Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Subassembly Manufacturing Yield Problem with Multiple Production Runs

Purushottaman Nandakumar, Jeffrey L. Rummel,

A SUBASSEMBLY MANUFACTURING YIELD PROBLEM WITH MULTIPLE PRODUCTION RUNS

PURUSHOTTAMAN NANDAKUMAR and JEFFREY L. RUMMEL

University of Connecticut, Storrs, Connecticut

(Received November 1992; revisions received July 1993, March 1994, November 1995, July 1996; accepted July 1996.)

We consider the assembly of a product that needs to exceed a certain length. The subassemblies are manufactured with a planned length setting. The resulting lengths of the subassemblies are not known with certainty, and vary about the planned setting. After the subassemblies are manufactured, those that fall short are rerun. This process is repeated until all subassemblies exceed the required length. We build and solve a model, to specify the planned length at each run, that minimizes the expected cost of production.

Most manufacturing processes contain uncertainty in yields, or characteristics, of the output. There exist many approaches to manage that uncertainty, and perhaps the most prevalent are safety stock models. One class of safety stock models schedules production for the desired outcome, and separately held safety stocks are used to make up for any shortfalls. Optimizing such a system typically requires the determination of the safety stock level, how units are removed from safety stock to replace defective units, and when to place special orders to replenish safety stocks. Another class of models schedules batches that are in excess of the requirement. In some batches there will be extra good units that are placed into a safety stock; in other batches, units are removed from safety stock to bring the production batch up to the requirement. A fundamental characteristic of such systems is that production runs balance out in the long run.

In the problem setting addressed in this paper, a product consists of a number of subassemblies, and all must meet or exceed a certain characteristic. Further, the subassemblies are nonsubstitutable. Separate safety stocks of common parts, therefore, cannot be utilized to replace the deficient unit. Similarly, one subassembly that is lacking in the requirement cannot be compensated for by another that exceeds the requirement. The compensation for variability needs to be integral to each unit produced, not pooled across units or accommodated by long-run averaging. In this paper, we will look at length as the characteristic of interest.

An example of this type of problem occurs in the manufacturing of fiber-optic cable. Customer orders for fiber-optic cables specify the number of individual fibers, the coloring and configuration of those fibers within the cable, and the desired length. The number of fibers can vary from just a few to 200 or more. In cables with a large number of fibers, sets of fibers are wrapped in different color tubes inside the final cable. The color of each fiber and content of each tube is specified by the customer and is different depending on the use for the cable. Because of this variability, cables are generally made to order. Cable lengths can vary from several meters to over 10 kilometers.

The production planner must specify the length of the fibers to be drawn and colored at the beginning of the process that will yield a finished cable of the appropriate length. There are several stages of the process that introduce scrap losses and measurement uncertainties, but a fundamental problem that needs to be modeled and solved can be described very simply. Suppose the planner had a distribution function that, for a given planned length, would specify the probability that the fiber will satisfy the target length. Given this function and the customer specifications, how much extra safety length should be planned for each fiber?

Some observations about this question are easily seen. As the length of the order changes, so does the amount of safety length, since failure on a long cable is more expensive than failure on a short cable. In addition, even though the safety length would be specified with respect to each fiber, those fibers will be bound together to make a cable. All fibers for a particular cable order are processed at the same time, usually in parallel, so they are all available for assembly. We assume setup costs for a production run, and the magnitude of the cost will also affect the planned length. The planned length setting, therefore, depends on the length of the order, the number of fibers, and the setup cost of the production run.

The cost of making a cable increases, in direct costs, as we increase the planned length of the fibers. This increased length, however, decreases the probability that each fiber comes out too short and must be redone. Redoing a fiber not only involves the cost of material that is discarded, but it also involves the cost of setting up the process for another run. Any additional production runs may involve a different number of fibers (only the ones that failed), and we allow for a different planned length at this stage that reflects the new economic tradeoffs.

The remainder of the paper is organized as follows. Section 1 is a brief discussion of related literature. Section 2

Subject classifications: Production scheduling: assembly yields. Dynamic programming: applications. Production/scheduling: stochastic.
Area of review: Manufacturing Operations.

0030-364X/98/4603-S050 $05.00

describes the model, and Section 3 is an exposition of the properties of the problem. Section 4 describes the optimal solution of the model and evaluates two heuristic solution methods with a computational study.

1. LITERATURE REVIEW

Our model falls, in a general sense, in the class of dynamic decision problems in production rework settings. The bulk of the literature on yield and rework problems have tended to concentrate on the lot-sizing issue. The decision maker has little influence over the process that generates defects, and adjusts production decisions to account for the variability. Nevertheless, it is useful to discuss the lotsizing literature to get insights about the yield models that have been used.

The lotsizing literature is vast, and a comprehensive review can be found in Yano and Lee (1995). In the early 1950s and 1960s, a series of papers addressed multiple production run problems for custom-ordered job lots. Several of these papers, such as Bowman and Fetter (1961), Llewllyn (1959), Goode and Saltzmann (1961), Hillier (1963), and Wadsworth and Chang (1964), to name a few, considered the problem of setting reject allowances for production runs. Beja (1977) formulates a model for the batching problem that uses a similar structure to the one we use. Wein (1992), Grosfeld-Nir and Ronen (1993), and Anily (1995) are some examples of papers that extend the basic model into other settings. In these models, a known custom demand needs to be met. Production occurs in batches, with random yield, and the processing of each batch involves a setup cost. The reject allowance is an excess quantity, over and above net requirements, that is processed to compensate for scrap. Interestingly, most of these papers use the Bernoulli model to capture the yield process.

The yield processes in a majority of these models are assumed to be outside the control of the decision maker, with a few exceptions. Several papers such as Porteus (1986), Rosenblatt and Lee (1986), and Lee and Rosenblatt (1988) attempt to incorporate models of imperfect production processes into a basic economic order quantity (EOQ) framework. The decision maker's control over the process is typically indirect and through inspection/repair or lot-sizing policies. Process investments and accompanying yield improvements have been modeled by Cheng (1991) and Gerchak and Parlar (1990) in an EOQ framework. The problem of diversification among suppliers, which implies some degree of control over process yields, has been analyzed by Gerchak and Parlar (1990) in an EOQ framework, and by Ilan and Yadin (1985) as a multiple production run problem. Structural results have been difficult to come by in problems where the decision maker has direct control over process yields.

2. PROBLEM FORMULATION

In modeling the problem, we need to make assumptions about three basic aspects of the problem. First, we assume the costs of processing fibers is a function proportional to the length of the fiber. This is reasonably accurate, since the bulk of the variable fiber cost is the direct material cost. Second, we assume that the economics of running an additional batch are captured by a standard setup cost model. The setup cost representation is fairly powerful, in that it can be used to model opportunity costs on the machine. During periods of high utilization, larger setup cost numbers could be used, reducing the number of reworks that result. Finally, we assume that the randomness in the length of each fiber is independent of all others. It is common that, although each fiber is drawn through a separate reel, several reels are mounted on the same machine. Systematic errors on one machine are rare, owing to careful process control and preventive maintenance. The length randomness occurs primarily from discrepancies in individual reels. Thus, while some fibers are processed on the same machine with an identical length setting, the yield process is equivalent to the fibers being drawn and colored on different machines. Consequently, the processes under which the fibers come out of tolerance are independent.

We formulate an infinite stage dynamic programming model for the assembly yield problem. At each stage, the number of fibers that need to be produced is n. The required length of the fibers is L, and the decision variable is the length of safety length to add, l. Given a planned length of $L + l$, each fiber can turn out shorter than L with probability $p(l)$. The following assumptions are made about the probability of failure, $p(l)$.

- There exists an $l_{\min}$ such that $p(l_{\min}) = 1$ (for example, set $l_{\min} = -L$).
- There exists an $l_{\max}$ such that $p(l_{\max}) = 0$ and $l_{\max} \leq L$.
- $p(l)$ is continuous, differentiable, decreasing in l, and the mapping of p to l is one-to-one in the closed interval $[l_{\min}, l_{\max}]$. This guarantees an inverse mapping $l(p) = p^{-1}(l)$.

Given these assumptions, the number of fibers that are started is exactly the number required. Since each fiber is dyed a different color, running two fibers of the same color is clearly more expensive than running one fiber of length $L + l_{\max}$ (which guarantees a fiber that meets requirements). Hence, two fibers of the same color are never started in the same production run, and at a stage requiring n fibers, exactly n fibers are started. The processing at each iteration results in i fibers that are short, and $n - i$ that meet the length requirements. Since the l chosen is the same for each fiber, the probability that exactly i ($i \leq n$) fibers are short is given by the binomial distribution, $\binom{n}{i} p^i (1 - p)^{n-i}$. Note that the decision maker has direct control over the yield process and can generate "zero defects" by choosing $l_{\max}$.

There are two costs associated with production. Each run involves a setup cost K, independent of the number of fibers. The cost of the fibers is proportional to the length of the fiber and is c per unit length. It includes both material and processing costs (incorporating the scrap value of short fibers is discussed in the conclusion).

Notice that picking a length, l, is equivalent to picking a probability of failure, p, since an inverse mapping $l^{-1}(p)$ exists. The analytic results are more easily derived by using p instead of l, and so we state the problem as a decision about the probability a fiber fails to meet the length requirement, p. Let $f(n)$ be the minimum expected cost to produce a cable with n fibers, and let $g(n, p)$ be the expected cost to produce n fibers given the decision to add length $l(p)$ to the fibers on the first run (and act optimally in subsequent production runs to replace fibers that are too short). The formulation can be written:

$$f(n) = \min_p g(n, p), \tag{1}$$

$$g(n, p) = K + cn(L + l(p)) + \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} f(i), \tag{2}$$

$$f(0) = 0. \tag{3}$$

The objective function $f(n)$ returns the least cost of making n fibers by minimizing over all values of p. $g(n, p)$ defines the cost of making n fibers and includes the setup cost, K, the variable cost of the fibers, $cn(L + l(p))$, and the expected cost of remaking the fibers that fall short. Notice that if there is a setup cost associated with each fiber of K_1, it can be accommodated easily by replacing L with $L + K_1/c$. Finally, define π_n to be the value of p that optimizes the n-fiber problem.

3. PROPERTIES OF THE PROBLEM

The general formulation of the problem, without further assumptions, has little structure that can be exploited. However, under certain conditions on the shape of the curve $l(p)$, the behavior of the optimal policy can be analyzed. The following two theorems provide insights that will be used to develop solution procedures. Proofs of the theorems are given in the Appendix.

Theorem 1. *The following results hold.*

1. $g(n, p)$ *and* $(g(n, p) - p^n f(n))/(1 - p^n)$ *have identical minima, and the value of the functions at the minima are identical.*
2. $f(n) \geq f(n-1) \quad \forall\, n \geq 1$.
3. $f(n+1) - f(n) \leq f(n) - f(n-1) \quad \forall\, n \geq 1$.
4. *If* $K = 0$, *the optimal solution is independent of the number of fibers*, n.
5. *If* $l(p)$ *is weakly concave in the closed interval* $[0, 1]$, *the optimal solution to* $f(n)$ *is* $p = 0$.

Theorem 1 is useful in characterizing the solution to the n-fiber problem. Notice that (1) defines $f(n)$ in terms of $g(n, p)$, but that $g(n, p)$ has a term including $f(n)$ for the case in which the first production run results in no usable fibers. This circular definition causes computational problems in the solution of the dynamic programming recursion. Part 1 of Theorem 1 provides us an elegant way around this complexity, in that it defines a way to find the optimal policy π_n in terms of only $f(0)$ through $f(n-1)$, independent of $f(n)$.

Parts 2 and 3 are useful for the derivation of the remaining results, and simply characterize the behavior of $f(n)$ as n increases. Part 4 shows the stationarity of the solution when the setup cost is zero. In the zero setup cost problem, the link between the fibers is removed, and the planning decision for each fiber can be made independently. The number of setups carried out is inconsequential, and the economies of batching the fibers are no longer present. Hence, each fiber will be processed as though it were an independent fiber, and the result follows.

Part 5 characterizes the solution of the problem when $l(p)$ is weakly concave in p. It is optimal for the firm to plan very long lengths and to ensure that each fiber is long enough when $l(p)$ is concave. The result comes from the fact that as the planned length is increased, in the interval $[l_{\min}, l_{\max}]$, the benefit of increased length increases at an increasing rate. Since the cost increase remains linear, every increase in l decreases total cost, and we choose $l_{\max}$.

If most $l(p)$ functions found in practice were indeed concave, the above analysis would have been sufficient. However, this is usually not the case. Typical $l(p)$ functions are either convex, or have both concave and convex sections. Several common representations of length variability (such as the assumptions of normal, log-normal, or exponential variations in the length from the planned length setting) lead to convex or near-convex $l(p)$ functions. Theorem 2, which follows, deals with such distributions.

Using Part 1 of Theorem 1, we can rewrite the solution to the one fiber problem:

$$\pi_1 = \operatorname*{argmin}_p \frac{K + c(L + l(p))}{1 - p}. \tag{4}$$

Let π_∞ be the solution to the problem with no setup costs and any number of fibers. Using Part 1 and Part 4 of Theorem 1, this can be written as:

$$\pi_\infty = \operatorname*{argmin}_p \frac{c(L + l(p))}{1 - p}. \tag{5}$$

We prove the following theorem for those $l(p)$ functions that do not have multiple optima π_n or multiple π_∞. Notice that this assumption holds for a large number of distributional assumptions for $l(p)$.

Theorem 2. *The following results hold.*

1. *The optimal solution to the n-fiber problem increases in n*: $\pi_1 \leq \pi_n \leq \pi_{n+1}$.
2. $\lim_{n\to\infty} \pi_n = \pi_\infty$.

The proof of this theorem is given in the Appendix. The theorem bounds the optimal solution using the results from (4) and (5). As the number of fibers becomes large, the solution tends to the solution of the problem without setup cost, although this result is not evident. Consider that as the number of fibers increases, the number of setups required to produce all fibers also increases. Therefore, it would appear that setups should play an increasingly important role in the determination of the safety length as the number of fibers increases.

Theorem 2 states that precisely the opposite effect occurs, as can be intuitively explained using marginal analysis. Consider increasing the number of fibers required from n to $n + 1$. As n increases, the probability that the last fiber is the only failure decreases. Hence, the expected setup cost incurred due to the extra fiber becomes smaller as n increases. The optimal solution when an infinite number of fibers are processed is the same as in the zero setup cost problem.

An immediate implication of Theorem 2 is for the design of heuristics. Since the probabilities converge to π_∞, the "zero setup cost" heuristic uses the solution to (5) for any n-fiber problem. On the other hand, the convergence might be over a long interval, and so the "one-fiber" heuristic uses the solution to (4) as the solution to the n-fiber problem. This heuristic is optimal for a one-fiber problem, and may be better than the zero setup heuristic for small values of n. A third simple heuristic, the "hybrid" heuristic, combines the first two heuristics by minimizing the following expression for an n-fiber problem:

$$\frac{K/n + c(L + l(p))}{1 - p}. \tag{6}$$

Notice that when $n = 1$, this reduces to (4), and when $n = \infty$, it reduces to (5). The inclusion of K adjusted for the number of fibers has the potential for providing solutions between the extreme points, and these intermediate solutions may dominate either of the extreme point heuristics. The next section develops a method for finding the optimal solution and tests the solution quality of these three heuristics.

4. OPTIMAL SOLUTION AND COMPUTATIONAL EXPERIMENTS

From Theorems 1 and 2, it is possible to bound the solution for a large interesting class of problems. If the number of fibers is few, it would be reasonable to use the one-fiber solution as a heuristic. Similarly, when the number of fibers is large, the zero-setup cost problem is likely to be near optimal. The hybrid solution may not be as close as either on a single problem instance, but its behavior may be more predictable over the entire range of problems. For a problem with a number of fibers between these extremes, determining the best heuristic requires the computation of the $g(n, p)$ matrix, which would reveal the optimal solution. In this section we describe the optimal solution technique, and then an experimental design to study the performance of the heuristics.

The algorithm for computing the optimal solution was coded in C and compiled on a Sun SPARCStation 2. The probability distribution is discretized over a number of intervals that can be chosen for the desired precision (we used 1,000 intervals for our experiments). Arrays for $f(n)$ and π_n are set to zero, and the values π_1, $f(1)$, and π_∞ are computed.

For each n starting with two, $f(n)$ is calculated as follows. A loop for the range $p \in [\pi_{n-1}, \pi_\infty]$ finds the minimum value of

$$g_1(n, p) = \frac{K + cn(L + l(p)) + \sum_{i=0}^{n-1} \binom{n}{i} p^i (1 - p)^{n-i} f(i)}{(1 - p^n)}, \tag{7}$$

and sets $f(n)$ equal to that value (and π_n to the corresponding p). Repeating this loop until reaching the desired value of n would yield π_n and $f(n)$. The ability to restrict the values of p makes the algorithm quite fast for the distributions we experimented with. Since we wanted to examine the behavior of the $g(n, p)$ functions, we calculated the entire matrix for all the experiments.

We developed a parameterized set of experiments to test the heuristics and to examine the cost behavior of the problem. We varied the ratio of setup cost to unit production cost, the length of the cable, and the distribution of the production process. To vary the distribution, we used two functional forms: a power curve and a truncated normal. The power curve related the added length and the probability of failure by the functional form $1 - (p^{1/k})$, and we let k take the values four and six. The truncated normal was computed by mapping a certain number of standard deviations of the standard normal to the finite interval of variability of cable length. The intervals of the standard normal we used $\pm 3\sigma$ and $\pm 5\sigma$. The range of cable length variability was specified as a proportion of the target length. For example, the low variability experiments had probability of failure equal to one if only the target length is started, and is equal to zero if one and one-half times the target length is started. The specific values used were as follows

	Low	Medium	High
Cost Ratio	1	10	50
Length	10	50	200
Variance	(1, 1.5)	(0.8, 1.8)	(0.7, 2.2)

We compare the one-fiber, zero-setup cost and hybrid heuristics with the optimal solution. In order to estimate the magnitude of savings these heuristics might provide, we define a fourth solution technique that implements "zero defects" by setting the length to its maximum, in order to have a zero probability of failure. These decisions are labeled "Maximum Length" in Tables I and II. The optimal and heuristic solutions for all combinations of experiment parameter values were computed in about two and one-half hours on the workstation.

Table I
Experiment Results Expressed as the Percentage Above the Optimal Solution

(n = 1 to 50, obs = 5,400)	Zero setup cost	One-fiber	Hybrid	Maximum Length
Overall Average	0.5820	0.1594	0.0027	24.8648
Short Length	1.5934	0.3076	0.0073	21.7424
Medium Length	0.1409	0.1478	0.0006	25.7686
Long Length	0.0117	0.0228	0.0001	27.0834
Small Setup Cost	0.0019	0.0045	0.0000	27.3938
Medium Setup Cost	0.1413	0.1488	0.0006	25.7130
Large Setup Cost	1.6027	0.3250	0.0073	21.4876
Fourth Power	0.7566	0.2875	0.0029	27.1204
Sixth Power	0.7250	0.2435	0.0029	40.1437
Normal $\sigma = 3$	0.4162	0.0638	0.0023	8.9266
Normal $\sigma = 5$	0.4301	0.0429	0.0025	23.2684
Small Variance	0.4633	0.0180	0.0029	10.9809
Medium Variance	0.6050	0.1373	0.0027	25.3509
Large Variance	0.6777	0.3229	0.0023	38.2626

Some sample plots of the $g(n, p)$ functions are shown in Figures 1 through 4. These plots demonstrate the behavior predicted by the theorems, with π_n (the optimal probability of failure) increasing in n. They also show that as the probability used moves away from optimal, the cost penalty can grow quickly.

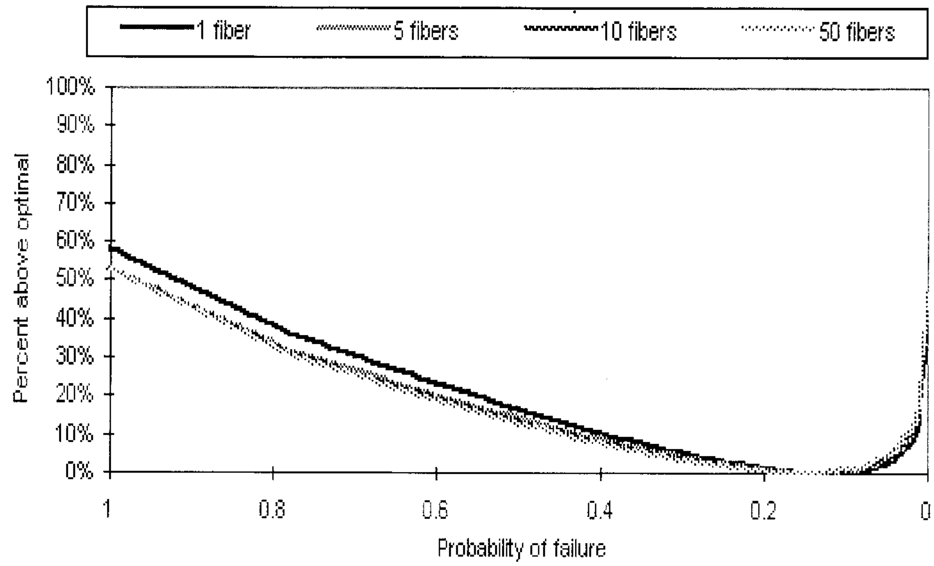
The comparison of the heuristics to the optimal solution was done in two ways, since the magnitude of the total costs for a 50-fiber problem is much greater than for a five-fiber problem. The percentage above the optimal solution, and the absolute size of the cost error, are computed for the comparisons. The average behavior of the heuristics over all parameter settings as the number of fibers increases is shown in Figures 5 and 6. As expected, the one-fiber heuristic performs well when the number of fibers is small and gets worse as n increases, while the zero-setup cost heuristic behaves in the opposite fashion. The hybrid is not always best, but it does perform well across the entire range and dominates the zero-setup cost heuristic.

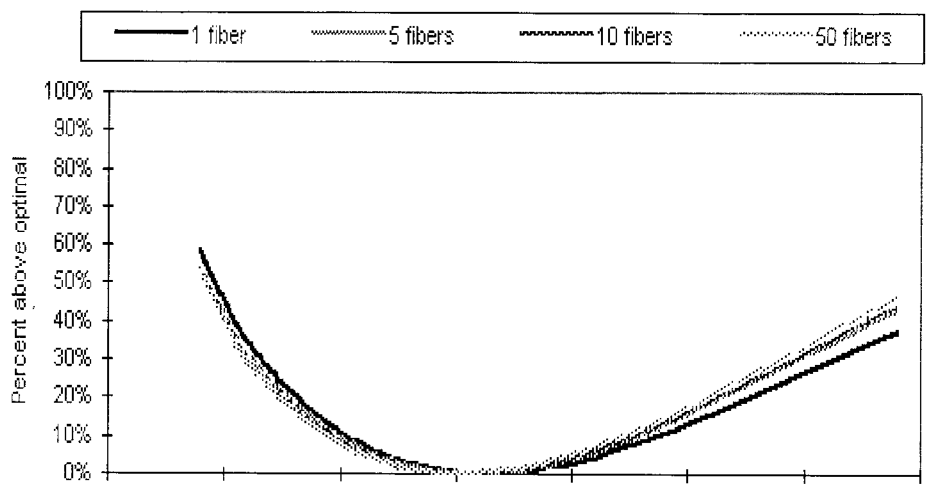
Tables I and II show the results for subsets of the experiments, holding one parameter of the experimental design constant. Note that over all the experiments, the hybrid heuristic dominates the other two, especially on a percentage basis. On an absolute error basis, the two extreme point heuristics have similar performance, but not on a percentage basis. This comes naturally from the fact that the zero-setup cost solution for a one-fiber problem is a much bigger percentage error (due to the low total cost) than the one-fiber solution is for a 50-fiber problem.

As one might expect, all the heuristics worked better for long lengths or small setup costs, since the difficulty for

Table II
Experiment Results Expressed as the Absolute Cost Above the Optimal Solution

(n = 1 to 50, obs = 5,400)	Zero Setup Cost	One-Fiber	Hybrid	Maximum Length
Overall Average	0.2496	0.2557	0.1252	84.7117
Short Length	0.5372	0.1839	0.2738	8.7607
Medium Length	0.1645	0.3720	0.0790	48.1646
Long Length	0.0470	0.2114	0.0228	197.2096
Small Setup Cost	0.0004	0.0026	0.0002	86.0985
Medium Setup Cost	0.0404	0.1129	0.0195	85.3769
Large Setup Cost	0.7080	0.6518	0.3557	82.6596
Fourth Power	0.3230	0.4486	0.1297	94.3493
Sixth Power	0.2836	0.3525	0.1238	123.4241
Normal $\sigma = 3$	0.1965	0.1358	0.1196	38.4251
Normal $\sigma = 5$	0.1952	0.0860	0.1274	82.6482
Small Variance	0.2080	0.0386	0.1500	34.8068
Medium Variance	0.2449	0.2379	0.1201	83.2160
Large Variance	0.2958	0.4908	0.1052	136.1122





heuristics is in adjusting the variable cost of production for the number of setups required. As the magnitude of the adjustment required decreases, the accuracy of the heuristics increases. Also note that the zero-setup cost heuristic has the worst behavior, since it ignores setup costs entirely.

The hybrid heuristic is remarkably unaffected by the probability distribution function for $l(p)$ and dominates the other two. The flatter curve of the normal distributions results in smaller cost penalties than the sharper power function distributions for the other two heuristics. We also see that the zero-setup cost heuristic is less sensitive to the distribution parameters. This is because the $g(n, p)$ curve is sharper to the right of the optimal solution (where π_1 lies) than to the left (where π_∞ lies).

5. CONCLUSIONS

In this paper we have developed a model for a subassembly manufacturing yield problem and provided a solution to the model, as well as some efficient heuristic methods. While it is possible to develop heuristics that more accurately capture the rate that π_n converges from π_1 to π_∞, the simple heuristics presented here perform quite well in our tests. This paper has also discussed the nature of safety