary to produce batches of several hundred or more). Management has already evaluated the setup cost-storage capacity constraint tradeoff, and has decided that a reasonable way to deal with the problem is to cycle through nearly all colors (with the exception of low volume colors) on a regular basis. Even within the limitations of the storage space constraint, there is some latitude in selecting the frequency with which to cycle through the colors.

The painting operation has a single "customer," which is the assembly area. The company recognizes the difficulties of yield variability in the paint area, and therefore specifies production targets for the paint area in advance. The goal of the paint area is to meet those production targets as closely, yet as economically as possible. After work is completed in the assembly area, the part is sent to another plant within the same company for assembly into the finished product. The ultimate user of the painted parts

is a fixed pace assembly line. Thus, the total demand for a part is relatively stable over time.

The situation at this facility differs from our problem assumptions in that demand is not truly constant, and consequently, it is not possible to ensure equal time between the start of production runs of a given part. Moreover, the storage capacity imposes additional constraints on the solution. Beyond this, however, the managerial policy of trying to cycle through all of the colors at regular intervals actually makes the problem very similar to the one described above.

The solution from the procedure described in section 3 could be used as the basis for setting planned cycle durations in this facility. It could also be used to evaluate the impact of different production policies (i.e., set of β_is) on inventory levels and service measures. We should note that the facility already uses yield adjustment factors. Thus, it would not be difficult to implement the solution from the model described above.

5.0 EXPERIMENTAL RESULTS

We were able to obtain _average_ yield rate data from the painting facility described earlier, but distributional information was not available. We decided to use this opportunity to investigate the effects of different yield rate distributions upon solutions to the problem. It was our hope that we would be able to obtain some insights into characteristics of optimal policies even if the yield rate distributions were not known precisely. It was well-known at this facility that colors with lower average yield rates also had the highest yield rate variances. Indeed, it was this combination of factors that made the lot-sizing decisions so difficult.

The products manufactured at the facility are similar with regard to cost of input materials and labor (at least up to and including the painting process). We estimated the ratio of the shortage cost to the inventory holding cost to be approximately 50 to 1. At the moment, the painting facility appears to be a bottleneck, so typically many other parts which, in aggregate, have considerable value, must wait for parts produced at this facility. One might think that such a high π/h ratio would lead to large adjustments for yield losses (i.e., small β_is). However, the tight capacity constraints preclude producing far ahead of demand. The ratio of the setup cost to the annual inventory holding cost for one unit was estimated to be 10 to 1.

Since the parts are similar in terms of both processing time and costs, production policies will tend to allocate more than a "fair share" of the capacity to colors having high yield rates. Such policies cannot continue indefinitely, however, since

backlogs of parts requiring low-yield-rate colors will eventually accumulate, forcing the system to produce those parts. Thus, the system will achieve equilibrium production levels over the long term, although short term policies will fluctuate. Nevertheless, in real systems, there is always a spectrum of parts, from easy and profitable to produce, to difficult and not very profitable to produce. The cost characteristics described above reflect such a system.

For the purposes of illustration we performed a single iteration of the procedure described in section 3. As the results will show, it appears that only one iteration is necessary. We explain the reasons for this in more detail later. We present results for a hypothetical four-product system which bears some similarity to the real system.

We chose to use uniform distributions for the yield rates because of their simplicity, and because they permit a great deal of flexibility in selecting means and variances without introducing complicating factors such as skewness, positive probabilities of negative yield rates, etc. The products have average yield rates of 0.75, 0.80, 0.85, and 0.90, respectively (with a maximum of 1.0 in all cases), and average daily demands of 100 apiece. The capacity of the system was varied from 500 to 750 units per day to permit us to examine the effect of capacity tightness on the solutions. Setup times are sequence independent and each consumes on the order of .005 of one day's available production time. Setting $h_i = 1$ for all i, we used the relative costs mentioned above rather than absolute costs because of confidentiality considerations.

The results for this set of problems (shown in Table 1) were quite interesting. First, we found that there appeared to be three distinctly different policies depending upon capacity tightness. When the capacity constraint was not binding, the β_i values were less than the average yield rates, indicating some provision of "safety input." There was a greater relative provision for the items with larger relative yield rate variances. That is, the item with the the smallest average yield and the largest yield rate variance had the smallest value of $F_i(\beta_i)$.

Table 1 here

Secondly, we found that when capacity was relatively tight, the optimal policy was to set β_i equal to the average yield rate for all i, and to adjust T to handle the capacity constraint. It is worthwhile to point out that the former occurred in spite of the very high shortage cost to holding cost ratio. Finally, in an intermediate range, where the capacity constraint was binding, but utilization of the system with the unconstrained solution was only slightly greater than 1.0, there was a tradeoff between setup costs and

the sum of inventory holding and shortage costs. The intermediate range appeared to be very small, however.

We also investigated the effect of the variance of the yield rate distribution by using slightly different problem parameters. We again used uniform yield rate distributions. The average yield rate is 0.75 for all parts, but the ranges are 0.5, 0.4, 0.3, and 0.2, respectively. The capacity level was varied from 550 to 800 units per day. All other parameters and costs are the same as above.

We found quite similar patterns in the results, although the range of the $F_i(\beta_i)$ values (when the capacity constraint was not binding) was slightly smaller (see Table 2). It appears that the differences among the $F_i(\beta_i)$ values are due principally to the differences of the yield rate variances among parts, and not to differences in the average yield rates. This conjecture is supported by experimental results (see Table 3) for problems with parts having the same yield rate variances but different average yield rates . Here, the ranges of the $F_i(\beta_i)$ values were fairly small, although the β_is were still smaller for the parts with lower average yield rates. Thus, it appears that the yield rate variance is a major contributor to relative "safety input" quantities. It is important to point out that in this third set of problems, on the first iteration, the policy was to set β_i equal to the lower limit of the yield rate support when the capacity is fairly tight. Thus, in this case, there would be no shortages, but inventory holding costs will be quite high. Recall, however that this represents results for the first iteration only.

Tables 2 and 3 here

Note that in the first two sets of results, it would generally be necessary to perform only one iteration of the procedure. When capacity constraints are loose (i.e., when the production rate is high in our examples), the β_i values remain unchanged as the capacity constraints become looser. Thus, when the demand rates are adjusted downward to account for initial inventory, the β_i values remain the same. (Decreasing demand has the same qualitative effect as increasing the production rate. This is intuitively clear, but we obtained empirical evidence of this as well.) Therefore, the initial β_i values will reflect the true net demand, and only the value of T must be modified to reflect the impact of initial inventory. Likewise, when capacity is sufficiently tight, the β_is remain the same since the net demands appear not to require any adjustment at all. It is only in the relatively small "intermediate" range of capacity tightness where iteration might be necessary.

In the third set of problems, it would be necessary to perform additional iterations, since the adjusted demand rates would be much lower than the original values, especially under the high production rate scenarios. This would tend to move the β_is closer to the midpoints of their respective ranges. This is, indeed, what happens on the second iteration. For the production rates 600 and 640, the β_is are now equal to the respective mean yield rates. For all higher production rates, the β_is are the same as for the higher production rate cases in Table 3. For the higher production rates, the solution converges on the third iteration. For the low production rates, the solution oscillates between the solutions obtained in iterations 1 and 2. Thus, manual intervention may be needed to find an intermediate degree of demand adjustment and consistent β_is, but this would not be difficult.

It is also useful to point out that, at least in the first two problem scenarios, a significant amount of capacity is required (despite the small setup time) before the solution moves away from the simple policy of adjusting for the average yield rates. Such a simple policy could lead to very high shortage costs. It appears that, in order to avoid these high shortage costs, the solution has a short cycle duration, which in turn, reduces the average *shortage duration*. In this way, shortage costs are controlled, but at the expense of more frequent stockout occasions. This suggests that when yields are random, it is important to plan the capacity of the system with a sufficient amount of slack to avoid being forced into a rather undesirable operating policy.

6.0 SUMMARY AND DISCUSSION

We have investigated a version of the product cycling problem in which the fraction of parts that are acceptable is characterized by a yield rate distribution. The major simplifying assumption is that inventory of each product is zero when its production run begins. A procedure with a fixed-point flavor was developed to determine the cycle duration and the associated production policies (yield loss adjustment factors). An iterative scheme was suggested to heuristically incorporate the impact of the yield adjustment factors on average initial inventory. In numerical experiments, however, it was found that in most situations only a single iteration is needed.

Here, it was assumed that the yield adjustment factors and cycle durations, once selected, could not be revised. Further work is needed to understand how to make these decisions dynamically, and in the presence of uncertain demand. Additional work is also

needed to investigate how these yield adjustment factors would perform, either alone, or with a modification to consider a target cycle length, in a dynamic setting.

Acknowledgment

References

Bassok, Y. and R. Akella (1988), "Combined Component Ordering and Production Decisions in Flexible Electronic Assembly Systems with Supply Quality and Demand Uncertainty," Working paper, Graduate School of Industrial Administration, Carnegie Mellon University.

Elmaghraby, S. (1978), "The Economic Lot Scheduling Problem (ELSP): Review and Extensions," *Management Science* 24 (6), 587-598.

Gerchak, Y., R.G. Vickson, and M. Parlar (1989), "Periodic Review Production Models with Variable Yield and Uncertain Demand," *IIE Transactions* 20 (2), 144-150.

Hodgson, T. and H.L.W. Nuttle (1987), "A Note on Linear Programming and the Single Machine Lot Size Scheduling Problem," *Inter. J. Prod. Res.* 24 (4), 939-943.

Hsu, W. (1983), "On the General Feasibility Test of Scheduling Lot Sizes for Several Products on One Machine," *Management Science* 29 (1), 93-105.

Jones, P. C. and R.R. Inman (1989), "When is the Economic Lot Scheduling Problem Easy?" *IIE Transactions* 21 (1), 11-20.

Leachman, R.C and A. Gascon (1988), "A Heuristic Scheduling Policy for Multi-Item, Single-Machine Production Systems with Time-Varying, Stochastic Demands," *Management Science* 34 (3), 377-390.

Karlin, S. (1958), "Steady State Solutions," in Studies in the Mathematical Theory of Inventory and Production, K.J. Arrow. S. Karlin, and H.E. Scarf, Eds., Stanford University Press, Stanford, CA.

Lawler, E.L., J.K. Lenstra, A.H.G. Rinooy Kan, and B.B. Shmoys (1985), The Traveling Salesman Probem, John Wiley & Sons, New York.

Maxwell, W.L. and H. Singh (1983), "The Effect of Restricting Cycle Times in the Economic Lot Scheduling Problem," *IIE Transactions* 15 (3), 235-241.

Mazzola, J.B., W.F. McCoy, and H.M. Wagner (1987), "Algorithms and Heuristics for Variable-Yield Lot Sizing," *Naval Research Logistics* 34 (1), 67-86.

Sepehri, M., E.A. Silver, and C. New (1986), "A Heuristic for Multiple Lot Sizing for and Order under Variable Yield," IIE Transactions 18 (1), 63-69.

Silver, E.A. (1976), "Establishing the Order Quantity when the Amount Received is Uncertain," INFOR 14 (1), 32-39.

Takagi, H. (1989), "Queueing Analysis of Polling Models," IBM Computer Science Research Report RT 0021, November 15, 1989.

Weber, R. and G. Weiss (1989), "On an Index Policy for Restless Bandits," Working Paper, School of Industrial and Systems Engineering, Georgia Institute of Technology.

Yano, C. A. and H.L. Lee (1989), "Lot Sizing with Random Yields: A Review," Technical Report 89-16, Department of Industrial and Operations Engineering, University of Michigan.

Table 1

Example Results

Problem Data:

$D_i = 100$ for $i = 1,...,4$

$P_1 \sim U[0.8,1.0]$
$P_2 \sim U[0.7,1.0]$
$P_3 \sim U[0.6,1.0]$
$P_4 \sim U[0.5,1.0]$

Minimum daily capacity required: 487.1 (not considering setups)

Results:

Production Rates (same for all products)	β	$F(\beta)$
760,800	(0.887,0.804,0.719,0.630)	(0.436,0.347,0.297,0.260)
720	(0.899,0.849,0.799,0.749)	(0.495,0.497,0.497,0.498)
560,600,640,680	(0.9,0.85,0.8,0.75)	(0.5,0.5,0.5,0.5)

Table 2

Effect of Different Yield Rate Variances and Capacity Tightness

Problem Data:

$D_i = 100$ for $i = 1,...,4$

$P_1 \sim U[0.5,1.0]$
$P_2 \sim U[0.55,0.95]$
$P_3 \sim U[0.6,0.9]$
$P_4 \sim U[0.65,0.85]$

Minimum daily capacity required: 533.3 (not considering setups)

Results:

Production Rates (same for all products)	β	$F(\beta)$
850,900,950,1000	(0.630,0.665,0.698,0.729)	(0.260,0.287,0.328,0.395)
800	(0.663,0.693,0.722,0.748)	(0.326,0.358,0.407,0.490)
600,650,700,750	(0.729,0.729,0.729,0.729)*	(0.458,0.448,0.430,0.395)

*The difference between these values and the average yield rate (0.75) may be due to rounding error and minor numerical instability when λ is large.

Table 3

Effect of Different Yield Rate Means
and Capacity Tightness
(First Iteration)

Problem Data:

$D_i = 100$ for $i = 1,...,4$

$P_1 \sim U[0.8,1.0]$
$P_2 \sim U[0.7,0.9]$
$P_3 \sim U[0.6,0.8]$
$P_4 \sim U[0.5,0.7]$

Minimum daily capacity required: 545.6 (not considering setups)

Results:

Production Rates (same for all products)	β	$F(\beta)$
800,840,880	(0.887,0.782,0.678,0.571)	(0.436,0.409,0.389,0.356)
600,640,680, 720,760	(0.8,0.7,0.6,0.5)	(0,0,0,0)

Appendix 1

Difference between Random Yield and Random Demand Problems

The key difference between random demand and random yield is that in the former case, production input is equal to production output (except perhaps for a constant multiplicative conversion factor), whereas in the latter case, it is not. Thus, when yields are random, any mathematical representations of production, inventory, and shortage quantities must distinguish between input and output quantities.

Even if demand is deterministic, it is not always possible to convert a problem with random yields into an equivalent problem with random demand. Suppose that we try to do this. A reasonable approach would be to define the distribution of the hypothetical random demand as the distribution of the quantity that would have to be input in order to satisfy demand. If the yield uncertainty is expressed in the form of a probability mass function, the conversion is straightforward, but it is not as simple if the yield uncertainty is expressed in another form, or if the demand itself is random.

Having completed this conversion, we now need to relate these hypothetical demand quantities to the consequent actual inventory or shortage quantities for every possible input quantity. In some cases, the mapping is not one-to-one, making it impossible to define an equivalent problem. For example, suppose there is the possibility of a zero yield outcome. Then there is a positive probability of an infinite hypothetical demand. What is the resulting shortage quantity? That information is now lost in the conversion. Of course, we could recover this information, but this would not be necessary if the two problems were actually equivalent. How do we represent the fact that even though the hypothetical demand is ostensibly infinite, the actual shortage is finite? It is not clear that this can be done. Let us consider another example. Suppose there are several input quantity-yield outcome pairs that give the same hypothetical demand quantity. The conversion process would aggregate information, which again would need to be recovered in order to compute actual inventory and shortage quantities.

In conclusion, random yield problems cannot be converted to random demand problems without a significant transformation of the objective function and constraints to reflect the disaggregation process described above. The resulting formulations would be more complex than the original fomulations.

Derivation of Terms in the Formulations

__P1__

We derive the second, third, and fourth terms of the objective function in (P1), and the first constraint.

Second term:

Starting with the safety stock level of $D_iT(a_i\bar{p}_i-1)/(2-a_i\bar{p}_i)$, we input $Q_i=a_i[D_iT-D_iT(a_i\bar{p}_i-1)/(2-a_i\bar{p}_i)]$ with a resulting output of $a_ip_i[D_iT-D_iT(a_i\bar{p}_i-1)/(2-a_i\bar{p}_i)]= a_ip_iD_iT(3-2a_i\bar{p}_i)/(2-a_i\bar{p}_i)$. Adding this to the safety stock quantity gives $D_iT[a_i\bar{p}_i-1+a_ip_i(3-2a_i\bar{p}_i)]/(2-a_i\bar{p}_i)$. This is less than the demand during the cycle if $p_i \leq a_i^{-1}$, which is the limit of integration. In this case, the average inventory during the cycle is one half the square of the maximum inventory level divided by D_iT. Multiplying by h_i gives the second term.

Third term:

The derivation follows that of the second term except that we now consider the case of $p_i \geq a_i^{-1}$. Here, inventory is non-negative throughout the entire cycle, so the average inventory level is the maximum level, $D_iT[a_i\bar{p}_i-1+a_ip_i(3-2a_i\bar{p}_i)]/(2-a_i\bar{p}_i)$, less $0.5D_iT$. A little algebra gives the expression in (P1).

Fourth term:

We again consider situations in which shortages occur. If $p_i < a_i^{-1}$, the shortage quantity is $D_iT - D_iT[a_i\bar{p}_i-1+a_ip_i(3-2a_i\bar{p}_i)]/(2-a_i\bar{p}_i)$, and the average time-weighted backorder level is one half the square of this value, all divided by D_iT. Some simplification leads to the expression in (P1).

First constraint:

The capacity constraint in general form is:

$$\sum_i (Q_i/K_i + \tau_i) \leq T.$$

Substituting for Q_i and simplifying leads to the expression in (P1).

P2

The derivation of the integral expressions in P2 are similar to those above. The only differences are that the quantities are expressed in terms of the Q_is rather than the a_is, and the initial inventory is assumed to be zero for all products.

In this appendix we show that for a fixed value of λ, $L(\mathbf{Q},T,\lambda)$ (see equation (5)) is unimodal in T at the optimal values of Q_i consistent with these values of T. From equation (9), a first order necessary condition for Q_i is:

$$h_i \bar{T} p_i + (\pi_i + h_i)(Q_i/D_i) \int_0^{D_iT/Q_i} p_i^2 f_i(p_i) dp_i + \lambda/K_i$$

$$= (\pi_i + h_i) T \int_0^{D_iT/Q_i} p_i f_i(p_i) dp_i$$

Substituting $Q_i \cdot \partial L(\mathbf{Q},T,\lambda)/\partial Q_i = 0$ into L for all i, we get,

$$G(\mathbf{Q}^*(T,\lambda),T,\lambda) = \sum_i [S_i - h_i D_i T^2/2 - (\pi_i + h_i)(Q_i^{*2}/2D_i) \int_0^{D_iT/Q_i^*} p_i^2 f_i(p_i) dp_i$$

$$+ (\pi_i + h_i) D_i T^2 F_i(D_iT/Q_i^*)/2]/T$$

$$+ \lambda(\sum_i \tau_i/T - 1)$$

where G is defined as L restricted to $\mathbf{Q}^*(T,\lambda)$ and $\mathbf{Q}^*(T,\lambda)$ denotes the vector of optimal values of Q_i given T and λ.

After some algebra we get:

$$\partial G(\mathbf{Q}^*(T,\lambda),T,\lambda)/\partial T = \sum_i \{ - (S_i + \lambda\tau_i)/T^2$$

$$+ D_i [-h_i + (\pi_i + h_i) (Q_i^*/D_iT)^2 \int_0^{D_iT/Q_i^*} p_i^2 f_i(p_i) dp_i$$

$$+ (\pi_i + h_i) F_i(D_iT/Q_i^*)]/2\}.$$

Also,

$$\partial^2 G(\mathbf{Q}^*(T,\lambda),T,\lambda)/\partial T^2 = (2/T^3) \sum_i [S_i + \lambda\tau_i$$

$$- (\pi_i + h_i)(Q_i^{*2}/2D_i) \int_0^{D_iT/Q_i^*} p_i^2 f_i(p_i)dp_i]$$

$$+ \sum_i (\pi_i + h_i)(D_i^2/Q_i^*)f_i(D_iT/Q_i^*) \qquad (A-1)$$

which will clearly be non-negative if the first term is non-negative, that is, if

$$\sum_i (\pi_i + h_i)(Q_i^{*2}/2D_i) \int_0^{D_iT/Q_i^*} p_i^2 f_i(p_i)dp_i < \sum_i [S_i + \lambda\tau_i] \qquad (A-2)$$

Observe that each term in the summation on the left hand side of the inequality is monotone nondecreasing in T, so the inequality will be satisfied for all T sufficiently small. Hence, in this domain, the function is strictly convex. We will show that the function is strictly increasing for all larger values of T.

Consider a value of T that does not satisfy (A-2). For this value of T we would have

$$\partial G(\mathbf{Q}^*(T,\lambda),T,\lambda)/\partial T > \sum_i \{0.5 \ D_i \ [(\pi_i + h_i) \ F_i \ (D_iT/Q_i^*) - h_i]\} \qquad (A-3)$$

which comes from appropriate substitution (of (A2) divided by T^2) into $\partial G(\mathbf{Q}^*(T,\lambda),T,\lambda)/\partial T$. Thus, $\partial G(\mathbf{Q}^*(T,\lambda),T,\lambda)/\partial T$ must be strictly positive (and T cannot be optimal) if the right hand side of (A-3) is non-negative. Let us consider the case where the right hand side of (A-3) is strictly negative. We would have

$$G(Q^*(T,\lambda),T,\lambda) = \sum_i [(S_i + \lambda\tau_i) - (\pi_i + h_i)(Q_i^{*2}/2D_i) \int_0^{D_iT/Q_i^*} p_i^2 f_i(p_i)dp_i]/T$$

$$+ \sum_i 0.5 \, D_iT[-h_i + (\pi_i + h_i) \, F_i \, (D_iT/Q_i^*)] - \lambda$$

$$< -\lambda \qquad\qquad\qquad\qquad (A-4)$$

The inequality follows from the facts that the first term is negative from the assumption that (A-2) is not satisfied and second term is negative by the assumption that the right hand side of (A-3) is strictly negative. Since costs must be non-negative, (A-4) cannot possibly be true.

Thus, for values of T where (A-2) is not satisfied (i.e., "large" values of T), it must be true that $\partial G(Q^*(T,\lambda),T,\lambda)/\partial T$ is strictly positive. Thus, since L is convex for "small" values of T and strictly increasing for "large" values of T, it is unimodal in T.