# Lot-Sizing With Random Yields:
# A Review

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

Hau L. Lee
Department of Industrial Engineering &
Engineering Management
Stanford University
Stanford, CA

Technical Report 89-16

April 1989

# LOT-SIZING WITH RANDOM YIELDS: A REVIEW

Candace Arai Yano
Department of Industrial &
Operations Engineering
University of Michigan
Ann Arbor, MI

Hau L. Lee
Department of Industrial Engineering &
Engineering Management
Stanford University
Stanford, CA

April 1989

**LOT-SIZING WITH RANDOM YIELDS:
A REVIEW**

Abstract

This paper reviews the literature on quantitatively-oriented approaches for determining lot sizes when production or procurement yields are random. We discuss issues related to the modeling of costs, yield uncertainty, and performance in the context of systems with random yields. We also present a taxonomy of lot-sizing problems with random yields which provides a framework for our review. Finally, we give a brief analysis of the existing literature and suggest directions for future research.

# LOT-SIZING WITH RANDOM YIELDS: A REVIEW

## 1. INTRODUCTION

The problem of determining production and procurement quantities and their timing, also known as *lot-sizing*, is one that faces every pure inventory and production-inventory system. A considerable amount of effort has been focused on solving these problems when demands and production rates are known. Many sophisticated procedures are now available to solve problems of this type for complex manufacturing systems optimally or near-optimally (e.g., Afentakis and Gavish 1986, Muckstadt and Roundy 1987). Also, good heuristic procedures, too numerous to mention here, have been developed and some have been tested extensively.

Much effort has also been spent on determining production and inventory policies when demand is uncertain. Graves (1987) provides a survey of many of the analytical models of these problems in manufacturing systems. Entire books have been written on inventory problems (e.g., Arrow, Karlin, and Scarf 1960, Love 1979, and Banks and Fabrycky 1987), with much of the focus on random demand.

Considerably less research has considered random yields in production or procurement, and much of this has been done recently. The goals of this paper are to classify and describe the research to date on lot-sizing in the presence of random yields. Realizing the need to limit the scope of such a review, we have chosen to focus on analytical models for pure inventory and production/inventory systems, and within those areas, on those that consider lot-sizing decisions. Consequently, we do not discuss quality control procedures for such systems, except as they relate to lot-sizing decisions. We also do not include performance evaluation models based on queueing networks, except where lot-sizing or inventory issues are of primary concern. It should also be noted that only a few studies based entirely on simulation are reviewed.

Many of the application areas for random-yield lot sizing models are well known: electronic fabrication and assembly, chemical processes, and procurement from suppliers that produce imperfect products. There are, however, many traditional discrete parts manufacturing processes that experience random yields. While the ultimate goal for yields is to "make it right the first time," the literature reviewed here can be used in a variety of ways. First, in the short term, the results can be used to help an operation run more effectively so that effort can be focused on improving performance, including yields. Second, process improvements and supplier selection decisions can be assessed more accurately and effectively if the system-wide effects of these decisions on yields are modeled appropriately and, where appropriate, optimized. Finally, these models can assist in capacity planning decisions when the yield randomness is expected to be a long-range concern, such as in industries with rapidly changing technology and product mix.

In the remainder of this section we discuss some modeling issues that arise in the context of systems with random yields. Section 2 presents our proposed taxonomy of lot sizing problems with random yields. Sections 3, 4, and 5 contain reviews of general literature, articles on continuous review models, and papers on periodic review models, respectively. The paper concludes with a brief analysis of the existing literature and a discussion of research directions.

## 1.1 Modeling Issues

Three important issues in modeling systems with random yields are:

1. The modeling of costs affected by the presence of random yields.
2. The modeling of yield uncertainty.
3. Measures of performance.

We will discuss each of these issues in turn.

### 1.1.1 Modeling of costs affected by random yields

The modeling of some costs, such as setup costs, shortage costs,.and salvage values, usually are not affected by the presence of random yields. However, variable unit costs and inventory holding costs must be modeled differently depending on the specific application.

Normally, a variable unit cost is attributed to each unit of the production or procurement quantity. When yields are random, the production *output* quantity may differ from the production *input* quantity, and the quantity *received* from a supplier may differ from the quantity *ordered.* Thus, one needs to be careful to define the variable cost as a function of the appropriate quantity. In manufacturing settings, costs are usually related to the production input quantity, whereas in pure inventory settings, the quantity received might be more relevant, especially if one has the option of returning defective items for replacement or refund.

The definition of variable costs may also be situation-specific. In manufacturing environments, there may be a variable cost of (raw) materials per unit of production input, as well as a variable cost per unit for processing. If a defective unit cannot be reworked and must be disposed or sold as scrap, then both the materials and processing cost per unit should be included. On the other hand, if it is possible to rework the part to make it acceptable, then the cost of initial processing and rework should be included, but the materials cost is constant and therefore need not be included.

For both production and procurement situations, inventory holding costs often depend upon the timing and nature of the inspection process. In procurement, one might perform 100% inspection upon arrival of an order from a supplier, inspect a small sample to accept or reject the lot, or not inspect at all. If 100% inspection is performed and if the inspection process is perfect, then defective units can be returned to the supplier or disposed (as appropriate), and inventory holding costs should be charged on only the good units. On the other hand, if the inspection process is imperfect, the modeling of inventory holding costs is more complicated. If sampling is used to accept or reject the lot, or if no inspection is done, then inventory holding costs are also charged on defective units while they are in inventory. Here, it would be reasonable to assume that each defective unit is identified soon after its sale or use in production, and is returned or disposed at that time.

Similar concepts apply to production environments, except that one cannot return a defective unit, but may have the opportunity to rework it. In some instances, it is also possible to inspect units one by one as they complete production and to control the input in accordance with the outcome of inspection. Hence, it is possible to produce exactly the required amount of good units if one chooses to operate the production facility in this way. Note that such an operation would have random duration production runs.

In both production and procurement situations, accounting for inventory holding costs can be further complicated by cost accounting rules and inventory valuation procedures. Sometimes, the cost associated with produced or unreturnable defective parts is considered a "period cost," which is charged immediately to cost of goods sold. In other cases, the cost of these parts is capitalized, and the inventory is valued by determining the average cost of a good unit. The tax ramifications differ between the two cases, and thus, affect the cash flow, which in turn should affect the manner in which inventory holding costs are modeled.

In some cases, the setup costs are affected by the presence of random yields. For example, some machine setups require trial and error adjustments during which the defective output must be scrapped. Here, most of the out-of-pocket cost attributable to setups is the result of random yields, and the setup cost itself is random.

### 1.1.2 Modeling of Yield Uncertainty

Before we describe models of yield uncertainty, it is worthwhile to point out the existence of circumstances where yields are random, but the randomness does not complicate the lot-sizing decision, *per se*. Consider a system that produces one unit at a time, and in which inspection is done immediately and instantaneously. Hence, immediate feedback on the yield outcome is available. Here, as long as one does not need to plan for raw material inputs or to allocate scarce capacity among several products, no alterations of the lot sizing decisions need to be made. One can simply continue to produce until the production target is met. In most circumstances, however, raw materials are required and capacity is limited, so lot-sizing decisions that incorporate yield uncertainty must be made somewhere in the organization.

Yield uncertainty has been modeled in several different ways in the literature. It is our opinion that each of these modeling approaches is applicable to certain classes of real systems. Some, of course, are more general than others, but they also have the disadvantage of being difficult to manipulate, both algebraically and numerically. Thus, the modeler must trade off accuracy with tractability. Below we describe five different ways of modeling yield uncertainty, and discuss some of their advantages and disadvantages. We also give examples of where each might be a fairly accurate representation of the underlying random process generating the yield distribution.

The simplest model of random yields assumes that the creation of good units is a Bernoulli process. Hence, the number of good units in a batch of size Q has a binomial distribution with parameters Q and p, where p is the probability of generating a good output from one unit of input. Implicit are the assumptions that the process is stationary and that the generation of one unit is independent of the generation of other units (i.e., no autocorrelation). One advantage of using this model is that one needs to specify only the value of p. One disadvantage is that it does not permit the specification of other forms of the variance of the fraction of good units. Indeed, in this modeling approach, the expected value of the fraction of good units is invariant with Q, but the variance of that fraction decreases with Q. This representation of yield uncertainty is appropriate for systems that are in control (from the standpoint of statistical process control) for long durations.

A second relatively simple way to model yield uncertainty is to specify the distribution of the *fraction* of good units (or yield rate). The main difference between this modeling approach and the binomial model is that it permits specification of both the mean and variance of the fraction good. On the other hand, it forces the distribution of the fraction good to be independent of the batch size. This model applies when relatively large batch sizes are used, or when the variation of the batch size from production run to production run tends to be small. It also applies to circumstances where yield losses occur because of limited capabilities of the production system to adapt to random environmental changes or variations in materials, etc. Here, the yield loss might be relatively predictable for any particular set of conditions, but the conditions are not predictable. Thus, the yield rate distribution reflects the relative likelihoods of the various possible conditions. In most cases the fraction of good units is bounded above by one. However, it is possible that this "fraction" is unbounded. This is the case when the inputs and the outputs of the production process are not measured in the same units.

A third modeling approach involves specifying the distribution of the time until a repetitive process becomes "out of control" and starts to make defective parts. Here, it is normally assumed that the process is in an "in control" state at the beginning of a production run. In this context, if one views the production decision as how long to let the process continue, then the fraction acceptable is stochastically decreasing with the length of the production run for any increasing failure rate (IFR) distribution. Prime examples of this are situations where the "failure" is the result of deterioration of the state of the production system during a production run (e.g., tool wear or breakage).

Sometimes the fraction acceptable is stochastically *increasing* with the length of the production run. This occurs when a startup of a process involves trial and error in setting values that affect the quality of the output. For example, situating a die properly, or choosing a temperature that is consistent with both environmental conditions and properties of the substance being heated often require some trial and error. For these situations, it might be possible to model the process in a similar fashion to the stochastically decreasing case. For instance, one might specify the distribution of time until the first acceptable unit is produced. Thus far, this has been done using a Bernoulli process where the process is assumed to be out of control until the first good unit is observed, at which point the process is assumed to be in control. Certainly, similar phenomena can be modeled using generalizations of these ideas.

Finally, there are several more general modeling approaches for both batch and repetitive processes, where independence, stationarity, and stochastically decreasing or increasing patterns are inapplicable. For batch processes, the most general approach used to date involves specifying *for each possible batch size* the probability that *each possible output quantity* will occur. This approach requires much experimentation and data collection, since it might be necessary to try batch sizes not ordinarily used. For each batch size, a reasonable sample size is necessary to obtain a good estimate of the distribution. A variety of non-stationary stochastic processes can be used to model yields for repetitive processes, but to the authors' knowledge, this has not been done in the context of making lot sizing decisions.

### 1.1.3 Measures of Performance

Throughout the literature on lot sizing with random yields, minimization of expected costs (or maximization of expected profit) is the stated criterion. The objective functions for cost minimization problems contain a subset of: setup costs, variable production or procurement costs, inventory holding costs (binary, per unit, and per unit per unit time), shortage costs (per unit, and per unit per unit time), salvage values, and disposal costs.

Although minimization of expected costs is certainly a reasonable objective, it is surprising that, with few exceptions, constraints on various measures of service have not been considered in lieu of shortage costs, especially since customer service measures are so popular in the literature on lot sizing with random demands. Perhaps these measures will be considered as the literature on random yields grows and matures.

In addition, to date, the issue of robustness has received little attention in the literature on random yields. We suspect the reason for this is that the random yield models and the associated solution procedures are somewhat more complex, both analytically and computationally, than models for random demands. This makes it difficult to analyze robustness issues. Yet, such investigations are needed since the same parameter estimation concerns exist in problems with random yields as well as those with random demands.

In the next section we present a taxonomy of lot sizing problems in the presence of random yields.

## 2.0 TAXONOMY OF LOT SIZING PROBLEMS WITH RANDOM YIELDS

This section describes a taxonomy of problems which is based on existing literature. We believe that the taxonomy is useful in identifying the characteristics of problems that have been addressed, and in pointing out where further research is needed. It will also be useful in organizing our review of the literature.

As with inventory models, the two major categories are continuous time models, and discrete time models. Within each of these categories, the problems are distinguished principally by the structure of the production or procurement system (e.g., single-stage, facilities in series, assembly). The continuous time models to date are infinite horizon models and assume that stationary policies are used. On the other hand, the discrete time models deal with single period, multi-period (finite horizon) and infinite horizon situations, and some permit non-stationary policies. There are also research efforts directed toward modeling complex manufacturing systems Our proposed taxonomy of research to date appears in Figure 1.

FIGURE 1

We now turn to a review of the literature. We have attempted to provide a complete review and apologize for any inadvertent omissions.

## 3.0 GENERAL PAPERS

Whybark and Williams (1976) use simulation to study the efficacy of safety stock and safety time as buffers against quantity and timing uncertainty of demand and supply in single-stage, single-product systems. Their results suggest that safety stock is preferred for quantity uncertainty and safety time is preferred for timing uncertainty.

Since random yields normally create quantity uncertainty rather than timing uncertainty, we might construe that safety stock would be a better buffering mechanism than safety time for the models considered in the next two sections. However, random yields can cause timing uncertainty if one needs to wait until enough good units are produced. Thus, there may be a role for safety time as well.

New and Mapes (1985) propose four different strategies to deal with random yields. For continuous production, make-to-stock situations, they recommend adjusting the production quantity for the average yield rate, and using fixed buffer stocks. For continuous production, make-to-order situations, they suggest a modification to the above policy to allow multiple production runs. They recommend the use of a service level (probability of satisfying the order) for a single custom order or for products having infrequent demand. Finally, for multiple-order custom products, they propose using safety time so that the production can be divided into smaller batches. This would enable more frequent inspection and monitoring of the number of good parts produced, reducing the risk of and size of overruns.

We now turn to a review of analytical models in a continuous review framework.

## 4.0 CONTINUOUS TIME MODELS

In continuous time models, demand for the product and decisions to produce or replenish the product can occur at any point in time. The literature on continuous time models can be classified according to demand characteristics: demand modeled as deterministic and constant; and demand modeled as a stochastic process.

### 4.1 Continuous Time Models with Constant Demand Rates

The Economic Order Quantity (EOQ) model is a classic example of an inventory model with constant demand rates. Although this assumption seems quite restrictive, it has been found that models such as the EOQ are quite robust with respect to the model parameters, and that sometimes the results obtained from these simple models are good starting solutions for more complex models.

Several papers have been written on models that incorporate the effect of yield uncertainty under the basic EOQ framework. The basic premise is that the amount received from an order is uncertain. Results differ depending on how the yield uncertainty is modeled.

One of the earliest papers on this topic was written by Silver (1976). In his model, randomness of the number of defective units is only one of the many sources of uncertainty affecting the quantity received. The other sources include human administrative errors, good production runs leading to larger than the required quantity, convenience in operations, and pilferage and damage. Hence, it is possible that the amount received is greater than the amount ordered. Silver considers two cases: one in which the standard deviation of the amount received is independent of the lot size, and another in which it is proportional to the lot size. The ratio of the expected amount received to the lot size is assumed to be constant, i.e., the expected yield rate is independent of the lot size. He refers to this ratio as the "bias factor." In both cases, it is shown that the optimal lot size is a slight modification of the EOQ.

Shih (1980) focuses on the case where yield uncertainty is caused by defective units. The yield rate is thus between 0 and 1, and is assumed to be invariant with the lot size. Such a representation of the yield rate also satisfies the assumption that the bias factor (cf. Silver 1976) is independent of the lot size. The resulting expression for the optimal lot size is again a simple modification of the EOQ formula. The modification involves only the first two moments of the yield rate distribution, and as expected, the optimal lot size is larger than the conventional EOQ.

Kalro and Gohil (1982) extend Silver's model to include complete and partial backlogging of demands. In the partial backlogging case, it is assumed that a known, constant fraction of demand is lost during a stockout period. Of course, since demands are deterministic and constant, the stockout period is also known and constant, and can be viewed as a decision variable itself. Mak (1985) extends this model even further by assuming that the fraction of demand lost during a stockout period is itself a random variable. He uses Shih's model of the yield rate distribution. Again, the optimal lot size and the optimal length of the stockout period are shown to be functions of the first two moments of the yield rate and the fraction of demand lost during a stockout.

In the above models, defective items received from the supplier are assumed to be identified and discarded (or returned for replacement) immediately after they arrive. In other words, a complete and perfect inspection mechanism is assumed to exist. Another line of research attacks the problem of whether inspection should be performed immediately upon arrival of an order, or when they are actually shipped to customers. The tradeoffs involve inspection and holding costs. Lee and Rosenblatt (1985) first explore this problem using the EOQ model under the assumption of a constant yield rate. Zhang and Gerchak (1988) extend the model to the case where the yield rate is not constant. Specifically they assume that each individual unit has a fixed probability of being defective.

Some recent research has attempted to model the production process from which defective products are produced. Porteus (1986) and Rosenblatt and Lee (1986a) have independently incorporated the effects of imperfect production processes into the basic EOQ model. In these papers, it is assumed that an order initiates the production of a batch of units. The production process is characterized by an in-control state under which defect-free units are produced, and an out-of-control state under which some or all of the units produced are defective. The production process begins in the in-control state after a setup is performed. At some point in time, the process goes out of control and remains in that state until the next setup, which puts the process back into control. Defective items are reworkable instantaneously at a cost. Porteus assumes that there is a constant probability that the process goes out of control while producing each unit (which gives a geometric distribution of the "time to failure"), whereas Rosenblatt and Lee assume that the time to failure is exponentially distributed. In the former paper, once the process is out of control, all of the items are assumed to be defective, whereas in the latter, only a fraction of the items produced is assumed to be defective. Both papers develop approximate formulae for the optimal lot size, which are shown to be smaller than the conventional EOQ. Thus, the presence of defective products motivates smaller lot sizes here.

It should be noted that the dichotomy of in-control and out-of-control states is a standard one in the quality control literature (see Juran and Gryna 1980 and Grant and Leavenworth 1988). Porteus's model of the shift from the in-control state to the out-of-control state is actually identical to that in an earlier work by Vachani (1969) on a production problem to meet a one-time demand. The assumption of an exponential time to failure is, of course, a classic one in the statistical process control literature (see Duncan 1956). Porteus's assumption of a geometric time to failure might be viewed as a discrete approximation of the exponential time to failure assumption. Lin *et al.* (1988) relax the exponential assumption to a general distribution. The cost function can be specified explicitly only for special cases, such as the uniform or Weibull distribution. Consequently, numerical optimization procedures are used to solve their problem. Extensions of the model of Rosenblatt and Lee (1986a) to include multiple out-of-control states, and the case where defective units may be produced even during the in-control state, are straightforward.

The possibility of intermediate ("process") inspections during a production run is considered in Lee and Rosenblatt (1987) using an approach in which the optimal production lot size and inspection intervals are jointly determined. They observe that, with the use of process inspections, the optimal production lot size can be larger than that without, as in Rosenblatt and Lee (1986a). Rosenblatt and Lee (1986b) compare the use of continuous process inspection and periodic process inspections under the same framework. In a later paper they (Lee and Rosenblatt 1988) allow for shortages, and consider the cost of restoring the process from an out-of-control state back into control as a function of detection delay.

In the model of Lee and Rosenblatt (1987) where inspections are allowed, if one assumes that the number of inspections within a production cycle and the lot sizes can be fractional, then the optimal solution can be obtained by treating the two problems separately and independently. Porteus (1988) shows that this is indeed the case, and attributes it to the assumption that inspections are instantaneous, i.e., the outcome of an inspection is known without delay. When there is a time lag between the inspection and its outcome, Porteus shows that the two problems are no longer separable, and determination of the optimal lot size and inspection intervals is much more complicated.

In the above EOQ-type models with imperfect yields, the yield distribution itself is assumed to be known and given. Investments to improve the production process can have an impact on this distribution. Cheng (1989) assumes that there is a constant yield rate, and the unit production cost of the item increases with the yield rate. For a specific form of the relationship of the yield rate to the unit production cost, a closed-form expression for the economic production quantity is obtained. Gerchak and Parlar (1988) consider the problem of jointly determining yield variability and lot sizes when yield variability can be reduced through appropriate investments. The approach of Silver

(1976) to model yields is used, where the "bias factor" is assumed to be constant. Gerchak and Parlar model a situation in which yield variability can be changed through appropriate investments. Closed form solutions for the optimal investment level and lot size can be obtained only for special cases of the investment function, where the investment function reflects the amount of investment needed as a function of the yield variability

The basic EOQ model framework also can be used to analyze the problem of diversification of suppliers (see Gerchak and Parlar 1988). Diversification will reduce the effects of yield variability. Consider two suppliers that have yield rates with different means and variances The problem is to determine the optimal lot sizes for both suppliers so as to meet demand and minimize costs. Gerchak and Parlar show that if the mean yield rates for the two suppliers are the same, then the optimal lot sizes should be proportional to the product of the ratio of the mean yield rates and the reciprocal of the ratio of the respective yield variances. Hence, the result is similar to one on portfolio optimization in the finance literature. Gerchak and Parlar also develop a condition under which diversification is worthwhile.

### 4.2 Continuous Time Models with Random Demands

Relatively little research has been done on continuous time models with random demands as well as random yields. For the case of random demands with backlogging, the model is tractable when one assumes that there is at most one replenishment order outstanding at all times, and when the backlog costs are a function of the number of units short. For such a case, Noori and Keller (1986a) consider the standard Reorder-Point/Reorder Quantity (r, Q) model (cf. Hadley and Whitin 1963), and use an approach similar to that of Silver (1976). Explicit results are developed when the bias factor is independent of the order quantity, and when the standard deviation of the amount received is linearly related to the order quantity. The optimal r and Q values are obtained using an iterative procedure.

For the general case where backlog costs are defined on a per unit per unit time basis, and where more than one replenishment order may be outstanding, Moinzadeh and Lee (1987) provide an exact analysis of the steady state operating characteristics for a (r, Q) inventory system with Poisson demands and constant replenishment leadtimes. There is no restriction on the form of the yield rate distribution in this analysis. However, the derivation of the steady state operating characteristics requires the solution of the limiting distribution of a Markov chain, and hence, it is difficult to determine the optimal r and Q. When the amount received is binomially distributed, Moinzadeh and Lee (1987) develop an effective approximation that involves adjusting the mean demand rate in the conventional (r, Q) model. Consequently, the optimal values of (r, Q) can be obtained easily.

All the papers reviewed in this section so far have assumed either that defective items are discarded upon arrival of an order, or can be repaired at a cost. In reality, it is possible that the defective items can be returned to the supplier, who would then replace these items with a new batch of items equal in size to the amount of defective items detected in the initial shipment. There is, of course, no guarantee that all the items in the second shipment are defect-free (which means there can be a third shipment, and so on). Suppose that the second shipment contains no defective items, which is a reasonable assumption if the second shipment is very small, or if the supplier pays special attention and care in this shipment. One can then think of such a situation as an inventory system where a replenishment order would arrive in two shipments, with the amount in the first shipment being a random variable. One can also think of such a model as equivalent to a case where defective items in an incoming lot can be repaired with a given repair leadtime, in which case the second shipment leadtime equals the first shipment leadtime plus the repair time. Moinzadeh and Lee (1989) develop exact expressions for the operating characteristics of an (r, Q) inventory system where the two shipment leadtimes are constant, and the intervals between successive demands form a sequence of independent identically distributed random variables. An approximate cost function for such a system is also developed based on the assumption of a negligible probability that more than one order is outstanding.

## 5.0 DISCRETE TIME MODELS

Most of the research on discrete time models has focused on models with a single period or a single stage (or both). We first review the literature on single-stage problems, beginning with the simplest of them, involving a single time period.

### 5.1 Single Stage Models

#### 5.1.1 Single Period Models

##### 5.1.1.1 Single procurement order or production run

The prototypical single-stage, single period model involves choosing a single lot size to procure or to produce to meet a single (known or random) demand, to minimize the sum of expected variable costs, inventory holding costs, and shortage costs. This problem was first addressed by Karlin (1958a). He considers two different models.

In the first model, it is assumed that the only decision available is whether to order (or produce), and that if an order is placed, a random quantity (with a known, continuous distribution) is delivered. The inventory holding costs and shortage cost are assumed to be continuous, increasing functions of the excess and shortage, respectively. Karlin shows that if the inventory holding and shortage cost functions are convex increasing in their respective arguments, then there is a single critical number (of initial on-hand inventory) below which one should place an order. Otherwise it is optimal not to order. He also shows that conditions on the holding and shortage cost functions can be relaxed with additional conditions on the distribution of the delivery quantity.

Karlin's second model permits a choice from among a finite number (n) of production levels, where each has an associated output quantity distribution. The holding and shortage costs are assumed to be convex increasing functions of their respective arguments. Karlin shows that if the distributions of the delivery quantities have a monotone likelihood ratio, then for any pair of production levels, there is at most one non-negative "breakeven" initial inventory value. (That is, below this value, one production level is preferred, and above this value, the other is preferred.) Therefore, there are intervals (of initial inventory values) for which a particular production level is dominant, and each production level is dominant in at most one interval.

Giffler (1960) considers a slightly different problem in which a binary, rather than a per unit penalty is assessed for shortages. He formulates the problem as one of finding a reject allowance which maximizes the savings relative to the alternative of no reject allowance. The proposed solution procedure is marginal analysis, and Giffler gives a condition for which marginal analysis is guaranteed to be optimal (i.e., a condition for concavity of his objective), and indicates that binomial and Poisson probabilities satisfy this condition. Levitan (1967) formally provides sufficient conditions for Giffler's procedure to be optimal and derives properties that reduce the number of first difference calculations required.

Gregory and Beged-Dov (1967) consider applications where almost all of the defective parts are produced during the setup process, and that no defectives are produced after a correct setup has been achieved. To model this phenomenon, they assume that the number of parts spoiled during a setup is geometric. In other words, there is a specified probability of achieving a correct setup for each part produced. Their objective function includes the cost of scrapping excess good units, and the cost of a single extra setup if demand is not satisfied by the first production run. They show that the objective function is unimodal and develop a closed-form expression for the shrinkage allowance. They also suggest that the approach can be extended to multiple production stages in series.

A profit maximization version of a problem with a single known demand and one production run is addressed by White (1970). He includes linear revenue, expected variable production costs, salvage values, and shortage costs. He shows that the objective function is concave and uses first differences to derive an expression which can be used to find the optimal solution.

To the authors' knowledge, no further research on the problem was published until the paper by Panda (1978). His single-period model is similar to Karlin's second model, except the production levels are not restricted and demand is permitted to be random. Also, linear holding, shortage, production, and salvage costs are assumed. For a situation in which the order quantity represents the location parameter for the supply distribution, he shows that the optimal policy has a single critical number which is the breakeven point between ordering and not ordering. He also derives additional conditions in which it is always or never optimal to order. In a closely related paper, Shih (1980) assumes that inventory holding costs and shortage costs are linear in their respective arguments and that the distribution of the fraction defective is invariant with the production level. Shih shows that the optimal production (input) quantity can be found using a variation of the newsboy model.

Gerchak, Parlar and Vickson (1986) obtain the same result for the profit maximization objective, and generalize the result to permit initial inventory. In a later paper, Henig and Gerchak (1987) prove the existence of an order point (for initial inventory) below which an order should be placed. They also show that the value of this order point does not depend upon the yield *variability* under very general conditions in a single period setting.

For the problem considered by Shih (1980), Noori and Keller (1986b) obtain closed form solutions for the optimal procurement quantity for uniform and exponential demand distributions and for various distributions of the quantity received. For a few specific cases, they derive expressions for the relative error that results from using the newsboy solution adjusted for average yield losses.

Ehrhardt and Taube (1987) consider a slightly more general problem in which the expected holding and shortage cost function is *assumed* to be convex, and ordering costs are linear in the replenishment quantity (not the quantity ordered). They demonstrate that for the case of zero setup cost, there is a single critical number (for on hand inventory) above which is it optimal not to order. Otherwise, there is an optimal order quantity which depends upon initial inventory. They also show that the form of the policy is similar when the setup cost is positive, but the setup cost must be accounted for in the determination of the smaller critical number.

Using a discrete uniform replenishment distribution and a negative binomial demand distribution , they evaluate three heuristic policies: (i) standard newsboy; (ii) newsboy solution divided by the expected fraction good; and (iii) solutions derived from the expression for the optimal policy when demand is uniformly distributed and certain other conditions are satisfied. The last heuristic uses the first two moments of the replenishment distribution. A computational study suggests that the second heuristic performs quite well.

Ehrhardt and McClelland (1987) analyze the single-period problem with linear purchase, shortage (lost revenue), and salvage costs, and permit positive setup cost. They assume that demand is random (with a known distribution), and that the distribution of the fraction defective is invariant with the order quantity. For the case of no setup cost and a continuous demand distribution, they evaluate two heuristic policies: (1) the standard newsboy solution, and (2) the newsboy solution adjusted for the average fraction defective. Their computational study shows that for shortage to holding cost ratios greater than approximately 2, the second heuristic performs very well. They introduce another heuristic which incorporates the effect of the variance of the fraction defective for situations in which demand takes on only discrete values. A computational study shows that the second heuristic performs better than the more complicated one.

For the objective of maximizing the minimum expected profit, Basu (1987) considers a class of single-period problems in which the demand distribution is unknown and the yield quantity has a distribution which is NBUE (new better than used in expectation) with mean $a + bQ$, $b > 0$, where Q is the production or procurement quantity. He shows that the optimal solution to the problem in which the yield quantity has an exponential distribution with the same mean as the original distribution is the unique optimal solution to the original problem. He also shows that as the number of observations of demand increases, the estimator of the optimal solution (obtained by using the empirical demand distribution) almost surely converges to the optimal solution.

#### 5.1.1.2 Multiple production runs

Several papers have been written on a similar problem where one has the option of making several production runs to satisfy a single demand. A series of papers in the late 1950's and early 1960's addresses the problem of determining *reject allowances* for custom-ordered job lots. The reject allowance is defined as the extra quantity processed to allow for possible defective units. In all but one of these papers, it is assumed that the manufacturing process is in statistical control. so that the production of good parts can be represented as a Bernoulli process. Thus, the number of good parts in a batch has a binomial distribution.

Bowman and Fetter (1961) discuss a reject allowance model in which the yield losses are described in terms of a (discrete) probability distribution of the *reciprocal* of the fraction good. (This is equivalent to specifying the distribution of the fraction good.) A variable cost per unit is charged on each unit of input, and a setup cost (for an additional production run) is charged if the output does not satisfy demand. The optimal policy has a critical ratio of consecutive cumulative distribution values.

Using the Normal and Poisson approximations to the binomial distribution, Llewllyn (1959) finds reject allowances that ensure satisfaction of an order, or to ensure that the capacity of the subsequent process is not exceeded, with a specified probability. The latter problem is motivated by situations where excess units must be scrapped if not used immediately. He also considers approximate cost minimization models where additional production runs with positive setup costs can be used to satisfy deficits from previous runs. His objective is to minimize expected unit production and setup costs. In the first model, he assumes that the second production run incurs a setup cost, but the variable costs in the second run and the costs associated with all subsequent runs are negligible. Using an example, he shows that increasing the setup cost results in larger reject allowances. In the second model, he also considers variable costs in the second production run and a setup cost for the third production run. From (only) one example, he concludes that the optimal decision does not change, so the second and higher order effects can be ignored.

Goode and Saltzman (1961) generalize Llewellyn's model by permitting an infinite number of setups, and incorporating salvage values (for excess good units) into the objective function. They describe an optimal procedure which is simply a dynamic programming algorithm for the problem (although they did not recognize it as such). For the same model, Hillier (1963) shows that under certain conditions, the expected cost functions are convex, so computations can be simplified by using first differences (rather than the more complex expected total cost function) to find optimal solutions. He also shows that the conditions are satisfied by the Normal approximation to the binomial distribution.

The model is further generalized by Wadsworth and Chang (1964) who include the scrap value of defective units in the objective function. They use incremental analysis and attempt to balance the marginal expected shortage cost and the marginal unit cost of an (additional) allowance. They develop a simple solution procedure which is based upon approximations to the marginal costs, and the assumption that the expected number of units short after the first production run is less than one.

White (1965) studies a version of this problem in which production runs are made until the order is satisfied. The objective is to minimize the total expected setup and variable production costs. Excess good units are assumed to be scrapped, but he suggests that the model can be generalized to incorporate salvage values. He gives a dynamic programming formulation for the case of a completely general yield probability mass function. White then formulates and analyzes a related problem in which the yield rate distribution is continuous, stationary over time, and independent of the input quantity. He establishes that Bellman's method of successive approximations (1957) converges under certain conditions on the yield rate distribution.

Further work on this problem was done by Delfausse and Saltzman (1966), who develop an exact (recursive) cost expression for total expected setup, variable production, and salvage costs for an order quantity as a function of the input quantity. They suggest that search procedures can be used to find the best input quantity for each order quantity. For various setup to variable cost to salvage cost ratios, they present optimal input quantities for a range of order quantities in the case of binomial yield quantities. Their numerical results suggest that the total expected cost function is relatively flat near the optimal solution.

Some subsequent work on the problem of satisfying a single known demand focuses on incorporating capacity constraints and other realistic considerations. Klein (1966) uses a Markovian decision framework to analyze two different problems. In the first, exactly N independent production runs are made to meet a future demand, and there is a capacity constraint on the number of units in each run. In the second, N is random because production can be terminated when the demand has been satisfied, and some dependence between production runs (such as learning effects) are permitted. The cost function in the first model includes production and inspection costs, holding costs, and shortage costs. They are permitted to be time dependent (i.e., varying with the index of the production run). The probability mass function of the number of good parts from a production run is allowed to vary with the production quantity and with the index of the production run. One unusual and interesting aspect of the models is that quantity tolerances are allowed, so shortage and overage costs are incurred only when the final quantity of good units lies outside a specified interval.

Klein models the first system (with N production runs) using a cyclic Markov chain where, after every sequence of N transitions, the system returns to its starting state. The cost function in the last of each of these N cycles represents the expected cost of excesses or shortages. Other costs are permitted to be completely general, but one must specify the expected cost of being in each state. Klein shows that this problem can be transformed into a linear program, which can then admit additional constraints. The problem with a random number (but no greater than N) production runs is modeled in a similar fashion, but with an absorbing state which is reached when demand has been satisfied or when N production runs have been completed, whichever occurs earlier. This solution procedure leads to randomize solutions, where the solution specifies the *probability* with which a particular action should be selected. Thus, the solution differs considerably from characteristics of most other solutions.

White (1967) examines situations in which the production of good units is a Bernoulli process, but the parameter p is unknown. He studies two cases: (1) total number of units processed is constrained (for example, because of raw material availability), and (2) the maximum production quantity in each production run is constrained. The solution procedure for the first case involves updating the distribution of p using standard Bayesian methods at the end of each production run. Then the optimal size for the next sublot is found using a linear program resulting from a cyclic Markov chain model. For the second case, White shows that the problem can be formulated as one (large) linear program, and suggests methods to reduce the state space so the linear program can be solved.

Possibly because of the generality of Klein's approach and the extensions of it by White, there appears to be a gap of nearly 10 years until the next paper on this problem was published. Beja (1977) examines systems having "constant marginal efficiency of production," which means that the unit cost of production divided by the probability of a good unit is constant over time. For constant unit production costs, the model is identical to a Bernoulli process. Using an objective function that includes unit production costs and setup cost, he presents a Markovian formulation where the decision at each stage is whether to produce another unit or to inspect the current output and re-setup if necessary. He proves some useful properties: (i) the optimal production quantity (in a production run) is monotonically increasing with the unsatisfied demand; and (ii) the objective function at every stage is a unimodal function of the production quantity. Beja then gives some conditions in which improved quality (with constant marginal efficiency) reduces costs, but also gives an example in which improved quality leads to increased costs.

Again, there appears to be a large time gap until the next article on this subject appeared. Sepehri, Silver, and New (1986) consider the Goode and Saltzman model (infinite number of production runs with setup costs) and formally formulate it as a dynamic program. They also consider the case of a finite number of setups, which is similar to Klein's model, but with stationary costs. Finally, they incorporate holding costs for good units produced "early" (i.e., before the final setup) into the finite setup model. They develop heuristics based upon incremental analysis, using Normal approximations to the Binomial distribution. A limited computation study suggests that the heuristics perform well, and that a simple rule-of-thumb policy of adjusting the production quantity to account for average yield losses performs poorly.

The Sepehri et al. heuristic is improved by Pentico (1988), who notes that there are two possible roots that solve a particular equation, and that in some cases Sepehri et al. choose the incorrect root. In a study of 360 different sets of problems, Pentico shows that the improved heuristic outperforms the rule-of-thumb policy tested by Sepehri et al., the marginal approaches of Bowman and Fetter (1961) and Wadsworth and Chang (1964), and a square root formula proposed by Orlicky (1973). This formula takes the lot size as the sum of demand and a constant multiplied by the square root of demand, where the multiplicative constant depends upon the probability of a defective unit and the setup and unit production costs.

Ilan and Yadin (1985) consider a single-period problem in which a known demand must be met, but lot splitting among different suppliers with random yields is allowed. This problem is similar to the one addressed by Gerchak and Parlar (1988) in an EOQ framework. The number of suppliers to be used, and their respective lot sizes are the decision variables of interest. This problem is extremely complex, and hence it is difficult to obtain structural results.

### 5.1.2 Multi-period Models

Relatively little research has been done on multi-period problems. All of the papers include assumptions of one production run per period, and linear holding and shortage costs, complete backlogging, and with one exception, zero lead times. Karlin (1958b) considers infinite horizon (steady state) problems where the objective is to minimize the average cost per period. Attention is restricted to policies having a single critical number representing the breakeven point of on-hand inventory between ordering and not ordering. He derives solutions for cases of demands having exponential and gamma distributions, and general continuous yield distributions that are permitted to depend upon the production quantity.

Gerchak, Vickson, and Parlar (1988) analyze a finite horizon problem with constant costs under the assumptions that the distribution of the fraction good is invariant with the production quantity and stationary in time, and that demand is stationary. Using a profit maximization objective, they show that myopic policies are not generally optimal, and that order-up-to policies are not optimal (i.e., the optimal production quantity is not necessarily a linear function of beginning-of-period inventory).

Henig and Gerchak (1987) further show that under similar assumption on the yield rate distribution, there exist critical order points for both the finite and infinite horizon problems. These critical order points are such that no order should be placed if the on-hand inventory level is above the critical order point; otherwise, an order should be placed. However, the order size is a complicated function of the system parameters. For the infinite horizon problem with stationary demand and costs, the critical order point is also stationary. Moreover, it is shown that this critical order point is greater than or equal to the one obtained when the yield is always 100%, i.e., the conventional model with no yield loss.

Yano (1986b) investigates both finite and infinite horizon problems in which demand is deterministic and the distribution of the yield rate is invariant with the input quantity. It is shown that under certain conditions the optimal policy is *multiplicative*. That is, the optimal input quantity is a real-valued multiple of net demand in a period. The conditions are essentially: (1) no speculative motive for holding inventory; (2) maximum yield rate is less than two times the minimum yield rate; and (3) the product is profitable to produce.

Mazzola, McCoy, and Wagner (1987) develop several heuristic policies for finite horizon problems with setup costs and deterministic demand. They permit completely general (discrete) yield quantity distributions in their formulations, but the computational study is based upon binomial yields. The better-performing heuristics in the first experiments involve finding a lot sizing solution using the Wagner-Whitin (1958) or Silver-Meal (1960) algorithm, then scaling up the lot sizes by multiplying by the reciprocal of the average yield rate. They also test modified heuristics based upon service level concepts (probability of satisfying demand) and report that the average cost error is 2.8%.

A heuristic procedure for the stationary finite horizon problem with setup costs, and with positive lead times is proposed by Ehrhardt and McClelland (1987). They suggest a method in which one first computes the two critical numbers (s and S) ignoring yield losses using standard techniques. Then, an approximate inventory position, y, is computed as the sum of on-hand plus the expected good in-transit inventory. If this value is less than s, an order is placed for a quantity equal to (S - y)/[1 - E(P)], where P is the random variable representing the fraction defective. They compare this heuristic with the best solution found from a search using simulation and the results indicate that the heuristic performs quite well.

5.2 Facilities in Series

Most of the papers on facilities in series consider single-period problems, and to date, only heuristic procedures are available for multi-period versions of the problem. Vachani (1970) considers a single period model of a serial production system in which all demand must be satisfied. He assumes that all of the good output of one stage is processed by the succeeding stage. Consequently, the only decision is the initial input quantity. At each stage a setup cost, and linear variable production and salvage costs are incurred. The yield distributions are assumed to have the characteristic that the *expected* fraction of units good is invariant with the batch size. For tractability, he makes the assumption that the total setup cost up through stage n is much larger than the variable production cost through stage n. This permits him to assume that there is at most one supplementary production run. He also assumes that the input quantity to the supplementary run is the shortage amount divided by the average yield rate. Under these assumptions, Vachani shows that the objective function is unimodal in the initial input quantity.

Vachani then considers a situation where more than one production run can be made at a stage before proceeding with the subsequent stage. Again, all good output of a stage is processed at the succeeding stage. He shows that under certain assumptions and conditions, the cost function is convex in the initial input quantity in the case of binomial yield quantities. He also points out difficulties in establishing convexity and monotonicity of various key functions in general.

Lee and Yano (1988) address a single-period problem for a serial system with deterministic demand and no setup costs. The distribution of the yield rate at each production stage is assumed to be invariant with the input quantity and independent of the yield rate distributions at the other stages. They assume that processing, holding and shortage costs are linear, that it is profitable to product the product, and that it is less expensive to hold inventory at one stage than to process it and hold the expected good output at the following stage. Under these conditions, they show that there is a single critical number representing the optimal target input quantity at each stage. These values can be determined in a sequential fashion, starting with the last production stage. The optimal policy is to input the target quantity if enough is available; otherwise, one should input whatever is available. These structural results are extended to the case of random demand in Yano (1986a).

Yano (1986c) generalizes the model of Lee and Yano by incorporating positive setup costs. The optimal policy for the one-stage problem is shown to have two critical numbers. The larger critical number is the optimal target input quantity, and the smaller critical number is the value of on-hand inventory below which it is preferable not to produce. If on-hand inventory lies between the two critical numbers, it is optimal to input everything. For two stages, the optimal policy has the same form, but only under certain restrictive conditions.

In the above papers, it is assumed that defective items at any production stage are scrapped. Spence (1988) extends the model of Lee and Yano to allow rework of the defective items at some production stages, at a cost. She assumes that all defective items are acceptable after being reworked. Her work is motivated by the wafer fabrication process for Application-Specific Integrated Circuits. There are multiple production stages in such a process, and the goal is to meet a target output level. Observing that the yield density function of the photolithography workcell of such a process is U-shaped in the lot size, Spence proposes a yield density function for each production stage which is a general polynomial function of the input quantity. Even with this complicated yield representation, the property of constant bias factor is preserved in this formulation. Moreover, the standard deviation of the yield is still a linear function of the input quantity. Hence, although very different in appearance, the polynomial representation of the yield density function has structural properties similar to those used by other researchers.

In Spence's model, after any given production stage n where rework is permissible, it is assumed that the number of items that have been processed at this stage, $Q_n$, and the number of "good" items among them, $y_n$, are known. The $Q_n$-$y_n$ items can be reworked. The basic result is that the input quantity to the next stage is characterized by two threshold quantities, $a_n$ and $b_n$. If $y_n$ falls below the lower threshold $a_n$, then the input quantity should be the minimum of $Q_n$ and $a_n$. Rework is used as necessary to arrive at this input quantity. If $y_n$ is above the lower threshold $a_n$, then the input quantity should be the minimum of $y_n$ and $b_n$. In this case, no rework is necessary.

Another model that is similar to a production system in series is described by Bassok and Akella (1988). It consists of two stages, one corresponding to raw material procurement, and the other to actual production. The yield uncertainty is represented by the quality of raw material supply, which is random. The problem is to simultaneously determine the amount of raw materials to purchase and the amount of machine capacity to reserve for production of multiple parts on the same machine, to satisfy demand for a single period. Given the amount of raw materials ordered, the amount that actually arrives is a random variable. The cost function for such a problem can be quite complex.

As an approximation, Bassok and Akella assume that the decision variables should be set such that the probability of material shortage for production is close to zero. (Implicit in this is the assumption that underutilization of capacity is very undesirable.) In this way, a simpler and tractable model results. Of interest is that Bassok and Akella show that their model results in producing smaller quantities than those obtained by solving the two decision problems separately. Hence, yield uncertainties in raw material supply appear to result in less production.

A recent study by Tang (1988) models three types of uncertainty in a multi-stage production line: yield, production rate, and demand. Buffer stocks are allowed between the stages, and the system operates as a "pull" system. The decision in each period is the production level. The proposed production rule starts with some target values for the buffer stocks at the various stages, and at the beginning of each period, restores the buffer stock at each stage to its target value in expectation. It thus minimizes the expected deviation of the buffer stock levels from their targets. However, the determination of target values is not explored.

### 5.3 Assembly Systems

Even less research has been done on assembly systems. Here, the amount of end-product that can be assembled is the minimum of the yields (in terms of quantity) of all the components. Thus far, only single period problems have been considered. Yano and Chan (1989) consider a situation in which simultaneous procurement decisions must be made for several assembly components with mutually independent random yield rates, where the yield rate distributions are independent of the lot sizes. They investigate properties of optimal solutions for the two-component case and develop heuristic procedures based on these properties. One heuristic policy which involves a search over a single parameter and has a newsboy-type form, is shown to perform quite well.

Versions of this problem with no salvage values for excess unassembled components are considered by Gerchak and Yano (1989). This assumption regarding salvage values permits them to retain concavity of the expected profit function while including random demand, positive assembly costs, and random assembly yields. Optimality conditions for various versions of the problem are derived.

A related problem with a different objective is considered by Yao (1988). The production or procurement cost for a component is assumed to be increasing convex in the input quantity of that component. The problem is to minimize total production costs, subject to the constraint on the probability that a target demand level is met. Yao makes the common assumption that the yield rate, p, is independent of the input batch size. Moreover, he also assumes that this yield rate for each component has the increasing failure rate (IFR) property, i.e., the hazard rate, defined as $g(p)/[1 - G(p)]$, where $g(p)$ and $G(p)$ are the p.d.f. and c.d.f. of p respectively, is increasing in p. Under such assumptions, the problem becomes a convex program, and a standard Lagrange multiplier approach can be used to solve it. Another performance measure, the expected yield, which Yao defines as the number of end-products that can be assembled, is usually very difficult to derive analytically. Yao considers the special case where the component yields follow shifted exponential distributions, and obtains the expected yield in closed form. The rationale for the use of shifted exponential distributions is that IFR random variables have coefficients of variation less than one, and can thus be approximated by shifted exponential random variable using a two-moment fit.

Lee (1987) considers a similar problem in which the procurement decisions are made sequentially. Again, the yield rates of the individual components are mutually independent and the yield rate distributions are independent of their respective lot sizes. A dynamic programming procedure is used to find optimal solutions, taking into account the actual yield rates associated with components already procured. The optimal sequence in which to make the procurement decisions can be found efficiently using a dynamic program. The solution procedures can be extended to include random demand, and random yields in the assembly process.

Singh et al. (1988) describe a problem in semiconductor wafer fabrication where the yield characterization is quite complicated because of the nature of the manufacturing process. Sites on wafers must be allocated to various chips and individual chips or entire wafers may be defective. The problem is to maximize the probability of meeting a set of orders while considering the wafer processing costs. The two types of decisions that must be made are: (i) how to allocate sites on wafers to the various types of chips and (ii) how many wafers should constitute a job. The first problem is formulated as one of maximizing the probability that the required chips are acceptable given a constraint on the number of sites. This nonlinear program is solved for various values of the number of sites. The second problem considers the impact of entire wafers being defective. Here, the problem is to choose the minimum number of wafers consistent with meeting demand for the set of orders. This problem is solved by iteratively increasing the number of wafers (thereby increasing the number of available sites) until the desired probability is attained.

## 6. COMPLEX MANUFACTURING SYSTEMS

In this section, we review modeling effort for more complex manufacturing systems. These complexities include real time dispatching capabilities, imperfect inspection, rework, and capacity-constrained production systems.

White (1970) shows how absorbing Markov chains can be used to determine the expected variable production cost per unit for some systems that can include multiple stages of production and rework. The expected variable production cost per unit is then used as a cost parameter in a single-period profit optimization model in which only the initial input quantity can be decided. In an example, he applies these concepts to a system with two production stages and the possibility of rework.

Akella and Rajagopalan (1987) study the problem of dispatching and batch sizing for a facility that produces multiple parts. The reason for batching in this context is the presence of positive setup times for the test equipment. Defective parts are reworked offline, and are sent through the testing stage again. Using a control theoretic framework, they attempt to minimize the sum of the cost of deviations of production from demand (which indirectly minimizes inventory holding and shortage costs) and the cost of resource over- and underutilization (which indirectly minimizes overtime costs). The cost function is assumed to be convex (quadratic), and this leads to optimal linear decision rules. Similar approaches are used for a facility consisting of a flexible assembly or manufacturing system. The production of printed circuit boards is used as an example (see Akella, Singh, and Bassok 1988).

Yano (1989) considers a finite horizon model with no setups costs where in each period, a "normal" and a "rework" production run are allowed (in that sequence). The variable cost for reworking a part is assumed to be less than the cost of producing one in a normal run. It is shown that the optimal production policy for the normal production run depends heavily upon the number of good and reworkable units on hand. Consequently, the optimal production quantity for the normal production run may be positive even when there are many reworkable parts on hand. Sufficient conditions are derived for optimality of "no rework" and "rework all" policies.

Motivated by imperfections of the wafer probe operation in semiconductor manufacturing, Lee (1989) incorporates the time required for rework explicitly in determining the optimal lot size for wafer fabrication. Both fixed setup and variable processing times for rework are included. It is assumed that the wafer probe process starts in the "in control" state and that the time until it becomes out-of-control has a geometric distribution. There is random self-detection of out-of-control problems during the production run. For the objective of minimizing the average total processing time per wafer, closed-form expressions for the optimal batch size are derived. Of interest is that the fixed rework times tend to favor large lots, while the variable rework times tend to favor large lots. The optimal lot size trades off these two effects.

Relatively little work has been done on more complicated systems. Meal (1979) suggests a heuristic for setting safety stock levels in Material Requirements Planning systems. He recommends that the variance of the usage of each part (resulting from both demand fluctuations and supply or yield fluctuations) be used in a single-stage inventory model to compute a safety stock level for that part.

Karmarkar and Lin (1986) consider a multi-period production planning problem in which demands and yields are random. The objective is to allocate resources (both consumable and renewable) to N products to minimize the sum of expected variable production, inventory holding, shortage, regular time, overtime, capacity acquisition, and capacity retirement costs. A linear cost structure is assumed. They develop methods to compute upper and lower bounds on the objective value. Computational tests indicate that the lower bounds are fairly tight. The lower bounds are based upon a Lagrangian relaxation which decomposes the problem into independent single-period convex programs, with multipliers linking the periods. Karmarkar and Lin suggest that the procedures on which the lower bounds are based might provide effective heuristic solutions.

Yano (1987) addresses the product cycling problem for products with random yields. Two aspects of the problem are considered. In the first, the concern is one of finding the optimal cycle duration for a *planned* rotation schedule. An approximate cost function is used, which is shown to be convex in the cycle duration, provided the optimal production quantities are used for the corresponding cycle duration. The solution is obtained using fixed point methods. A method to allocate capacity to the products during the current cycle is also presented.

Extensions of the continuous time models of section 4 to multiple items are complicated. Based on the model of Lee and Rosenblatt (1988) in which the time until the system goes out of control has an exponential distribution, Muckstadt (1988) considers the joint problem of determining lot sizes and inspection intervals with multiple items processed on the same machine. Here, it is assumed that the inspection process is perfect. An additional constraint is included to ensure that the total setup and processing time does not exceed the time available for production. The basic results of the problem are very similar to those of Lee and Rosenblatt. The various products, however, may have very different production cycles. Muckstadt treats the sequencing and scheduling problems heuristically, and assumes that a power-of-two policy is used for both the production cycles and inspection intervals. This leads to very tractable solution procedures for these two decision variables.

Muckstadt also extends this model framework to the case where the inspection process is imperfect. When the production process is out of control, it is assumed that the detection time is exponentially distributed. Such an assumption essentially is equivalent to the assumption of a constant probability that an out-of-control process is detected.

We note here that the paper by Spence (1988) described in section 5.2 also falls into this category.

## 6.0 ANALYSIS AND DISCUSSION

The literature review makes evident the sparsity of research on realistic discrete time models with multiple time periods, multiple production stages, capacity constraints and the resulting production lead times, and multiple production runs at a stage within a single time period. The discrete time models have the advantage of being able to incorporate non-stationarity of yields, demand, and costs. This can be especially important during the early portion of a process (and product) life cycle.

The main drawbacks of existing models of continuous time systems are the assumptions of stationary demand, unlimited capacity, and zero or constant lead times. The models are, however, more tractable analytically and computationally than discrete time models, and so it may be easier to extend them to more general settings.

The issue of variable yields in multi-stage systems (specifically MRP systems) was pointed out by Wagner (1980), but it is only in the last few years that progress has been made in this area. We believe that the reason for this is that problems with random yields become enormously complex, both theoretically and computationally, as the size and complexity of the underlying manufacturing system increase. This is true partly because input and output quantities are different, whereas they are the same when only demand is uncertain.

Because of problem complexity, much of the research in the past two years has focused on heuristics that have their foundations in simple optimization models. We believe that most realistic problems cannot be solved optimally, and often, it will even be difficult to derive the *form* of the optimal policy. Thus, there is a need for heuristic solution procedures that are computationally inexpensive, and easy to implement. Moreover, existing heuristics need to be tested over wider ranges of parameter values and especially over a greater variety of yield rate and demand distributions.

There is also a need for appropriate performance measures for heuristics. Several papers report heuristic solutions that give costs within a few percentage points of the minimum cost solution. It is important to realize that because of the nature of the problem, variable production or procurement costs must be included in the objective function either explicitly or implicitly. In most cases, the unavoidable variable production costs constitute a major portion of total costs. Thus, although a heuristic procedure may provide a solution within a few percentage points of the minimum *total* cost, it may provide poor performance relative to minimizing *controllable* costs. Consequently, questions must be raised about appropriate performance measures for heuristics.

Finally, there is still a great deal of controversy about how randomness of yields should be modeled. We believe that there is *not* one correct way of modeling yield variances, but that each of the modeling approaches described in Section 2 is applicable to a class of applications. Further empirical work is needed to describe commonly observed yield distributions, and additional computational work is needed to test the robustness of existing solution procedures to misspecification of the form of the yield distribution.

Acknowledgment

The work of Candace A. Yano was supported in part by National Science Foundation Grant 85-04644 to the University of Michigan.

# REFERENCES


1. Afentakis, P. and B. Gavish (1986), "Optimal Lot-Sizing Algorithms for Complex Product Structures," *Operations Research* 34(2), 237-249.

2. Akella, R. and S. Rajagopalan (1986), "Part Dispatch in Random Yield Multi-Stage Flexible Test Systems for Printed Circuit Boards," Working paper, Graduate School of Industrial Administration, Carnegie-Mellon University.

3. Akella, R., M.R. Singh, and Y. Bassok (1988), "Real Time Part Dispatch in Flexible Assembly, Test and Manufacturing Systems, " Working paper, Graduate School of Industrial Administration, Carnegie-Mellon University.

4. Arrow, K.J., S. Karlin, and H. Scarf (1958), Studies in the Mathematical Theory of Inventory and Production, Stanford, CA: Stanford University Press.

5. Banks, J. and Fabrycky, W.J. (1986), Procurement and Inventory Systems Analysis, Englewood Cliffs, NJ: Prentice-Hall, Inc.

6. 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.

7. Basu, S.K. (1987), "An Optimal Ordering Policy for Situations with Uncertainty in Supply," *Naval Research Logistics* 34(2), 151-160.

8. Beja, A. (1977), "Optimal Reject Allowance with Constant Marginal Production Efficiency," *Naval Research Logistics Quarterly* 24(1), 21-33.

9. Bellman, R. (1957), Dynamic Programming, Princeton University Press, Princeton, NJ.

10. Bowman, E. H. and R.B. Fetter (1961), Analysis for Production Management, Homewood, IL: Richard D. Irwin.

11. Cheng, T.C.E. (1989), "EPQ with Process Capability and Quality Assurance Considerations," Working Paper, Dept. of Actuarial and Management Sciences, University of Manitoba.

12. Delfausse, J., and S. Saltzman (1966), "Values for Optimum Reject Allowances," *Naval Research Logistics Quarterly* 13(2), 147-157.

13. Duncan, A.J. (1956), "The Economic Design of $\bar{X}$ Charts Used to Maintain Current Control of a Process," *J. of American Statistical Association*, 51, 228-242.

14. Ehrhardt, R. and M. McClelland (1987), "Inventory Management with Random Replenishments," paper presented at the ORSA/TIMS Conference in New Orleans, May 1987.

15. Ehrhardt, R. and L. Taube (1987), "An Inventory Model with Random Replenishment Quantities," *International Journal of Production Research* 25(12), 1795-1804.

16. Gerchak, Y. and M. Parlar (1988), "Yield Variability/Cost Tradeoffs and Diversification in the EOQ Model," Working Paper, Department of Management Sciences, University of Waterloo.

17. Gerchak, Y., M. Parlar, and R. G. Vickson (1986), "A Single Period Production Model with Uncertain Output and Demand," Proceedings of the 25th IEEE Conference on Decision and Control, 1733-1736.

18. Gerchak, Y., R.G. Vickson, M. Parlar (1988), "Periodic Review Production Models with Variable Yield and Uncertain Demand," *IIE Transactions* 20(2), 144-150.

19. Gerchak, Y. and C.A. Yano (1989), "Lot Sizing in Assembly Systems with Random Component Yields," Technical Report 89-9, Department of Industrial and Operations Engineering, University of Michigan.

20. Giffler, B. (1960), "Determining an Optimal Reject Allowance," *Naval Res. Log. Quarterly* 7(2), 201-206.

21. Goode, H.P. and S. Saltzman (1961), "Computing Optimum Shrinkage Allowances of Small Order Sizes," *Journal of Industrial Engineering* 12(1), 57-61.

22. Grant, E.L. and R.S. Leavenworth (1988), Statistical Quality Control, McGraw-Hill Book Co., New York.

23. Graves, S. C. (1987), "Safety Stocks in Manufacturing Systems," *Journal of Manufacturing and Operations Management* 1(1), 67-101.

24. Gregory, W.R. and A. Beged-Dov (1967), "On the Determination of Optimal Shrinkage Allowance in a Job Shop," *Journal of Industrial Engineering* 17(4), 284-288.

25. Hadley, G. and T.M. Whitin (1963), Analysis of Inventory Systems, Prentice Hall, Englewood Cliffs, NJ.

26. Henig, M. and Y. Gerchak (1987), "The Structure of Period Review Policies in the Presence of Variable Yield," Working Paper, Department of Management Sciences, University of Waterloo.

27. Hillier, F.S. (1963), "Reject Allowances for Job Lot Orders," *Journal of Industrial Engineering* 14, 311-316.

28. Ilan, Y. and M. Yadin (1985), "Second Sourcing and Transfer of Technology--Tradeoffs Under Uncertainty," Working paper, Department of Operations Research, Statistics, and Economics, The Technion, Israel.

29. Juran, J. and F. Gryna (1980), *Quality Planning and Analysis*, McGraw Hill Book Co., New York.

30. Kalro, A. H. and M.M. Gohil (1982), "A Lot Size Model with Backlogging when the Amount Received is Uncertain," *International Journal of Production Research* 20(6), 775-786.

31. Karlin, S. (1958a), "One Stage Models with Uncertainty," Chapter 8 in [3].

32. Karlin, S. (1958b), "Steady State Solutions," Chapter 14 in [3].

33. Karmarkar, U. and S. Lin (1986), "Production Planning with Uncertain Yields and Demands," Working paper, William E. Simon Graduate School of Business Administration, University of Rochester.

34. Klein, M. (1966), "Markovian Decision Models for Reject Allowance Problems," *Management Science* 12(5), 349-358.

35. Lee, H.L (1987), "Lot-Sizing for Assembly Systems with Variable Yields," Paper presented at the 1987 Multi-Echelon Inventory Conference, Wharton School, University of Pennsylvania.

36. Lee, H.L. (1989), "Lot-Sizing in a Production Process with Defective Items, Process Correction, and Rework," Working Paper, Dept. of Industrial Engineering and Engineering Management, Stanford University.

37. Lee, H.L. and M.J. Rosenblatt (1988), "Simultaneous Determination of Production Cycle and Inspection Schedules in a Production System," *Management Science* 33(9), 1125-1136.

38. Lee, H.L. and M.J. Rosenblatt (1985), "Optimal Inspection and Ordering Policies for Products with Imperfect Quality," *IIE Transactions*, 17, 284-289.

39. Lee, H.L. and M.J. Rosenblatt (1987), "A Production and Maintenance Planning Model with Restoration Cost Dependent on Detection Delay," *IIE Transactions* (to appear).

40. Lee, H. and C.A. Yano (1988), "Production Control for Multi-Stage Systems with Variable Yield Losses," *Operations Research* 36(2), 269-278.

41. Levitan, R.E. (1960), "The Optimum Reject Allowance Problem," *Management Science* 6(2), 172-186.

42. Lin, M.J., S.T. Tseng and T.M. Lin (1988), "Optimal Production Cycle Length in the Imperfect Production System, " Working Paper, Department of Industrial Engineering, National Taiwan Institute of Technology.

43. Llewellyn, R.W. (1959), "Order Sizes of Job Lot Manufacturing," *Journal of Industrial Engineering* 10(3), 176-180.

44. Love, S.F. (1979), Inventory Control, McGraw-Hill Book Co., New York.

45. Mak, K.L. (1985), "Inventory Control of Defective Product when the Demand is Partially Captive," *International Journal of Production Research* 23(3), 533-542.

46. 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.

47. Meal, H.C. (1979), "Safety Stocks in MRP Systems," Technical Report 166, Operations Research Center, MIT.

48. Moinzadeh, K. and H.L. Lee (1987), "A Continuous Review Inventory Model with Constant Resupply Time and Defective Items," *Naval Research Logistics* 34(4), 457-468.

49. Moinzadeh, K. and H.L. Lee (1989), "Approximate Order Quantities and Reorder Points for Inventory Systems Where Orders Arrive in Two Shipments," *Operations Research* 37(2) 277-287..

50. Muckstadt, J. (1988), "Establishing Reorder Intervals and Inspection Policies when Production and Inspection Processes are Unreliable," Working Paper, School of OR and IE, Cornell University.

51. Muckstadt, J.A. and R. Roundy (1987), "Multi-Item, One-Warehouse, Multi-Retailer Distribution Systems," *Management Science* 33(12), 1613-1621.

52. New, C. and J. Mapes (1984), "MRP with High Uncertain Yield Losses," *Journal of Operations Management* 4(4), 315-330.

53. Noori, H. and G. Keller (1986), "The Lot-Size Reorder Point Model with Upstream-Downstream Uncertainty," *Decision Sciences* 17, 285-291.

54. Noori, H. and G. Keller (1986), "One-Period Order Quantity Strategy with Uncertain Match Between the Amount Received and the Quantity Requisitioned," *INFOR* 24(1), 1-11.

55. Orlicky, J.B. (1975), <u>Material Requirements Planning</u>, Mc-Graw Hill Book Co., New York, 134-135.

56. Panda, R.M. (1978), "Static Inventory Model with Uncertainty in Supply," *Calcutta Statistical Association Bulletin*, 27, 141-148.

57. Pentico, D.W. (1988), "An Evaluation and Proposed Modification of the Sepehri-Silver-New Heuristic for Multiple Lot Sizing Under Variable Yield," *IIE Transactions* 20(4), 360-363.

58. Porteus, E. L. (1986), "Optimal Lot Sizing, Process Quality Improvement, and Setup Cost Reduction," *Operations Research* 34(1), 137-144.

59. Porteus, E.L. (1988), "Optimal Inspection, Lot Sizing and Setup Reduction," Working Paper, Graduate School of Business, Stanford University.

60. Rosenblatt, M.J. and H.L. Lee (1986), "A Comparative Study of Continuous and Periodic Inspection Policies in Deteriorating Production Systems," *IIE Transactions* 18(1), 2-9.

61. Rosenblatt, M.J. and H.L. Lee (1986), "Economic Production Cycles with Imperfect Production Processes," *IIE Transactions* 18(1), 48-55.

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

63. Shih, W. (1980), "Optimal Inventory Policies when Stockouts Result from Defective Products," *International Journal of Production Research* 18(6), 677-685.

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

65. Silver, E.A. and H.C. Meal (1969), "A Simple Modification of the EOQ for the Case of a Varying Demand Rate," *Production and Inventory Management* 10(4), 52-65.

66. Singh, M.R., C.T. Abraham, and R.Akella (1988), "Planning for Production of a Set of Components When Yield is Random," *Proceedings of the IEEE International Electronic Manufacturing Symposium.*

67. Spence, A. (1988), "Yield Variability in Manufacturing: Rework and Scrap Policies," Unpublished Ph.D. dissertation, Graduate School of Business, Stanford University.

68. Tang, C.S. (1988), "The Impact of Uncertainty on a Production Line," Working paper, Graduate School of Business Administration, UCLA.

69. Vachani, M. (1970), "Determining Optimum Reject Allowances for Multistage Job-Lot Manufacturing," *IIE Transactions* 2(1), 70-77.

70. Vachani, M. (1969), "Determining Optimum Reject Allowances for Deteriorating Production Systems," *Naval Research Logistics Quarterly* 16(3), 275-286.

71. Wadsworth, H.M. and S.H. Chang (1964), "The Reject Allowance Problem: An Analysis and Application to Job Lot Production," *Journal of Industrial Engineering* 15 (3), 127-132.

72. Wagner, H.M. (1980), "Research Portfolio for Inventory Management and Production Planning Systems," *Operations Research* 28(3), 445-475.

73. Wagner, H.M. and T.M. Whitin (1958), "Dynamic Version of the Economic Lot Size Model," *Management Science* 5(1), 89-96.

74. White, D.J. (1965), "Dynamic Programming and Systems of Uncertain Duration," *Management Science* 12(1), 37-67.

75. White, J.A. (1970), "On Absorbing Markov Chains and Optimum Batch Production Quantities," *AIIE Transactions* 2(1), 82-87.

76. White, L.S. (1967), "Bayes Markovian Decision Models for a Multi-Period Reject Allowance Problem," *Operations Research* 15(5), 857-865.

77. Whybark, D.C. and J.G. Williams (1976), "Material Requirements Planning Under Uncertainty," *Decision Science* 7(4), 595-606.

78. Yano, C.A. (1986a), "Controlling Production in Serial Systems with Uncertain Demand and Variable Yields," Technical Report 86-5, Department of Industrial and Operations Engineering, University of Michigan.

79. Yano, C.A. (1986b), "Optimal Finite and Infinite Horizon Policies for a Single Stage Production System with Variable Yields," Technical Report 86-32, Department of Industrial and Operations Engineering, University of Michigan.

80. Yano, C.A. (1986c), "Optimal Policies for a Serial Production System with Setup Costs and Variable Yields," Technical Report 86-24, Department of Industrial and Operations Engineering, University of Michigan.

81. Yano, C.A. and T.J. Chan (1989), "Production and Procurement Policies for an Assembly System with Random Component Yields," Working Paper, Department of Industrial and Operations Engineering, University of Michigan.

82. Yano, C.A. (1987), "The Product Cycling Problem in Systems with Uncertain Production Yields," Technical Report 87-32, Department of Industrial and Operations Engineering, University of Michigan.

83. Yao, D.D. (1988), "Optimal Run Quantities for an Assembly System with Random Yields," *IIE Transactions* 20(4), 399-403.

84. Zhang, X. and Y. Gerchak (1988), "Joint Lot Sizing and Inspection Policy in an EOQ Model with Variable Yield," *IIE Transactions* (to appear).

| Review | Process Configuration | Time Horizon | Demand | Section |
|---|---|---|---|---|
| Continuous | Single Stage | Infinite | Constant and Deterministic | 4.1 |
| | | | Random | 4.2 |
| | Multiple Sources | Infinite | Constant and Deterministic | 4.2 |
| Periodic | Single Stage | Single Period, Single Order or Production Run | Deterministic or Random | 5.1.1.1 |
| | | Single Period, Multiple Production Runs | Deterministic or Random | 5.1.1.2 |
| | | Multiple Periods | Deterministic or Random | 5.1.2 |
| | Facilities in Series | Single or Multiple Periods | Deterministic or Random | 5.2 |
| | Assembly Systems | Single Period | Deterministic or Random | 5.3 |
| Continuous and Periodic | Complex | Various | Constant, Deterministic or Random | 6 |

**Figure 1**

**Taxonomy of Random Yield Problems**