MANAGEMENT SCIENCE
Vol. 14, No. 8, April, 1968
Printed in U.S.A.

JOB-SHOP LOT RELEASE SIZES*†

GEORGE SCHUSSEL

Brown Engineering, A Teledyne Company, Huntsville, Alabama,

A new model (ELRS) for determining optimal job-shop lot sizes is presented. Since the setting of the lot release size is one of the most important controlling parameters in the scheduling of a job-shop, the improvement in technique described in the article should result in substantial savings to job-shop operation costs. In those cases where management is already using an EOQ method for determining lot release sizes, the input data and system required for the implementation of ELRS is basically similar and therefore the new and improved analysis can be substituted at relatively minor inconvenience to the user. For those job-shops which set lot release sizes by informal means, the ELRS concept may provide enough of an improved technique to induce this use of this analysis.

1. Objectives

This paper discusses a general economic model which can be used to determine optimum production lot sizes. The objective of the development of this model was to develop an analytical tool which could balance the general categories of production, setup, inventory and holding costs so that a job shop could produce efficiently and maintain inventory at the lowest possible level. The model is quite general and can apply to any regularly scheduled job shop type of operation requiring repetitive releases of the same part or subassembly.

The model presented in this paper was derived exclusively to solve the production problem. Some companies have used standard Economic Order Quantity (EOQ)[1] analysis in the production area. The EOQ model was derived for the problem of ordering from an outside vendor, not for inhouse production, which has very different problems. The Economic Lot Release Size (ELRS) model is specifically designed to solve the problem of lot sizes in a production environment.

2. Description of Model

The model presented in this report has the capability of identifying an optimal policy for a production inventory problem where the costs are four different broad types. For reference, these costs are called Production, Warehouse, Capital and Setup; however, all costs involved in such a process, such as machine setup, inventory carrying, direct labor, variable overhead, storage, paper work, interest, insurance, material and taxes are handled by placing the costs in appropriate categories.

The heart of the ELRS model is a trapezoidal figure which is used to represent dollars of inventory value from the start of production on a release to the usage of

* Received September 1966; revised August 1967.

† The work on which this paper is based was done at Northrop Corporation, where the author directed the work of the Management Science Staff.

[1] Many references discuss EOQ. For example see [2].

the last piece on that release. The EOQ model is a triangle, where it is assumed that production, or an order, happens at one instant and a consumption occurs linearly after that point. Figure 1 illustrates this.

The ELRS model explicitly recognizes that production occurs over a span of time, that there is some holding time from the end of production to the start of consumption, and that consumption occurs over time. This article takes the point of view that production output from the job shop is in the form of parts and subassemblies that are subsequently consumed by a production line making final products. The mathematical model presented here, however, is very general and may be interpreted in other fashions (such as consumption representing the sales of a product). The holding time refers to the safety and travel time that is typi-

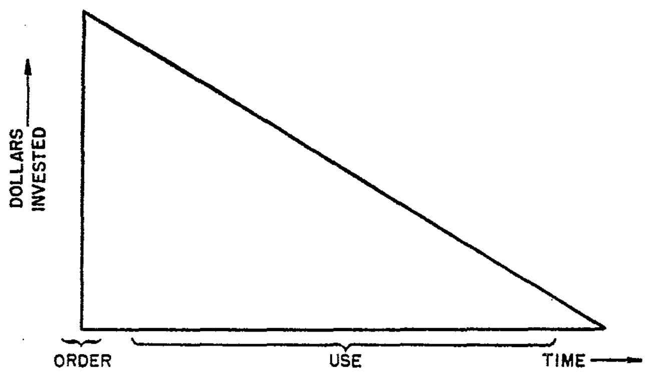

Fig. 1. EOQ Model of Inventory Value

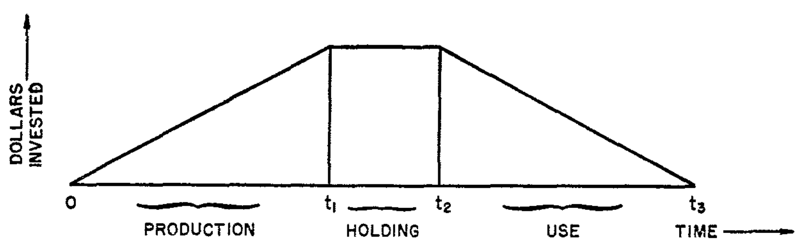

Fig. 2. ELRS Model of Inventory Value

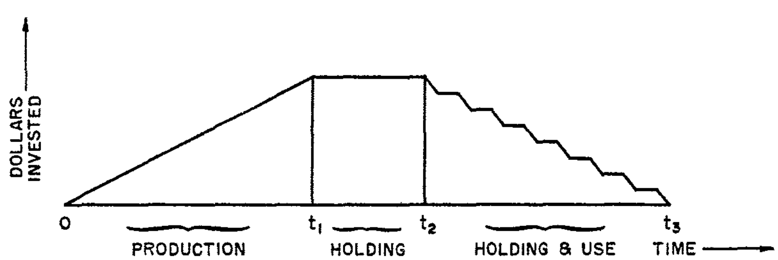

Fig. 3. ELRS Model of Inventory Value

cally scheduled between the end of production and the start of consumption (on a production line or elsewhere). Figure 2 illustrates this.

If the part is one that is used in later subassemblies, ELRS has the capability of representing the staggered repeated usage shown by Figure 3. This staggered usage model is analyzed in Section 9.

3. Different Models of Discounting Cash Flows

Using the trapezoidal model, this report goes into three analyses of the general problem, each successively more complex in its handling of the discounting of future cash flows.

Model A is the general ELRS model, except for ignoring the concept of discounted future cash flows. This model is very useful in itself because unless the production and use of a lot release occur over a very long period of time (over a year or two), or the cost of capital is very high (100% per year or so), the recommended production lot sizes of Model A are very close to those of the more complex models.

Model B uses the approximation of solving for equivalent median time periods and treating the cash flows as if they occurred at their medians. This method is only an approximation to correct discounting because it weighs cash flows that occur in the latter half of the cycle too heavily.

Model C uses a discounting factor and integrates continuously discounted cash flows over time.

The output recommendations from all three models are suggested lot sizes. Because of the fact that the cost of not having the exactly optimal production sizes is very small unless the error is large, the output from the analysis should be treated as a recommendation and not as an absolute fiat. Considerations of workload balancing should be applied to the recommendations made by this system and production lot sizes within 50% of the recommended lot size will usually be efficient. Once the production lot size for a part is determined, it does not need to be redetermined every time the part is made. An updating, based upon new cost information, is the only reason to rerun the system on a part.

4. Term Definitions

- C The total cost/unit which we are trying to minimize. This includes production and carrying costs.
- C' The equivalent cost figure which is a function of Q and which, when it is minimized with respect to "Q", also determines the optimal release quantity.
- D The decimal rate which is the sum of carrying and capital percentage costs on a basis of percentage of value/unit time.[2]
- E The decimal rate cost per unit time which is the sum of carrying and warehousing costs on the basis of size.[2]

[2] These terms are required as input for the computerized system which implements the analysis described in this paper.

F The final storage area requirements of one complete unit.[2]

K The total dollar cost of the capital/unit of any production inventory run. It is proportional to the value of the item.

P The sum of the variable production costs over the various required machine operations for the entire production cycle. This includes material, labor, and machine costs.

Q The size of the increment release. "q" will be used as the corresponding running variable under an integral sign.

R The usage rate of the finished part, in units/unit time.

r The cost of capital as a decimal rate of invested funds.[2]

S The sum of the setup costs in dollars required for all of the production machine operations. This includes such items as the paperwork costs of writing an order.[2]

t_1 The time at the end of all production processes on the increment release, (with start time $\equiv 0$)[2].

t_2 The time at the start of the production line use of the increment release. With the theory used in the report, $t_2 = t_1 +$ (the scheduled holding period from the end of production to the start of usage).[2]

t_3 The time at the end of the use period of the increment release. It is equal to $(t_2 + Q/R)$.

t_m The derivable middle point in our production, hold and use cycle such that

$$\int_0^{t_m} W\,dt = \int_{t_m}^{t_3} W\,dt \quad \text{or} \int_0^{t_m} K\,dt = \int_{t_m}^{t_3} K\,dt.$$

V An approximation to the total production costs for each item. "V" is equal to $(P/Q + S/Q)$.

W The carrying cost/unit including warehousing, taxes, insurance, handling, and risk of obsolescence. Some of these factors are proportional to the size and others to the value of the item.

5. Model A

The various costs of repetitively manufacturing any given increment quantity will vary as shown in Figure 4 with respect to lot size. The downward slope of the P curve reflects the fact that as the lot size increases, the average production cost per unit decreases. This is due to the inherent efficiencies of long production runs. The downward slope of the S curve simply reflects the fact that the same setup charges are being distributed over more production units as the lot size increases. The upward sloping W curve is due to the fact that as the production lot size gets greater, the average amount of inventory will be larger. With a larger inventory level, the holding costs will rise. The upward sloping K curve reflects the same facts. As the average inventory level rises, the cost of supporting the working capital tied up in inventory rises.

The desire is to minimize the sum of these four costs,

$$C = \frac{S}{Q} + \frac{P}{Q} + W + K, \tag{1}$$

where everything except S is a function of Q. In general, the shape of the total cost curve will be the familiar unimodel "U" shaped curve.

The setup charge "S" is assumed to occur during the period $0 - t_1$ at a level rate.

The production cost "P" is likewise assumed to occur at a level rate through the period $0 - t_1$. It is a function of Q.

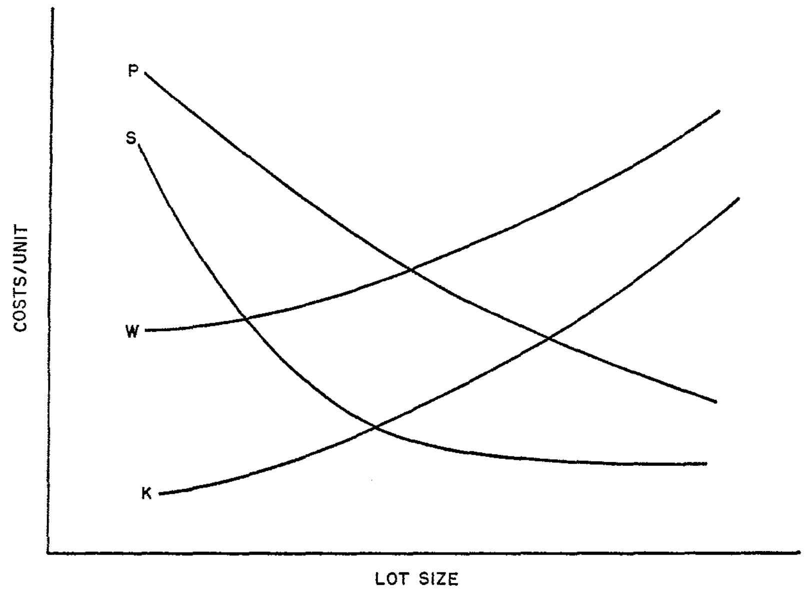

Fig. 4. Cost Per Unit Variation

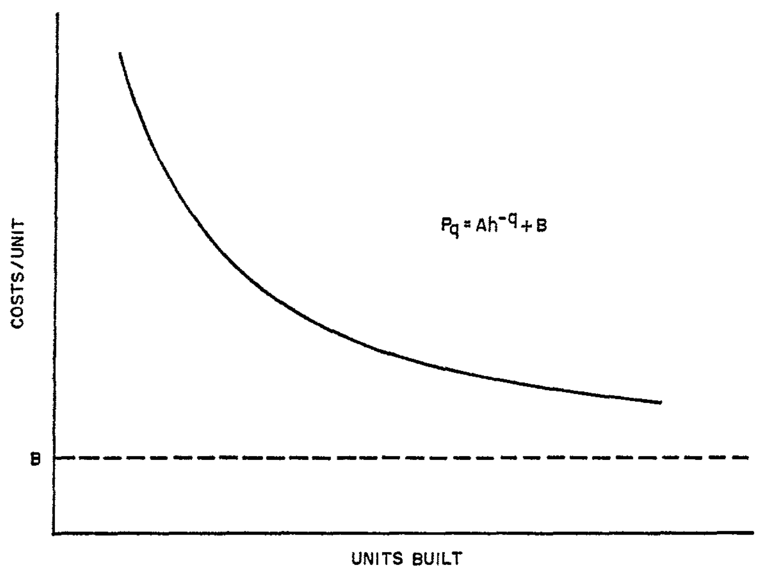

Fig. 5. Cost of Production

The warehousing cost "W" occurs throughout the entire cycle $0 - t_3$ at differing levels. It is a function of Q.

The cost of capital "K" likewise occurs throughout the entire cycle $0 - t_3$ at differing levels and is also a function of Q.

5.1 Derivation of Setup Cost "S"

The total setup cost as used in the following derivation is simply the sum of the individual setup costs on the various machines that the part or subassembly has to be channeled through. The setup cost per unit, then, is simply this total divided by Q, the number of units released at any one time.

(2) $$\frac{S}{Q} = \frac{1}{Q} \sum_j S_j,$$ where j covers all of the machine setups that have to be performed on the lot release.

5.2 Derivation of Production Cost "P"

ELRS does not assume a constant production cost but makes the more common assumption that production cost per unit decreases with the number of units in any release according to the exponential learning curve concept.[3] P/Q is, therefore, not constant, but a function of "q". This may be illustrated by the curve in Figure 5. This graph illustrates the concept that as the lot release becomes larger, the cost per unit drops because of the inherent learning process taking place during production. In the equation on the graph "q" is the running variable and "Q" is the production lot release. "B" is the asymptotic cost of the qth unit as "q" approaches infinity.

"B" is approximately the "cost" of producing a terminal unit if the production run is very long. This "cost" is the sum of these terminal unit costs over the j various machine processes. "A" and "h" are two constants which specify the learning curve. In practice, these constants may be determined by getting the P_q, q coordinates of two points, e.g., the first and tenth or the first and fifth units. With the knowledge of "B" and these two other (P_q, q) points, we can set up two simultaneous equations substituting and solving for A and h. If we do this, the expression giving the total production costs of any unit is

(3) $$\frac{P}{Q} = \frac{1}{Q} \int_0^Q (Ah^{-q} + B)\, dq.$$

The above expression for P/Q is integratable in closed form. The result is

(4) $$\frac{P}{Q} = \frac{A(1 - h^{-Q})}{Q \log (h)} + B,$$

where $\log(h)$ is $\log_e (h)$.

Occasionally, such as in the case of a numerically controlled machine, the learning curve concept is not applicable. In the computerized system for ELRS, this special case is handled by testing to see if the costs of production do not change with q.

[3] "Learning Curve" is the name commonly applied to this concept in the aerospace industry. Other industries may have other names, such as "shakedown" to apply to the same concept.

5.3 Definition of Warehousing "W" Capital "K" Costs

In the ELRS formulation, the warehousing costs, W, are assumed to be made up of two components. One of these components varies with the size of the production item and the other component varies with its value. Items such as warehousing and handling costs are at least partially made up of factors that can be assumed to be proportional to the size of the product. Other costs such as taxes, insurance, and risk of obsolescence may likewise be assumed to be proportional to the dollars invested in the product. The dollar cost of capital, K, is a percentage of value, because the cost of capital is proportional to the length of the time and level of dollars invested. ELRS, therefore, assumes all of the costs in both W and K to be proportional to either dollars invested or size. We now define two new factors.

D is the sum of the ratio factors of W and K which are cost-wise related to value.

E is the sum of the ratio factors of W which are cost-wise related to size.

Therefore, we can say that the total cost of W and K is equal to

$$W + K = (DV + EF)\cdot(\text{average time held}). \tag{5}$$

The average time that a unit is held may be derived by remembering the trapezoidal inventory model and using the fact that the third time period, t_3, may be defined as

$$t_3 = t_2 + Q/R. \tag{6}$$

If we sum the areas of the two triangles and the rectangle that make up the trapezoid and divide by Q, we see that the average dollar unit in inventory is held for a time period equal to

$$\text{average time held} = t_2 - \tfrac{1}{2}t_1 + Q/2R, \tag{7}$$

and therefore,

$$W + K = (DV + EF)\cdot(t_2 - \tfrac{1}{2}t_1 + Q/2R). \tag{8}$$

5.4 Cost Equation for Model A

By summing the three expressions just derived, and by neglecting the actual timing at which the various cash flows occur, we can say that the total cost per unit of producing and maintaining our inventory is defined as

$$C = \frac{S}{Q} + \frac{A(1 - h^{-Q})}{Q \log h} + B + (DV + EF)(t_2 - \tfrac{1}{2}t_1 + Q/2R). \tag{9}$$

The Q which minimizes this function is the optimal production release quantity. Certain costs in the above equation do not vary with Q and therefore one can eliminate them and equivalently minimize the following function to determine the optimal Q.

$$C' = \frac{S}{Q} + \frac{A(1 - h^{-Q})}{Q \log h} + \frac{Q}{2R}(DV + EF). \tag{10}$$

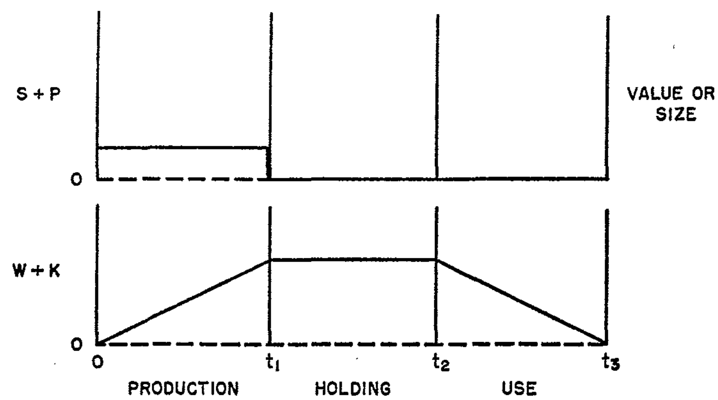

FIG. 6. Cash Flows During Inventory Cycle

Taking the derivative of C' with respect to Q and setting the resulting expression equal to zero, we obtain the following relationship:

$$0 = \frac{-S}{Q^2} + \frac{AQh^{-Q}\log h - A(1 - h^{-Q})}{Q^2 \log h} + \frac{DV + EF}{2R}. \tag{11}$$

This equation is not directly solvable for "Q"; however, we may numerically determine the solution by a Newton-Raphson, False Position[4] or other iterative numerical process.

Since the total cost curve is unimodal, it is also possible to use an efficient search technique such as dichotomous or Fibonacci search (see [7]), to rapidly converge on the minimum of Equation (10). This type of technique should very rapidly converge to values of Q corresponding to minimum C. This is because the cost structure of economic lot release or EOQ problems is such that the costs of a moderately incorrect Q are almost negligible. One has to err substantially on the lot size before the cost of the error is noticeable.

6. Timing of Cash Flows

We now proceed to develop two models, analogous to the one just presented, except that they incorporate discount factors which apply to future revenues. The first approach involves an approximation to a discount model which determines the median points of the cash flows and treats the cash flows as if they occurred at their median points in time. The second approach will involve integration of the cash flows as they are continually discounted by rate of return factors.

The model of inventory values and physical size that all three models use is typified by the trapezoid shown in Figure 2.

The ELRS model assumes both the setup and production costs occur linearly over time during the period $0 - t_1$. The holding costs, on the other hand, occur throughout the entire inventory cycle and are proportional to the amount of

[4] See any numerical analysis text, such as [4].

inventory or dollars invested in inventory. These two types of cash flows may be represented by Figure 6.

No production or setup charges are incurred during the holding and use periods of the cycle. Since the holding and capital costs are proportional to the value of the inventory at any time and the value of the inventory follows the trapezoid shape, the spending on these charges will be trapezoidal in form over time also.

We assume that:

1) the use time is proportional to "Q" and is given by Q/R;

2) the production and holding times are invariant with respect to Q. This assumes that the sequential process of production on sequentially oriented machines is not production time limited.

Both of these assumptions will be fairly good in a production operation where scheduled buffer times are relatively large. The production and use times are defined from the start of the process on the first unit to the end of the process on the last unit.

For S and P, letting $f(t)$ be the cash flow that occurs over time, we state that from time 0 to time t_1,

$$f(t) = \text{const.} = S/t_1 \quad \text{or} \quad P/t_1, \tag{12}$$

and that from time t_1 to time t_3,

$$f(t) = 0. \tag{13}$$

Analogously, for W and K, where a_1, a_2, and a_3 are various constants, we find that,

$$\begin{aligned} &\text{from } 0 \text{ to } t_1, && f(t) = a_1 t; \\ &\text{from } t_1 \text{ to } t_2, && f(t) = a_1 t_1 = \text{const.}; \\ &\text{from } t_2 \text{ to } t_3, && f(t) = a_2 - a_3 t. \end{aligned} \tag{14}$$

7. Model B

The idea of Model B is to use the approximation that the entire cash flow of a particular cost occurred at its median point in time. It is a simple exercise in analytic geometry to show that the median point in time for cash flows S and P occur at $t_1/2$. Likewise it can be shown that the median point for cash flows W and K is t_m, which is defined as follows.

If

$$3t_1 \geqq t_2 + t_3, \tag{15}$$

then

$$t_m = \sqrt{t_1(\tfrac{1}{2}t_2 + \tfrac{1}{2}t_3 - \tfrac{1}{2}t_1)}. \tag{16}$$

If

$$t_3 \geq 3t_2 - t_1. \tag{17}$$

then

$$t_m = t_3 - \sqrt{\tfrac{1}{2}(t_3^2 - t_2^2 + t_1t_2 - t_1t_3} . \tag{18}$$

And if neither condition holds, then

$$t_m = \tfrac{1}{4}(t_1 + t_2 + t_3). \tag{19}$$

Accordingly, Model B uses the approximation that the cash flows S and P occur at $t_1/2$ while the cash flows W and K occur at t_m. Therefore, the equivalent total cost of production, C, for Model B is

$$C = \left(\frac{S}{Q} + \frac{A(1 - h^{-Q})}{Q \log h} + B\right)\left(\frac{1}{1 + r}\right)^{t_1/2} + (DV + EF)(t_2 - \tfrac{1}{2}t_1 + Q/2R)\left(\frac{1}{1 + r}\right)^{t_m}. \tag{20}$$

This equation includes all of the costs and conditions which were handled in Model A and also does an approximate discounting with respect to time of the cash flows.

Analogously to the first case, the "C" may be solved by numerically solving for the zeros of the derivative of the cost expression after it is set equal to zero or by efficiently searching for a minimum.

8. General Discounted Flow Model—Model C

We now consider a discounting model which exactly reflects the various cash flows as they occur.

Again letting $f(t)$ be the flow of cash over time

$$C = \int_0^{t_3} \frac{f(t)}{(1 + r)^t}\, dt \tag{21}$$

is the objective function that should be minimized if all cash flows occur in the period zero to t_3.

For our model, the above expression may be decomposed to

$$C = \int_0^{t_1} \frac{f(t)}{(1 + r)^t}\, dt + \int_{t_1}^{t_2} \frac{f(t)}{(1 + r)^t}\, dt + \int_{t_2}^{t_3} \frac{f(t)}{(1 + r)^t}\, dt, \tag{22}$$

which is the objective function to be minimized.

8.1 Production Cost

If we are to integrate with respect to time, "t," then all costs which are either explicitly or implicitly functions of time should be expressed as explicit functions of time. For example, if we make the production cost a function of time,

$$q = \frac{Qt}{t_1}, \tag{23}$$

where "Q" is the lot size. The equation for the production cost per unit becomes, therefore,

$$\Big|_0^{t_1} \frac{P}{Q} = \frac{1}{t_1}\int_0^{t_1} \frac{(A(h)^{-Qt/t_1} + B)\,dt}{(1+r)^t}, \tag{24}$$

Outside of the interval 0 to t_1, P equals zero.

8.2 Setup Cost

The setup cost as a function of time is

$$\Big|_0^{t_1} \frac{S}{Q} = \frac{S}{t_1 Q}\int_0^{t_1} \frac{dt}{(1+r)^t}, \tag{25}$$

and is equal to zero outside of the interval 0 to t_1.

8.3 Carrying Costs

If "V" and "F" are functions of time, then

$$\Big|_0^{t_1} V(t) = \frac{Vt}{t_1};\ \Big|_{t_1}^{t_2} V(t) = V; \quad \text{and} \quad \Big|_{t_2}^{t_3} V(t) = \frac{V(t-t_3)}{t_2-t_3}. \tag{26}$$

The same expressions apply to "F" with "F" replacing "V." Therefore, the W plus K terms appear as follows,

$$W + K = \int_0^{t_1} \frac{(DV+EF)t\,dt}{t_1(1+r)^t} + \int_{t_1}^{t_2} \frac{(DV+EF)\,dt}{(1+r)^t} + \int_{t_2}^{t_3} \frac{(DV+EF)(t-t_3)\,dt}{(1+r)^t(t_2-t_3)}. \tag{27}$$

8.4 Total Costs

Summing up the previous mathematical expressions, we arrive at an equation for the total costs of producing and maintaining a production lot,

$$C = \frac{S}{t_1 Q}\int_0^{t_1} \frac{dt}{(1+r)^t} + \frac{1}{t_1}\int_0^{t_1} \frac{(AH^{-Qt/t_1} + B)\,dt}{(1+r)^t} + \frac{DV+EF}{t_1}\int_0^{t_1} \frac{t\,dt}{(1+r)^t} + (DV+EF)\int_{t_1}^{t_2} \frac{dt}{(1+r)^t} + \frac{DV+EF}{t_2-t_3}\int_{t_2}^{t_3} \frac{(t-t_3)\,dt}{(1+r)^t}. \tag{28}$$

The first term represents the total discounted setup cost; the second term is the total discounted production cost; and the other terms are the discounted carrying, holding, and capital costs. As "Q" changes, the third and fourth terms in the above equation remain constant and therefore it is possible to choose an optimal policy by minimizing an equivalent cost expression which excludes these terms. The fifth term varies with "Q" because t_3 is defined in terms of "Q". However, since C is evaluated for fixed Q, terms containing t_3 may be taken outside the integral. We fix the Q and then integrate over time to obtain the cost associated with this Q.

The first term in equation (28) becomes

$$\frac{S}{t_1 Q}\int_0^{t_1}\frac{dt}{(1+r)^t}=\frac{S((1+r)^{t_1}-1)}{t_1 Q(1+r)^{t_1}\log(1+r)}. \tag{29}$$

Integrating by parts, the second integral becomes

$$\frac{A((1+r)^{t_1}-h^{-Q})}{(1+r)^{t_1}(Q\log+t_1\log(1+r))}+\frac{B((1+r)^{t_1}-1)}{t_1((1+r)^{t_1}\log(1+r))}. \tag{30}$$

The fifth term of the C equation can also be integrated by parts, and becomes

$$\frac{R(DV+EF)}{Q\log^2(1+r)}\left[\frac{1}{(1+r)^{t_3}}+\frac{\frac{Q}{R}\log(1+r)}{(1+r)^{t_2}}-\frac{1}{(1+r)^{t_2}}\right], \tag{31}$$

where

$$t_3=t_2+\frac{Q}{R}. \tag{32}$$

If an expression for total cost/part is desired in addition to the optimal Q, we would be interested in the third and fourth terms of equation (28) even though they do not vary with Q. These terms are also integrated by parts. The third term becomes

$$\frac{DV+EF}{t_1\log^2(1+r)}\left[1-\frac{1}{(1+r)}-\frac{t_1\log(1+r)}{(1+r)^{t_1}}\right], \tag{33}$$

while the fourth term is

$$\frac{DV+EF}{(1+r)^{t_2}\log(1+r)}\left[\frac{(1+r)^{t_2}}{(1+r)^{t_1}}-1\right]. \tag{34}$$

Remembering that t_3 is equal to t_2+Q/R, the "Q" which minimizes the following C expression, which is the sum of the closed integrals, is optimal:

$$\begin{aligned} C = {} & \frac{S((1+r)^{t_1}-1)}{t_1 Q(1+r)^{t_1}\log(1+r)}+\frac{A((1+r)^{t_1}-h^{-Q})}{(1+r)^{t_1}(Q\log h+t_1\log(1+r))} \\ & +\frac{B((1+r)^{t_1}-1)}{t_1(1+r)^{t_1}\log(1+r)}+\frac{DV+EF}{t_1\log^2(1+r)}\left[1-\frac{1}{(1+r)}\right. \\ & \left.-\frac{t_1\log(1+r)}{(1+r)^{t_1}}\right]+\frac{DV+EF}{(1+r)^{t_2}\log(1+r)}\left[\frac{(1+r)^{t_2}}{(1+r)^{t_1}}-1\right] \\ & +\frac{R(DV+EF)}{Q\log^2(1+r)}\left[\frac{1}{(1+r)^{t_3}}+\frac{\frac{Q}{R}\log(1+r)}{(1+r)^{t_2}}-\frac{1}{(1+r)^{t_2}}\right]. \end{aligned} \tag{35}$$

Since this equation is not simple to differentiate with respect to Q, it is suggested

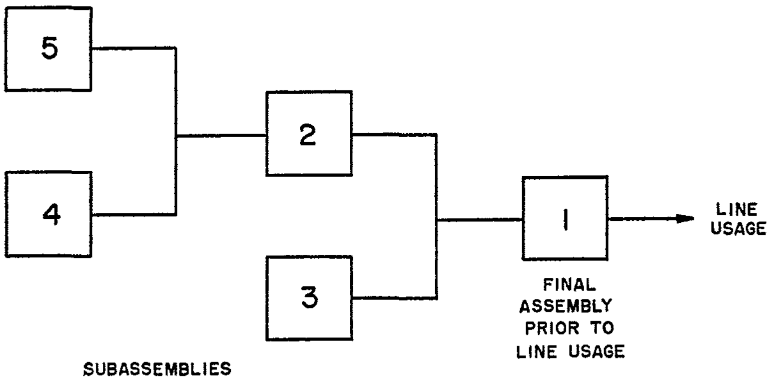

FIG. 7. Interrelationships Among Subassemblies

that the optimal "Q" be determined by one of the earlier mentioned search techniques.

9. Generalization of ELRS to Processes Including Subassemblies

9.1 *Policy Determination*

Often the production problem in a fabrication shop is not viewed as a continuous process, the output of which is one final product; rather, it is viewed as the production of several subassemblies which may then be combined into other subassemblies, ad infinitum, until they are finally combined into the final product. In a production process such as this, lot release sizes on the various subassemblies may, in fact, be different from each other. The problem then becomes one of determining optimal lot release sizes for each one of the subassemblies. Each subassembly, however, cannot be viewed independently of the other subassemblies because of the interdependencies that exist between the various subunits in the production process. This succession of subassemblies into a major assembly may be illustrated by Figure 7.

At first thought, determining the proper production policy out of all of those available may seem like a very difficult task; however, simple reasoning eliminates most of the available policies. From basic relationships, we can draw the following conclusions.[5]

1. *Never produce earlier subassemblies in smaller lot releases than later ones.*[6] If

[5] Considerations of workload balancing can change any of these conclusions. These conclusions, as all those made in this paper, are made with the assumption that workload balancing considerations will be applied after the ELRS is determined. These considerations may change the lot release sizes. Problems of safety stocks and spoilage of parts can be handled within the framework of this analysis. If the expected spoilage of a part is 10%, then the initial release must be 10% higher than the optimum recommended by this analysis. A safety stock will be an additional amount that is made once and then carried along with the production policy applying to oscillations on top of this safety stock.

[6] An earlier subassembly is one that fits as a part into a later subassembly. The subscript "i" will be used to number the subassemblies, "$i + n$" denoting an earlier subassembly and "$i - n$" connoting a later subassembly.

we did produce earlier subassemblies requiring multiple releases of an earlier part for one release of a later one, a better policy would always be to release the earlier subassembly in lot sizes equal to that of the later subassembly. This is because being produced less often, the total cost per unit of $S/Q + P/Q$ will be less. The total amount of time that the earlier subassembly is carried in stock will also be less, because the earlier releases on the subassembly cannot be used until later releases on the subassembly bring the total amount of the part available up to what is necessary for making one release of the later assembly. A better policy would be to produce the entire requirement of the earlier subassembly at the time of the later release, just prior to the requirement for its usage.

2. *Always produce earlier subassemblies in integer multiples of the lot release size of the final assembly.* There is no need to produce earlier subassemblies in non-integer multiples of later subassemblies or the final assembly because the non-integer parts of the subassembly release will be unable to completely satisfy the demand for that subassembly in any release of a subsequent assembly. A non-integer multiple release will always require more production setups and runs than will the next higher integer multiple (or any larger) release. The only advantage of producing with more releases is that inventory carrying costs are lower. However, the noninteger part of the release ends up being carried until the next release is made, a relatively long time that did not enter the ELRS solution, and this extra carrying time negates the advantage of producing exactly the optimal quantity. The non-integer part is therefore just carried longer in inventory and provides no advantage whatsoever. An economic analysis of this extra carrying time is presented later in this section.

3. *We therefore have narrowed the choice of reasonable policies to those policies which determine lot release sizes of earlier assemblies to be equal to integer multiples of the lot release sizes of later subassemblies and of the final assembly.* On the following pages, a description of how we can explicitly determine a production policy for interrelated subassemblies is presented.

9.2 Analysis

From Conclusion 3 above,

$$n_i = kn_{i-1}. \tag{36}$$

When part "i" goes into part "$i - 1$", where "k" is an integer, and where the

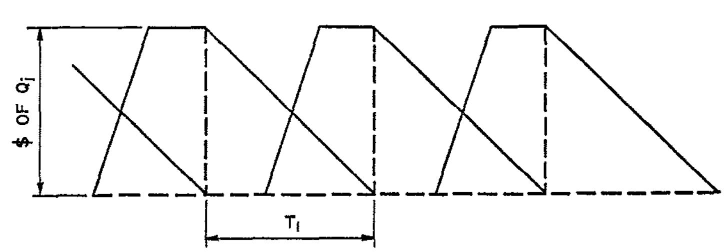

FIG. 8. Inventory Fluctuation of Fabricated Part Experiencing Continuous Usage

"n's" are defined from the optimal lot release quantities,

$$Q_i = n_i Q_1 , \tag{37}$$

Q_1 being the lot release size of the completed assembly ready for installation on the line.

Since the final assembly is used at a uniform rate on the line, the production cycle of the final assembly has to make sure that parts are always available and would resemble Figure 8. Notice that each new release is completed and ready for usage on the line just at the time that the old release runs out (the addition of safety stocks would not change the basic cycle, but merely move all ordinate values upwards).

The time between successive lot releases, T_1, is equivalent to the time between the dotted lines,

$$T_1 = Q_1/R. \tag{38}$$

(R should be defined in terms of final production units, so that all subassemblies in an interrelated case will have the same R. If Z identical parts or subassemblies are required for one final unit, then the ELRS output should be multiplied by Z.)

It will become evident that the time period T_1 is important because for optimal inventory control all levels of lower assemblies or parts that go into assembly 1 must be released in integer time multiples of T_1.

Now consider subassembly 2, which is required as a part for assembly 1. From earlier statements,

$$Q_2 = n_2 Q_1 , \tag{39}$$

where n_2 is an integer.

The fluctuation of inventory value over time for a single lot release of item 2 would be like Figure 9 if $n_2 = 4$. The abscissa represents time and the ordinate is either dollars invested in inventory or units of inventory. Each time that subassembly 1 is released, $\frac{1}{4}$ of the total amount made of subassembly 2 is used up. This staggered usage would cause the inventory value cycle of subassembly 2 to have the staircase effect of the figure above, each drop corresponding to a production period of subassembly 1. This basic cycle looks different from the previously postulated trapezoidal shape for inventory. However, it is possible to modify the previously derived ELRS formulas to take into account this new shape.

For simplicity in the time variable, we shall work with Model A and adjust the earlier ELRS formula by modifying the "average time/unit" expression, which helped to determine the carrying charges. Model C would also be used, giving a slightly more accurate policy recommendation; however, part of the required input would be a list of the specific times that production would commence on each subassembly. The author feels that in most industrial situations the improved recommendations made possible by the use of Model C would not be worth the extra trouble it would take to get this additional input data. Therefore, in the many subassemblies case, we use Model A and ignore time discounting.

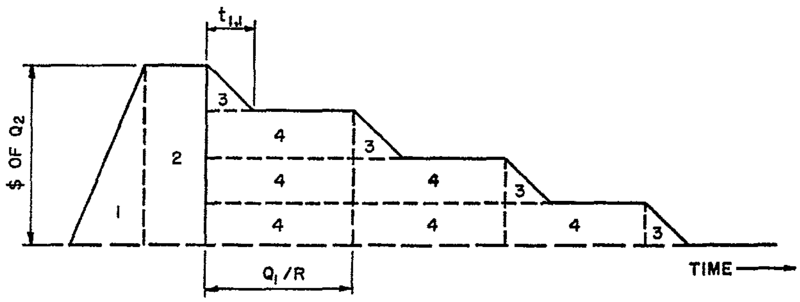

FIG. 9. Inventory Fluctuation of Subassembly 2

The cost equation for Model A of ELRS was

$$(9) \qquad C = \frac{S}{Q} + \frac{A(1 - h^{-Q})}{Q \log h} + B + (DV + EF)\left(t_2 - \frac{1}{2}t_1 + Q/2R\right).$$

The last expression in this cost equation,

$$(7) \qquad t_2 - \tfrac{1}{2}t_1 + Q/2R,$$

was derived from consideration of the trapezoid and was a representation of the time that an average unit was held. In general, the av time/unit = 1/units produced (units × time/unit).

For the new cycle, we simply have to consider the new geometric shape. We can break the entire cycle up into small geometric patterns whose areas we know. By summing these small figures, we get the area of the new cycle. According to the way our axes are defined, this area is the integral of (units × time/unit) and when divided by the number of units produced yields an expression for average time/unit. For this particular new cycle, the average time per unit is[7]

$$(40) \qquad \text{av time/unit} = \frac{1}{Q_2}\left[\frac{Q_2\, t_{1,2}}{2} + (t_2 - t_1)Q_2 + n_2\left(\frac{t_{1,1}\, Q_2}{2n_2}\right) + \frac{Q_2}{n_2}(T_1) + \frac{Q_2}{n_2}(2T_1) + \cdots \frac{Q_2}{n_2}(n_2 - 1)(T_1)\right].$$

The second line of terms in equation (40) may be summed to

$$(41) \qquad \frac{Q_2}{n_2}\left[\frac{n_2(n_2 - 1)T_1}{2}\right].$$

Therefore, the

$$(42) \qquad \text{av time/unit} = \frac{1}{Q_2}\left[\frac{Q_2\, t_{1,2}}{2} + (t_2 - t_1)Q_2 + \frac{Q_2\, t_{1,1}}{2} + \left[\frac{Q_2(n_2 - 1)T_1}{2}\right]\right],$$

or, for subassembly 2

$$(43) \qquad \text{av time/unit} = \left[t_2 - \frac{t_1}{2} + \frac{t_{1,1}}{2} + T_1\,\frac{(n_2 - 1)}{2}\right].$$

[7] Definition Note: $t_{i,j}$ means t_i for subassembly j. Where the second subscript is omitted, it means that it should be the subscript for the assembly whose Q is presently being derived.

As mentioned earlier,

$$T_1 = Q_1/R. \tag{44}$$

The four terms comprising equation (43) represent the areas of the various geometric shapes that make up Figure 9, which shows the cycle of inventory value for subassembly 2.

1. The first term represents the initial triangle for production of subassembly 2.
2. The second term is the rectangle for the holding period.
3. The third term is the sum of the little triangles which occur during the use of this part in assembly 1.
4. The remaining terms which are algebraically combined into one term are the long narrow rectangles which make up the rest of the area.

The ELRS formula, with the new value for the average time that one unit is held, gives the cost of producing subassembly 2 according to the optimal policy derivation made earlier.

A little thought shows that the time between successive lot releases of subassembly 2 is given by

$$T_2 = n_2 T_1 . \tag{45}$$

This is because the time between successive releases is governed by the use period and does not depend on the production or hold times. We can consider subassemblies that are earlier (larger number) and we now illustrate that, in general,

$$T_i = n_i T_1 , \tag{46}$$

where "n_i" is defined as

$$n_i = \frac{Q_i}{Q_1} . \tag{47}$$

The basic shape of the inventory cycle for subassembly 3, if subassembly 3 were required as a part of subassembly 2, would be similar to the trapezoid with steps, representing the inventory cycle of subassembly 2. All subassemblies will have this same basic shape. For instance, if Q_3 were six times as large as Q_2, then the inventory cycle for subassembly 3 would have six steps like Figure 10, regardless of what n_2 was. Q_2, however, is important in determining the actual height of the figure since n_3 is defined in terms of Q_1. If Q_3 was equal to $6Q_2$, while Q_2 was $4Q_1$, then Q_3 would equal $24Q_1$ (i.e., $n_3 = 24$).

Notice that the length of the steps in 3's case would be equal to the time between lot releases of 2, T_2, which was proportional to the product of T_1 and n_2.

As another example, if $Q_3 = 2Q_2$, while $Q_2 = 3Q_1$, the basic cycle on subassembly 3 will be twice as long as that of 2 (remember basic cycle time is determined only by the use period of the cycle) and six times as long as T_1. The ratio Q_i/Q_1 is equal to the ratio T_i/T_1 in general because the usage rate, R, is the same for all subassemblies; if twice as many are produced at any one time, we only need to produce one-half as often.

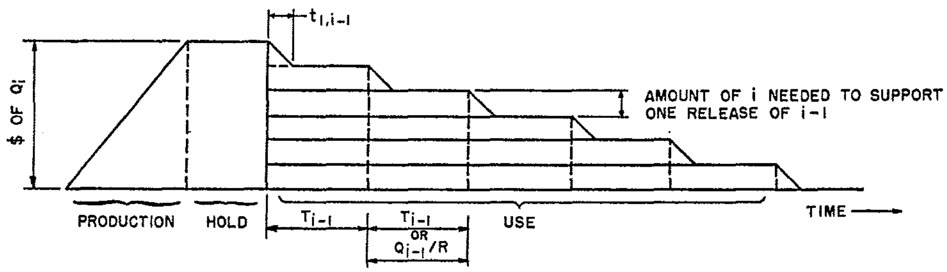

FIG. 10. General Representation of Inventory Fluctuation

In general, then, the coefficient "n" is the relevant factor for determining the basic cycle period of each subassembly

$$T_i = n_i T_1 . \tag{46}$$

The stepping stone use of 2 does not change general conclusion (48). The important factor is "n" and the time between the production releases on 2. If n_3 equals n_2, the time between successive lot releases, T_3, is equal to T_2. If n_3 equals $2n_2$, then the inventory value curve will have one step and T_3 will equal $2T_2$ or $6T_1$ when $n_2 = 3$.

Now return to the analysis of the average time per unit in the inventory cycle. For subassembly 2 this was

$$\text{av time/unit} = t_{2,2} - \frac{t_{1,2}}{2} + \frac{t_{1,1}}{2} + \frac{T_1(n_2 - 1)}{2} , \tag{48}$$

with the last term accounting for the area under the steps.

The most important step in the logic was that each use period in 3 corresponded to a complete production period in 2.

In general, the value figure for subassembly i has n_i/n_{i-1} steps. Therefore, we can generalize the formula for av time/unit by allowing the replacements shown in Table 1.

So, in general, the average time spent in inventory/unit for subassembly "i" is

$$\left[t_{2,i} - \frac{t_{1,i}}{2} + \frac{t_{1,i-1}}{2} + \frac{T_{i-1}\left(\frac{n_i}{n_{i-1}} - 1\right)}{2} \right] \tag{49}$$

and ELRS, with this expression for average time, validly determines the lot release size of assembly i in the many subassemblies problem.

9.3 Solution Algorithm

The above theory provides an adequate base for the determination of an algorithm which will lead to a near optimal production lot size for several interrelated subassemblies.

1. Starting at the lowest level (or highest numbered or earliest subassemblies), determine the Q_i's and associated total C_i's. Using the regular ELRS Model A

TABLE 1

Generalization of Equation 49

T_1	*to be replaced by*	$\frac{T_{i-1}}{n_i}$
n_2	*to be replaced by*	n_{i-1}
$t_{1,1}$	*to be replaced by*	$t_{1,i-1}$
$t_{1,2}$	*to be replaced by*	$t_{1,i}$
$t_{2,2}$	*to be replaced by*	$t_{2,i}$
	where $n_1 \equiv 1$	

formula, derive the Q_i's and C_i's of all subassemblies working from the earlier ones toward the later ones and using the costs of all prerequisite subassemblies as material costs for later subassemblies. Finish the entire list of Q's and C's ending up with a value for Q_1 and C_1.

2. Now, given Q_1 and C_1, work backwards, developing new Q_i's and C_i's from the optimum policy restrictions such as $Q_{i+1} = kQ_i$, where k is an integer. The new ELRS formula is used in this step but the subassembly material costs are the same ones derived in the previous step. Numerically, we only need to test integer values for "n" starting with 1. As soon as a cost figure, C_i, is reached that is higher than the C_i determined with $n - 1$, then we can say, without testing any further, that Q_i corresponding to the previous value is the cheapest. We can do this because the cost curve derived from the ELRS formula has a positive second derivative everywhere and therefore a unique minimum. Now we have an entire set of Q_i's and C_i's derived from the policies mentioned earlier.

3. Using the lowest level C_i's just obtained from Step 2, now derive the next higher level C_i's for the constrained Q_i's from Step 2. This computation uses the lowest level C_i's as material costs in conjunction with the new Model A formula and the most recent values for T_i and n_i. Using earlier subassemblies' costs as material cost inputs to later subassemblies, compute new C_i's for the constrained Q_i's through the entire pyramid of prerequisite relationships back towards the final subassembly. The purpose of Step 3 is to modify the costs associated with Step 2 to reflect the constrained production policy. When the new costs of all of the subassemblies prior to the final one are determined, compute a new Q_1 and C_1 for the final subassembly.

4. Now go back and repeat Step 2 and continue iterating between Step 2 and Step 3 until two successive Q_1's or C_1's from Step 3 are within a small arbitrary percentage of each other, such as 5%. When this point is reached, complete Step 2 one last time. This final set of Q_i's and associated C_i's provides the optimal policy for production in the many subassemblies problem. If looked at independently, the production of any particular subassembly might be cheaper at a different Q than that determined by this method; however, as it has been shown above, the logical interrelationships make this overall policy the best.

10. Numerical Test of ELRS

Two computer programs were written so that numerical tests could be made of the various models proposed in this paper. The purpose of the first program was to permit an exhaustive testing of the differences between the recommended policies of Model A, Model B and Model C. These three different analyses were simultaneously tested on forty separate data cases, half of which were fairly reasonable representations of actual production data and the other half of which were extreme, unrealistic data cases which were used to test the limiting behavior of the ELRS model. The average IBM 7090 total run time per case was less than one second.

The second program implemented the generalization of ELRS to processes including subassemblies. The primary purpose of this program was to test the reasonableness of recommended policies in the many subassemblies case and to test simultaneously the effectiveness and rapidity of convergence of the solution algorithm proposed for the many subassemblies case.

The results obtained from all of the test cases in both programs were very reasonable. The cost curves obtained were of the typical broad flat EOQ type. This type of cost curve reflects conditions where small to medium size deviations from an optimal policy result in minor cost penalties.

The only exception from the reasonable behavior of the models was the performance shown by Model B under extremely large interest rates. With a discount rate of several hundred percent per year, the cost curve from Model B lost its U shape and showed continually decreasing costs as Q increased. Although this degeneracy occurred under a test of very unrealistic conditions, it is sufficient cause to suggest that Model B not be a recommended analysis. Both Model B and Model C require the same type of input data and since Model C is a conceptually superior analysis, there is really no need for the Model B approach. Model A and Model C behaved very reasonably for all of the test cases.

Table 2 lists some typical values that were determined for actual parts fabricated at the Norair Division of Northrop Corporation.

As the discount rate increased while all other conditions remained comparable, the production policy recommended by all of the models decreased. On the other hand, in every case where the recommended Q by each of the models was different, the sizes recommended by Models B and C were larger than that recommended by A. Under reasonable discount rate assumptions, the recommendations by all of the models were the same; however, when the discount rate got up to approximately 100% per year and over, differences in recommended Q's were ob-

TABLE 2

Norair Examples

Optimum Release Quantity	Setup and Production $ Costs
138	1
78	2
59	8
29	10
68	20
63	25

TABLE 3

Model C Recommendations

Low Q	Optimal Q	High Q
17	27	48
24	36	85
26	54	105
30	48	90

served. The worst discrepancy in a policy recommended by Model A from that recommended by Model C was found on a data case where the discount rate was 1000% (maximum tested) per annum. In this case, the recommended Q of Model A was 38% less than that of Model C; however, the cost penalty on Model C of using this 38% lower Q, was only 1.5%.[8] The next worst cases involved deviations in recommended policies of 25% and 30% with cost penalties in both cases of less than 1% as figured by the Model C analysis.

Because of these small differences, it was concluded that under most reasonable production circumstances the added value of the sophistication introduced by the discounting of future costs was relatively minor. On the other hand, the only additional data required for the Model C solution are t_1 and t_2 (r is not a new data requirement because it is included as a part of D). If this information is readily available and the evaluation is done on a computer, then there is no reason not to use the more correct approach, Model C.

The broad flat area on the cost curve around the optimum production lot size may be effectively illustrated by a couple of examples. Where both "HIGH Q" and "LOW Q" are defined as the Q's on each side of the optimal Q corresponding to a cost 5% greater than the optimal Q, Table 3 represents typical output data. It is obvious that the production policy can vary substantially around the optimal while incurring only small cost penalties. If the optimal production policy cannot be followed, it seems to be better to err by producing too many at each lot release rather than too few.

The advantage of an upward bias over a downward one is reinforced if we are using the recommendations of Model A instead of Model C. This is because Model A, which is really a simplification of the more correct analysis, Model C, tended to be slightly lower in its recommended lot sizes.

It may be helpful to restate the fact that Model A was used as the basis for the generalization of ELRS to the interrelated parts case because all of the data for all subassemblies would have to be introduced in a time dependent manner for Model C to function properly. As mentioned in Section 9.2, it was felt that

[8] This is not a misprint. All other things remaining equal, an increase in r always caused the recommended Q from Model C to decrease. However, as unplausible as it may sound, Model C, which considered discounted values always recommended a larger Q than Model A, which ignored r. Total costs at any Q were smaller under Model C than Model A but the minimum cost always occurred at a larger Q.

this type of input data would not be readily available in most industrial situations. Data on sequential time relationships are not required as input for Model A; and, since the policy recommendations were identical with those of Model C under reasonable circumstances, Model A was used as the basis for the generalization.

The second program and the associated tests showed that the ELRS model for several interrelated subassemblies and the proposed solution algorithm were reasonable, feasible and very efficient in terms of computer run time. The solution method on all of the data cases ran and converged very rapidly, never requiring more than two cycles. The policy recommendations always seemed very reasonable. Usually the suggested lot size of the final subassembly prior to line usage was very close to what it would have been not considering any of the relationships. The solution algorithm seemed to typically round off earlier lot release sizes to the integer multiple of the successor subassembly that was nearest to what the independent Q would have been for the predecessor subassembly (within constraints). For example, one test was run with five levels of hierarchical dependency with the same data applying to each subassembly. The only difference at each step, therefore, was in the cost of materials; the later subassemblies included the cost of the earlier subassemblies in their costs of materials. In this case, two out of the four predecessor levels had "n's" that were double that of successors, so that the first level subassemblies were released in lot sizes four times as great as the final subassembly. Other tests showed similarly reasonable behavior and the author was unable to uncover any numeric examples that the generalized ELRS process was unable to analyze.

It may be of interest to note that the twenty real test cases were randomly selected from the Norair Division of Northrop Corporation and calculations showed that production according to ELRS sizes averaged a total of setup, production, and holding costs that was $2\frac{1}{2}$ percent lower than the then current figures. For a moderately large job-shop that had expenditures of $20 million a year, this reduction in total costs would amount to $500,000.

11. Conclusions and Limitations

This paper has presented an analysis of the cost of repetitively producing and maintaining parts and subassemblies in a fabrication or job shop. The objective of this analysis was to determine the cheapest lot sizes for producing and storing these parts and subassemblies. The result of this analysis was a mathematical economic model that took into account various broad classes of costs, such as setup, production, carrying, capital, direct labor, warehouse, storage, paperwork, taxes, insurance, interest, material, etc.; and determined the lot release size that would make the total of these costs the lowest possible. Some other important considerations such as production schedules and correct inventory levels may be determined as a direct outgrowth of the analysis presented.

Some companies use the standard EOQ formula,

$$Q = \left(\frac{2RS}{A}\right)^{1/2}, \tag{51}$$

$$R \equiv \text{Usage Rate}$$

$$S \equiv \text{Setup Cost}$$

$$A \equiv \text{Carrying cost/unit/time,}$$

to determine production lot release sizes. We have seen, however, that compared with standard EOQ, ELRS has the following advantages.

1. Production and Setup costs are spread over time—EOQ assumes that all of these charges occur at one moment.
2. Production and Setup costs are spread over the machines involved in the production process—EOQ does not include this.
3. ELRS treats production costs on a learning curve basis—EOQ does not analyze this.
4. ELRS takes into account scheduled holding periods after the production process is completed and before scheduled use—EOQ does not consider this.
5. ELRS has the capability to discount future costs by the interest rate—EOQ completely ignores this point.
6. ELRS's maintenance costs are based on value and physical storage size—EOQ only considers value.
7. ELRS establishes an optimum overall production policy for any group of interrelated parts and subassemblies. If part A is required to make subassembly B, this fact is taken into account in determining an overall optimum production policy for the entire group of interrelated parts and subassemblies—EOQ does not consider this at all.

The recommendations made by the model presented in this report should not always be followed exactly because some factors, such as physical production constraints, have not been included in the analyses. If the optimal production release size is 50 and production restraints limit the release size to 10, then the lot size has to be 10. Special conditions can also influence the applicability of the model. For example, the analysis is only really applicable to the production of a part whose design is fairly stable. If a part were so often technologically changed that the design was only frozen for a short time period, then it would be a waste of time to analyze its production by the ELRS model. The model also assumes that the usage rate and the amount held in safety stock are fairly stable.

ELRS analysis should not be applied to a part the first few times it is released. Only when a certain amount of production history and experience has built up would a model such as ELRS become applicable. After the production policy recommendations of the ELRS model are made available, they should be modified by allowances for:

1. Workload balancing.
2. Manpower fluctuations.
3. Technical restraints, such as machine capacity.
4. Emergencies such as new requirements with very short lead times.

The recommendations of ELRS cannot be followed blindly, but must be analyzed for reasonableness in light of these other criteria.

In short, the model presented in this report is offered as a significantly better approach to determining production release sizes than any other available, and it is offered as a way of substantially reducing the cost of production and inventory in a job shop.

References

1. Ackoff, R. L., *Progress in Operations Research*, John Wiley & Sons, New York, 1961, Chapter 3. (This chapter concludes with a long reference list on Inventory Theory.)
2. Dalrymple, J., "Controlling Retail Inventories," *Journal of Retailing*, Spring, 1964, p. 9.
3. Hadley, G., and Whitin, T. M., *Analysis of Inventory Systems*, Prentice-Hall, Englewood Cliffs, 1963.
4. Hildebrand, F. B., *Introduction to Numerical Analysis*, McGraw-Hill, New York, 1956.
5. Holt, C. C., et al., *Planning Production, Inventories, and Work Force*, Prentice-Hall, Englewood Cliffs, 1960.
6. IBM Corporation, *IMPACT—Inventory Management Program and Control Techniques*, Publication E20-8105, 1962.
7. Wilde, D. J., *Optimum Seeking Methods*, Prentice Hall, Englewood Cliffs, 1964.