COWLES FOUNDATION DISCUSSION PAPER NO. 23

Programming of Economic Lot Sizes*

Alan S. Manne 1/

January 31, 1957

Revised April 18, 1957

1/ The author is deeply indebted to E. Greenwood, general manager, and also to C. Correll, A. Pastick, and A. Goldman, all of Norden-Ketay, Inc., Milford and Stamford, Connecticut.

---

* Research undertaken by the Cowles Foundation for Research in Economics under Contract Nonr-609(01) with the Office of Naval Research.)

Programming of Economic Lot Sizes

Summary

This paper studies the planning problem faced by a machine shop required to produce many different items so as to meet a rigid delivery schedule, remain within capacity limitations, and at the same time minimize the use of premium-cost overtime labor. It differs from alternative approaches to this well-known problem by allowing for setup cost indivisibilities.

As an approximation, the following linear programming model is suggested: Let an activity be defined as a sequence of the inputs required to satisfy the delivery requirements for a single item over time. The capacity input coefficients for each such activity may then be constructed so as to allow for all setup costs incurred when the activity is operated at the level of unity or at zero. It is then shown that in any solution to this problem, all activity levels will turn out to be either unity or zero, except for those related to a group of items which, in number, must be equal to or less than the original number of capacity constraints. This result means that the linear programming solution should provide a good approximation whenever the number of items being manufactured is large in comparison with the number of capacity constraints.

1. Background

It is common knowledge that the presence of "setup costs" in a manufacturing process raises questions of indivisibilities [4], and that such indivisibilities constitute a formidable obstracle to any attempt to phrase economic lot size problems in terms of linear programming. In economist's language, this amounts to saying that the presence of economies of scale contradicts the assumptions of marginal analysis, along with such economic theories as linear programming, which are so deeply rooted in marginalism.

This paper reports upon the successful use of linear programming in a special instance involving setup costs and economic lot sizes.*

---

* If the total costs of producing x units of a particular item in a single lot are given by: $a\delta + bx$

$$\text{where } x \begin{pmatrix} > 0 \\ \\ = 0 \end{pmatrix} \qquad \text{implies } \delta \begin{pmatrix} = 1 \\ \\ = 0 \end{pmatrix}$$

then the constant $\underline{a}$ is said to represent the "setup cost" for that item and the constant $\underline{b}$ the "incremental unit cost." Evidently setup costs are minimized by concentrating an entire production requirement into a single lot, rather than by splitting up that requirement among several lots.

Unlike a number of the more recent proposals [e.g., 1], the particular model is a non-stochastic one, and in this sense is of less general applicability. The distinctive features of the approach outlined here is that capacity limitations - and hence interdependence between individual items - are treated as an explicit part of the economic lot size decision.

Since the background of the individual problem makes it possible to justify a number of simplifications in the mathematical model, it seems worthwhile to summarize the leading features of that background:

The plant in question sells most of its output to the armed services of the United States. The "end items" of this plant are in turn an input for other manufacturing activities, and the timing of deliveries takes on even greater importance here than in the routine production of consumers' goods. Indeed, in his competitive bid for any particular product, the manufacturer stipulates not only the price at which he will undertake to produce the item, but also the dates at which individual units will be delivered to the Government. Because the actual timing of deliveries affects the manufacturer's reputation -- hence his ability to obtain future contract awards -- one assumption that underlies all production planning is that the manufacturer will adhere to the promised delivery dates.

Since the plant does not produce large quantities of any individual end item, the productive process is of the batch type rather than continuous. As is typical of many metal-working establishments, the first step is to produce individual parts in the plant's own machine shop and to procure certain parts from other manufacturers. Once all the parts for a particular finished unit are on hand, these are assembled, tested, and the item is turned over to the Government. Note that if the Government contract calls for delivery of 25% of the finished units over each of four successive months, only 25% of the total requirement for each individual part needs to be available at the time actual assembly is initiated.* This means that one of the significant choices

---

* In practice, "buffer stock" considerations may indicate that more than 25% of the total for each part ought to be available before final assembly is initiated.

---

to be made is that of splitting production lots for individual parts so as to meet the initial delivery requirements, but still defer a portion of the machining work until the latter part of the delivery cycle. Lot-splitting does, of course, bring about an increase in setup costs, and so an optimum lot size decision entails an economic balance between the advantages of reducing setup costs, and the advantages of smoothing out the production program over time.

Although there is some overlap between planning for the machine shop and for the final assembly and testing operations, this paper is primarily concerned with the machine shop itself, and only to a secondary extent with the problems created by this overlap. Actual planning of the machine shop's activities takes place at two distinct echelons of the plant's management, and correspondingly at two different levels of abstraction. Short-range scheduling is concerned solely with such details as which parts are to be manufactured, and which individuals and machines are to be used. Long-range scheduling (up to eighteen months ahead) is concerned with the general problem of whether the machine shop's existing resources will be able to meet the company's future delivery commitments, and if not, what policies should be adopted to supplement the existing resources: over-time work, recruiting and training of additional personnel, and outside procurement of certain parts.* This paper is concerned almost entirely with the long-range

* Decisions on the purchase of new equipment constitute an additional degree of freedom, but since the payout period for such equipment normally extends over several years, the company's policies on equipment purchase may be regarded as fixed -- at least as far as an eighteen month production schedule is concerned.

problem, as distinct from the day-to-day operation.

Within the company, the traditional procedure for long-range production planning has emphasized calculations made upon the assumption that each of the parts was to be run off without splitting any of the lots -- despite the fact that lot-splitting is far from a rare occurrence. Once the simplifying assumption is made, it is then largely a matter of arithmetic to take the end item delivery schedules, pool this information with the "operation sheet" machining time estimates, and arrive at the man-hour requirements for machining during each of the months in which parts are to be produced for a particular end item. Given these estimates of requirements, in turn it is possible to subtract off the straight-time man-hours available from the existing work force, and come up with a figure for the deficit or surplus of manpower over requirements. In case of an impending deficit, it is up to the long-range planning group to recommend whether to order overtime work, to attempt outside procurement for certain of the parts that would normally have been made in the company's own plant, or to alter the initially stipulated schedule of parts deliveries to the final assembly operation. In practice, of course, a tight scheduling problem will force the planning group to depart from the assumption of no split lots, and thereby to depart from minimizing setup costs.

In the linear programming calculations - just as with current methods of scheduling - two simplifying features of this particular manufacturing operation are exploited. Neither is essential to the use of linear programming, but both are highly convenient for expository purposes: (1) Limitations on the availability of specific machines have been disregarded. It will ordinarily be true that if a particular production plan stays within the limitations of the manpower available with a particular time period, the plan will also be within the capabilities of the plant's machine tool equipment. (2) Inventory-holding costs have been neglected. Physical storage costs are quite low, and the contractual arrangement of Goverment "progress payments" makes the interest cost element a minor one.

2. A linear programming formulation

The linear programming model of the machine shop's operations is intended to provide numerical answers to the following general problem: Given a large number of individual parts to be machined, and given delivery requirements for each of these parts over a series of time periods, determine how many of each of the parts should be machined in each time period - taking account of the fact that there are limits upon the amount of straight-time and of overtime productive capacity available during the individual periods, and also that lot-splitting increases the total amount of setup time required.

In determining an output schedule, the objective is assumed to be the minimization of overtime labor requirement. This criterion for choice among alternative production plans implies: (a) that the straight-time services from the projected work force represent a fixed commitment on the company's part, and that nothing can be saved by failing to use up these services; (b) that the total labor requirements fall within the man-hours available from straight-time plus overtime work so that the question of outside procurement does not arise;* and (c) that the only remaining variable costs are those that

---

* From the viewpoint of model formulation, the question of outside procurement is an inessential complication. Ordering from an outside supplier differs from internal production only in that it costs money and imposes no drain upon the internal availability of labor.

From this same viewpoint, the question of recruiting and training new personnel is also an inessential complication. An activity of this sort could be incorporated directly within the model - provided that the training cost per man was known.

---

increase with the total number of overtime man-hours.

Underlying this linear programming formulation is the definition of an activity as a <u>sequence</u> of inputs over time that satisfies the delivery requirements for a particular part. (Individual parts are distinguished from one another by the subscript <u>i</u>, and the alternative sequences for the i <u>th</u> part by the subscript <u>ij</u>.) Since, in general, there will be more than one sequence that is feasible from the viewpoint of delivery requirements for the i <u>th</u> part, the linear programming variables x_{ij} refer to the fraction of the total requirement for the i <u>th</u> part that is supplied by the j <u>th</u> sequence of inputs for that part. Although no physical meaning can be attached to a fractional value of x_{ij} * (e.g., a solution specifying that half the requirements for a given

---

* For the special but nonetheless interesting case in which delivery requirements recur at a steady rate over the indefinite future, there is a meaningful interpretation that can be attached to fractional values of x_{ij} - i.e., that the actual lot size be intermediate between the quantities specified in the initial definition of the x_{ij} alternatives. With this interpretation, the non-linear inventory problem with storage and capacity restrictions described by Rifas in the Churchman, Ackoff, and Arnoff volume [2,ch.10] can be transformed into a straightforward exercise in linear programming.

---

part are to be met by a one-lot sequence of output and half by a split-lot sequence), there is no guarantee that such proper fractions will be absent from a linear programming solution. All that can be guaranteed is that if there are T time periods distinguished within the model, there will be at most T parts for which the x_{ij} fractions turn out to be intermediate between zero and one. (A proof of this assertion is given in section 5.)

Unknowns, coefficients, and constants for the programming model are defined in the following way:

(a) <u>unknowns</u>

x_{ij} = fraction of the total requirement for the i th part to be supplied by the j th alternative sequence of inputs. ($i = 1,\ldots,I; j = 1,\ldots,J$.)

ℓ_t = number of hours of overtime labor required during time period t ($t = 1,\ldots, T$.)

s_t = "slack" variable for straight-time labor during time period t. (all t)

v_t = "slack" variable for overtime labor during time period t. (all t)

(b) coefficients

β_{ijt} = labor input required during period t in order to carry out the j th alternative production sequence for part i. (all i, j, and t)

(c) constants

S_t = maximum availability of straight-time labor man-hours during the t th time period. (all t)

V_t = maximum availability of overtime labor man-hours during the t th time period. (all t)

With these definitions, the linear programming model becomes:

$$\text{(2.1)} \qquad \text{Minimize } \sum_t \ell_t$$

subject to:

$$\text{(2.2)} \qquad \sum_j x_{ij} = 1 \qquad (i = 1,\dots,I)$$

$$\text{(2.3)} \qquad \sum_i \sum_j \beta_{ijt}\, x_{ij} - \ell_t + s_t = S_t \qquad (t = 1,\dots,T)$$

$$\text{(2.4)} \qquad \ell_t + v_t = V_t \qquad (t = 1,\dots,T)$$

$$\text{(2.5)} \qquad x_{ij}\,,\ \ell_t\,,\ s_t\,,\ v_t \geq 0 \qquad (\text{all } i\,,\ j\,,\ t)$$

Expression (2.1) indicates the minimand - the sum of overtime labor requirements - and conditions (2.2) - (2.5) list the constraints that must be satisfied by the unknowns x_{ij}, ℓ_t, s_t, and v_t. Equations (2.2) say that the total requirement for the i th part must be met by a combination of one or more sequences of production for that part. (2.3) ensures that within each time period the total number of man-hours required to satisfy the individual output programs will not exceed the amount available of straight-time labor plus the overtime to be ordered for that period. Equations (2.4) place upper bounds upon the use of overtime labor. And finally, conditions (2.5) impose the usual nonnegativity requirements upon all unknowns.

To define more precisely what is meant by the x_{ij} variables and the β_{ijt} coefficients, it is easiest to refer to a three-period numerical example. ($T=3$.) Let the specific part under discussion be part 1, ($i = 1$), and let the deliveries scheduled, the setup time, and the incremental labor requirements for that part be:

a_1 = 10 man-hours = setup time for part 1.

b_1 = .9 man-hours / part = incremental labor required per unit of output of part 1.

R_{11} = 30 units = delivery requirements for part 1 at end of period 1

R_{12} = 30 units = " " " 2

R_{13} = 40 units = " " " 3

With these numerical values, there are exactly four alternative sequences for labor input and parts output that need to be considered explicitly within a linear programming model.* These four are distinguished from one another by

---

* If the programming model distinguishes between T time periods, there will be at most 2^T-1 distinct combination of periods within which some production could occur -- hence at most 2^T-1 "activities" for each parts category i. Furthermore, if the first period's delivery requirements are greater than zero ($R_{i1} > 0$), this upper bound becomes 2^{T-1}. Even for $T = 6$, $2^{T-1} = 32$, an easily manageable number of activities. Although strict logic compels the enumeration of all such lot-splitting possibilities, in practice it should not be difficult to reduce the number substantially by common-sense inspection.

the index $\underline{j}$.

j index	1	2	3	4
number of separate lots	1	2	2	3
delivery requirement to be produced in period 1.	$R_{11} + R_{12} + R_{13} =$ 100 units	$R_{11} =$ 30 units	$R_{11} + R_{12} =$ 60 units	$R_{11} =$ 30 units
delivery requirement to be produced in period 2.	0	$R_{12} + R_{13} =$ 70 units	0	$R_{12} =$ 30 units
delivery requirement to be produced in period 3.	0	0	$R_{13} =$ 40 units	$R_{13} =$ 40 units

It can be seen that each of the four output sequences just listed corresponds to one of the four possible combinations of periods in which a production setup occurs. $(2^{T-1} = 4)$ Once a particular combination is specified, the j th plan is uniquely determined by the rule that each delivery requirement is to be satisfied out of production during the nearest preceding period in which setup costs for that part are being incurred. It is not at all self-evident that the only production sequences deserving consideration are those indicated by this rule. At a later point, however, we shall prove that this is indeed the case, and that no reduction in overall costs can be achieved by substituting other output sequences in place of these. (see Appendix.) Hence, for the three-period model, the four output plans are said to "dominate" all others. Corresponding to these alternatives, the period-by-period inputs of labor required to satisfy the delivery requirements for part 1 - that is, the β_{1jt} coefficients are:

j index	1	2	3	4
x_{ij} unknown	x_{11}	x_{12}	x_{13}	x_{14}
β_{1j1}, period 1 input coefficients	$a_1 + 100\ b_1 =$ 100 man-hours	$a_1 + 30\ b_1 =$ 37 man-hours	$a_1 + 60\ b_1 =$ 64 man-hours	$a_1 + 30\ b_1 =$ 37 man-hours
β_{1j2}, period 2 input coefficients	0	$a_1 + 70\ b_1 =$ 73 man-hours	0	$a_1 + 30\ b_1 =$ 37 man-hours
β_{1j3}, period 3 input coefficients	0	0	$a_1 + 40\ b_1 =$ 46 man-hours	$a_1 + 40\ b_1 =$ 46 man-hours

3. <u>Aggregation of individual items</u>

It has already been emphasized that in order for the model described by (2.1)-(2.5) to be a useful one, the number of individual parts I must be quite large in relation to the number of time periods, T. (In the particular machine shop, this qualification creates few difficulties. The number of parts required for any one finished item would seldom amount to less than 100 distinct pieces.) But since the number of equations in the system equals $(2T + I)$, this also means that any conventional simplex computations of (2.1)-(2.5) would involve substantial costs. In general, there are two ways around a difficulty of this kind: One might be to construct a computing routine expressly designed to exploit the special structure of this linear programming matrix.* The other would be to

---

* If he examines the detached coefficients matrix associated with equations (2.2)-(2.4), the reader will observe that every basis of rank $(2T + I)$ that can be formed from this matrix may be partitioned in the following way:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

where A is an <u>identity</u> matrix of rank $(T + I)$, and D is a square matrix of rank T. The matrices B and C are rectangular - the former with $(T + I)$ rows and T columns, the later with T rows and $(T + I)$ columns. The numerical difficulties connected with solving a system of such equations are much closer to the order of magnitude of T, rather than of $(2T + I)$.

---

aggregate the original model in some suitable way, obtain an optimal linear programming solution to the aggregative model, and then translate this solution back into a detailed production plan for each part. This second course is the one that will be followed here.

The aggregation principle that seems most natural for this problem is to say that two parts belong to the same production category if they have a similar ratio of setup labor to total single-lot labor time, and if they also have a similar pattern of delivery requirements. In other words:

let R_{it} = delivery requirements for part i at end of period t

a_i = setup time for part i

b_i = incremental labor required per unit of output of part i

Then two parts ($i = 1$ and 2, respectively) are said to be in the same production category k if and only if there are two factors of proportionality α and λ such that:

$$(3.1) \qquad \frac{R_{2t}}{R_{1t}} = \lambda \qquad \text{(all } t\text{)}$$

and

$$(3.2) \qquad \frac{a_1}{a_1 + b_1 \sum_t R_{1t}} = \frac{a_2}{a_2 + b_2 \sum_t R_{2t}} = \alpha$$

If conditions (3.1) and (3.2) hold, then the labor input coefficients for the two parts will be related to one another by a single factor of proportionality - a factor equal to the ratio of the two setup cost coefficients:

$$(3.3) \qquad \frac{\beta_{2jt}}{\beta_{1jt}} = \frac{a_2}{a_1} = \frac{a_2 + b_2 \sum_t R_{2t}}{a_1 + b_1 \sum_t R_{1t}} \qquad \text{(all } j,t\text{)}$$

In other words, if conditions (3.1) and (3.2) apply, and if the j th setup sequence is an optimal one for part 1, it will also be an optimal one for part 2. Hence there is no reason to distinguish between the two parts within a linear programming model. All that needs to be done is to adopt one of them (e.g., part 1) as a standard unit of measurement, and then to express the aggregate requirement for that class of parts in equation (2.2) as:

$$(3.4) \qquad \text{aggregate requirement} = 1 + \frac{a_2}{a_1} = 1 + \frac{a_2 + b_2 \sum_t R_{2t}}{a_1 + b_1 \sum_t R_{1t}} \quad .$$

By following this principle of aggregation, it will ordinarily be possible to make a substantial reduction in the number of equations listed in (2.2), and so reduce the burden of computations without lessening the inherent accuracy of the linear programming model.

In practice, the aggregation conditions (3.1) and (3.2) do not seem unduly stringent. Conditions (3.1) say, e.g., that if at the end of period $\underline{t}$, 25% of the total requirement for part 1 is to become available for final assembly, then 25% of the total for part 2 must also become available at that time. When both parts are required for the same end item, the timing of delivery requirements will usually be identical, and so there should be no difficulty in constructing a small number of groups such that each part within a given class will satisfy conditions (3.1). Once this kind of preliminary grouping has been effected, it should be easy to define production categories $\underline{k}$ that also satisfy conditions (3.2) - at least to whatever degree of approximation is warranted by the goodness of the original estimates of a_i, b_i, and R_{it}. Table 1 illustrates this point for the case of one typical end item actually produced by our manufacturer - an end item requiring 110 distinct parts, each with the same pattern of delivery requirements. Here the quality of the raw data was such that the six-category classification scheme shown for these parts in Table 1 appeared entirely satisfactory for purposes of long-range production planning.

In following through the aggregation procedure just described, it seems convenient to define the unit of measurement - i.e., the "standard" part in each production category $\underline{k}$ - as one for which the total of setup time plus single-lot running time equals one hour. The requirement for all parts in category $\underline{k}$ may then be expressed in terms of this standard as follows:

$$
(3.5) \qquad Q_k = \text{aggregate number of "standard" hours' worth of parts in category } \underline{k}
$$

$$
= \sum_{i \epsilon k} \left(a_i + b_i \sum_t R_{it}\right)
$$

Table 1.

A system of aggregation for 110 distinct parts, as classified by setup labor ratio α_i

Production category k	Class interval for the setup labor ratio $\alpha_i = \dfrac{a_i}{a_i + b_i \sum_t R_{it}}$ (all $\underline{i} \epsilon \underline{k}$)	Number of distinct parts within category $\underline{k}$	Maximum number of "standard" hours required for any single part in category $\underline{k}$	Total "standard" hours required for all parts in category $\underline{k}$ $= Q_k$
1	$0 \leq \alpha_i <$ .10	9 parts	1,046 man-hours	4,064 man-hours
2	.10 " .20	33 "	567 "	4,774 "
3	.20 " .30	32 "	176 "	2,097 "
4	.30 " .40	21 "	90 "	654 "
5	.40 " .50	11 "	66 "	286 "
6	.50 " 1.00	4 "	34 "	98 "
	total	110 "		11,973 "

To summarize: Given the labor input coefficients a_i and b_i and also the delivery requirements R_{it} for each of many distance parts <u>i</u>, a small number of production categories will ordinarily suffice for the purpose of an aggregative linear programming model. Furthermore, once an optimal solution has been calculated, there should be no difficulty in translating the aggregative results back into a detailed plan for the production of each distinct part.

What makes such a translation possible? e.g., what if the linear programming solution called for 1,000 hours' worth of parts in a given category to be produced in a single lot, and 1,000 hours" worth by a split-lot plan? It is perfectly true that no sense could be made of a detailed plan that called for producing half of <u>every</u> distinct part with a single-lot program and half with split lots. But it would make perfectly good sense to translate the aggregative solution into a detailed plan that called for producing one distinct group of parts according to the single-lot plan and another group according to the split-lot plan -- provided that the total "standard" time for parts in each of these two groups came to 1,000 hours apiece. The whole trick consists of observing that when the number of distinct parts in large, and that when one is dealing with groups of such parts, the alternative production programs are not mutually exclusive, and that under these conditions, one can always spell out a meaningful detailed plan for any convex combination of the stated alternatives.

Whenever the number of distinct parts in a production category exceeds more than a handful, there should be no serious difficulty in translating the aggregative solution back into a detailed program for the output of each part.*

---

* If the reader insists upon some precision in the definition of a "handful," and if he is willing to recognize that the a_i, b_i, and R_{it} parameters are each a bit fuzzy, I would venture the guess that no real translation difficulities will occur if the number of distinct parts within a given category $\underline{k}$ exceeds 10, and if the maximum time required for any single part in a given category is less than 20% of the total. On this score, see Table 1.

The formulation of the model ensures that not only will the total parts requirement be satisfied in terms of "standard" units, but also that the production of each of these parts can be time-phased in such a way as to satisfy the initially stipulated delivery requirements.

4. A numerical example

This illustrative example will refer to a case involving three time periods and five production categories. In following through the calculations, the first step is to obtain numerical values for the setup time ratios α_k and the percentage delivery requirements $R_{kt} \div \sum_t R_{kt}$ within each of the five production categories $\underline{k}$. These parameters, along with the constants Q_k, S_t, and V_t, are all listed in Table 2. With this information available, it is then a straightforward matter to construct the β_{kjt} labor input coefficients, and then the matrix of detached coefficients (Table 3) for the linear programming model indicated abstractly by conditions (2.1)-(2.5).*

---

* Except for the position of the decimal point, the β_{1jt} that appear in Table 3 are identical with those calculated on p. (12) above. All other β_{kjt} were obtained by a similar process.

The only change introduced by the aggregation procedure is the replacement of the index $\underline{i}$ by the index $\underline{k}$ ranging in value from 1 to K. That is, instead of x_{ij} variables which represent the fraction of the requirement for the i th part supplied by the j th alternative sequence, we now have x_{kj} variables which represent the total number of "standard" hours' worth of parts in category $\underline{k}$ that are to be produced by the j th sequence. Along with this change, it is, of course, necessary to replace the constants of unity in equations (2.2) with the Q_k, the total number of "standard" hours' worth of parts required in category $\underline{k}$.

Altogether this system involves 25 unknowns and 11 equations. Of the unknowns, 16 are of the x_{kj} type, and there are three each of the ℓ_t, s_t, and v_t type.*

---

* The reader may wonder why only two alternative programs ($\underline{j} = 1$ and 2) are listed for parts categories 4 and 5. In strict logic, even though no delivery requirements for these parts exist during period 1, one should still consider the possibility of producing them during that period as well as during 2 and 3. But period 1 production of these items would only be profitable if, in an optimal solution, the "shadow price" associated with labor in period 1, turned out to be lower than that associated with labor in period 2. Since the *a priori* considerations were against this outcome, all activities corresponding to positive amounts of period 1 output were omitted from the linear programming tableau shown in Table 3. As things worked out, the optimal solution substantiated these conjectures, and so nothing was lost by discarding the possibility of period 1 output for parts categories 4 and 5.

Since the matrix shown in Table 3 indicates non-zero coefficients only, the first row of that matrix (numbered 0) contains just three entries -- the cost coefficients associated with the three overtime labor variables ℓ_t in the minimand, expression (2.1). Following the minimand, the next five rows correspond to equations (2.2), the requirements for output in each of the five parts categories.

Then come the three rows numbered 6,7, and 8 -- one for each time period -- constraining the input of labor to fall within the number of man-hours available. Rows 6,7 and 8 correspond therefore to equation group (2.3). And finally, the last three rows (numbers 9,10, and 11) coincide with equations (2.4) -- the upper bound conditions upon the use of overtime work in any one time period.

Using the simplex method of calculation, and taking advantage of the special structure of the matrix shown in Table 3, it proved to be an easy matter to calculate an optimal solution to this model, and to determine that the optimum was unique. The optimal solution, along with the corresponding "shadow prices" or "dual variables," is shown in Table 4. According to this solution, it pays to use a one-lot production plan for the output of every part in categories 3 and 5. (Since these two categories are the ones for which the setup cost parameter α_k is largest, this outcome is an entirely reasonable one.) All parts in category 2 are to be produced in two lots -- 60% of the output in period 1, the remainder in period 3, and none at all in period 2. And in the case of both categories 1 and 4, it pays to combine two lot-splitting plans. That is, 1,915 "standard" hours' worth of parts in the first category are to be produced by splitting production between time periods 1 and 3, and the remaining parts in that category by splitting production among all three time periods. Similarly, 1,479 hours' worth of parts in category 4 are to be turned out in a single lot during period 2, and the remaining output of 3,321 is to be obtained by splitting production between periods 2 and 3.

Table 2

Parameters and constants for the numerical example

	Parts category $\underline{k}$				
	1	2	3	4	5
α_k	.1	.2	.3	.2	.3
$R_{k1} \div \sum_t R_{kt}$	.3	.3	.3	0	0
$R_{k2} \div \sum_t R_{kt}$	.3	.3	.3	.4	.4
$R_{k3} \div \sum_t R_{kt}$	.4	.4	.4	.6	.6
Q_k	3,500	4,100	2,900	4,800	3,200

Time period, $\underline{t}$	S_t	V_t
1	6,000	1,500
2	6,000	1,500
3	6,000	1,500

Table 3.

Matrix of detached coefficients. A numerical example of the linear programming model (2.1) -- (2.5)

Equations \ Unknowns		x_{kj}																ℓ_t			s_t			v_t			Constants Q_k, S_t, V_t	
		x_{11}	x_{12}	x_{13}	x_{14}	x_{21}	x_{22}	x_{23}	x_{24}	x_{31}	x_{32}	x_{33}	x_{34}	x_{41}	x_{42}	x_{51}	x_{52}	ℓ_1	ℓ_2	ℓ_3	s_1	s_2	s_3	v_1	v_2	v_3		
0) Cost coefficients																		1.00	1.00	1.00							-	(0
1) Output of parts, category	1	1.00	1.00	1.00	1.00																						3,500	(1
2) " " "	2					1.00	1.00	1.00	1.00																		4,100	(2
3) " " "	3									1.00	1.00	1.00	1.00														2,900	(3
4) " " "	4													1.00	1.00												4,800	(4
5) " " "	5															1.00	1.00										3,200	(5
6) Labor input, period	1	1.00	.57	.64	.57	1.00	.44	.68	.44	1.00	.51	.72	.51					-1.00			1.00						6,000	(6
7) " " "	2		.73		.57		.76		.44		.79		.51	1.00	.52	1.00	.58		-1.00			1.00					6,000	(7
8) " " "	3			.46	.46			.52	.52			.58	.58		.68		.72			-1.00			1.00				6,000	(8
9) Overtime, period	1																	1.00						1.00			1,500	(9
10) " "	2																		1.00						1.00		1,500	(10
11) " "	3																			1.00						1.00	1,500	(11

Table 4

Values of dual variables and of non-zero primal variables in the optimal solution

a) Non-zero primal variables

Parts category i	Initial parameters and constants		Values of variables, in "standard" hours		
	α_k	Q_k	One-lot plans	Two-lot plans	three-lot plans
1	.1	3,500 hours	-	$x_{13}=$ 1,915	$x_{14}=$ 1,585
2	.2	4,100	-	$x_{23}=$ 4,100	-
3	.3	2,900	$x_{31}=$ 2,900	-	-
4	.2	4,800	$x_{41}=$ 1,479	$x_{42}=$ 3,321	-
5	.3	3,200	$x_{51}=$ 3,200	-	

man-hours

Time period t	ℓ_t (overtime)	s_t (slack)	v_t (slack)
1	1,500	-	-
2	992	-	508
3	-	-	1,500
Minimand $= \sum \ell_t =$	2,492		

Table 4 (cont'd)

b) Dual variables, change in minimand per unit change in value of the constant associated with the particular equation.

Output requirement equations (2.2)

$u_1 = 1.202$	overtime hours / "standard" hour's worth of parts category	1
$u_2 = 1.299$	" "	2
$u_3 = 1.370$	" "	3
$u_4 = 1.000$	" "	4
$u_5 = 1.000$	" "	5

Labor availability equations (2.3)

$u_6 = -1.370$	overtime hours / hour's worth of straight-time labor in time period	1
$u_7 = -1.000$	" "	2
$u_8 = -.706$	" "	3

Overtime limitation equations (2.4)

$u_9 = -.370$	overtime hours / hour's worth of overtime labor in time period	1
$u_{10} = 0$	" "	2
$u_{11} = 0$	" "	3

The dual variables for each equation $(u_1, u_2, \ldots, u_{11})$ measure the potential change in the minimand (overtime labor cost) per unit change in the constant associated with that equation. An extra hour's worth of straight-time labor available in period 2, for example, would make it possible to reduce the total amount of overtime by exactly one hour. Hence $u_7 = -1$. But an extra hour available in period 1 could be employed so as to avoid a substantial amount of lot-splitting, and for this reason $u_6 = -1.370$.* Such values are immediately suggestive

---

* Although it will not necessarily always be true $u_6 \leq u_7 \leq u_8 \leq 0$, this ranking will hold whenever: (a) inventory costs are negligible, and (b) the output sequences are defined so that production of each item is permitted in any of the time periods prior to delivery.

---

of "break-even" points for the worth of additional labor in the machine shop beyond the amounts already assumed available. Similarly the dual variables $u_1, \ldots, u_5$ - those associated with the output requirement equations (1)-(5) - are indicative of the incremental worth of any external supply of parts in each of these five categories.

One theorem about the properties of the first five dual variables will be asserted without proof: If all parts within two categories satisfy condition (3.1) but not (3.2), and if the first category's setup time parameter is lower than that of the second, then the "implicit cost" of meeting an additional hour's worth of output requirements in the first category will be no higher than that of an hour's worth in the second category. Hence $0 \le u_1 \le u_2 \le u_3$. Also $0 \le u_4 \le u_5$.

If one were concerned purely with the formal aspects of this economic lot size problem, the discussion of the numerical example could end at this point. Given the optimizing criterion and the constraints listed in (2.1) -- (2.5), an optimal solution has been produced for the aggregative scheduling problem, and in principle it has been shown how this could be translated back into a detailed time-phased plan for the output of each distinct part. But if one's interest is with the actual managerial problem that is represented by this model, something more needs to be said. In our idealization of the machine shop's activities, all interactions have been neglected between the machine shop and the final assembly area. In particular, we have ignored the possibility that by splitting parts production within the machine shop, we may disrupt the smooth flow of final assembly work on any one series of end items. To the extent that this intermittent pattern of final assembly costs more than a continous flow of work, the "sub-optimization" calculated for the machine shop is a misleading one. This does not mean that the linear programming analysis is useless -- only that the results of this analysis have to be integrated with what is also known about the final assembly operation.

Here, for example, one of the men actually responsible for production planning suggested that it might be possible to transfer skilled final assembly machinists from their usual jobs, and to bring them temporarily into the machine shop to help meet the initial period's peak demand there. Since this proposal would make it possible to avoid all lot-splitting, it contains several attractive features -- not only the obvious reduction in setup costs,* but also the very

---

* If manpower were available early enough to make single-lot production possible for all parts, the actual time of 20,492 man-hours $(\Sigma S_t + \Sigma \ell_t)$ could be reduced to the "standard" time of 18,500 hours $(Q_1 + Q_2 + Q_3 + Q_4 + Q_5)$. The excess labor requirement for split-lot production amounts therefore to 1,992 man-hours.

---

real benefits to be derived in the final assembly area by having 100% of every part available at the time that final assembly is initiated for any one series of end items. Against both of these prospective benefits, it is, of course, also necessary to evaluate the immediate cost of disrupting final assembly activities by such a temporary transfer. The linear programming analysis of the machine shop cannot by itself indicate that such transfers would be in the best interests of the plant as a whole, but it can at least indicate the order of magnitude of the direct labor savings inherent in single-lot production of all parts. Surely this calculation is not the only thing relevant to the question of whether workers ought to be transferred temporarily, but it does represent one of the pieces of information needed in order to arrive at a sound decision.

5. A theorem on the occurrence of fractional values for the x_{ij} variables

At an earlier point, it was convenient to assert without proof that the applicability of the linear programming proposal did not depend upon the possibility of aggregating distinct parts into the output categories defined by (3.1) and (3.2), but only upon the existence of a large number of distinct parts $\underline{i}$ - each of them with a small labor input requirement by comparison with the total availability of labor. The precise form of this assertion is as follows: Consider the model described by (2.1)-(2.5). Then if there are I parts and T time periods, in every basic feasible solution there will be at least (I - T) parts for which exactly one x_{ij} variable is operated at a positive level. Thus, except for at most T parts, the linear programming solution will immediately indicate a detailed feasible time-phased plan for the output of each item in the machine shop. For each of these T parts, there is indeed the possibility that the linear programming solution will require half the lot to be produced according to a one-lot plan and half according to a split-lot plan. The physical absurdity of such a solution is obvious, but if T is sufficiently small in relation to I, the few parts that will be affected should cause no difficulty from the viewpoint of long-range planning. Another way to state this result is to say that when the number of parts to be scheduled far exceeds the number of individual time periods (a reasonable enough assumption when one end item alone may contain 110 distinct components), the very multiplicity of parts acts in such a way as to smooth out the "lumpiness" associated with setup costs.

The preceding theorem may be restated as follows:

(5.1) If, in a basic feasible linear programming solution to the model indicated by (2.1) -- (2.5), there are $\underline{m}$ parts for which exactly one x_{ij} variable appears at a positive intensity in that solution, then $m \geq I - T$.

Proof Let n represent the number of the variables ℓ_t and v_t operated at positive levels in the particular basic feasible solution. (In order for equations (2.4) to be satisfied, $n \geq T$.) The expression $(I - m)$ represents the number of parts for which two or more x_{ij} variables are operated at a positive intensity in the particular solution. Now since there are altogether $(2T + I)$ restraint equations listed in this model, at most $(2T + I)$ variables will appear at positive level in the solution. I.e.:

(5.2) $2T + I \geq n + m + 2(I-m)$

and since $n \geq T$

(5.3) $T + I \geq m + 2(I-m)$

$\therefore$ (5.4) $m \geq I - T$, which was to be proved.

6. Summary

This paper may be recapitulated as follows: Starting with a production scheduling problem that involves indivisibilities in the form of setup costs, a linear programming model has been constructed that is not identical with the original problem, but which provides an excellent approximation when the number of distinct parts is large in comparison with the number of time periods, T . In this approximation to the original problem, the variables do not refer to the size of each production lot within each time period, but rather to the fraction of the total requirement for any given part that is satisfied by a particular sequence of production for that part. The linear programming formulation ensures that, except for at most T individual parts, these fractions will all turn out to be either zero or one. With this exception, therefore any "basic feasible" solution will automatically avoid the possibility of meeting one portion of the requirement for a given part by a one-lot program of output and another portion of the requirements with a split-lot program. Although this physically absurd option is built into the model, a theorem ensures that the option will be exercised only rarely.

How serious a distortion of reality is implied by a linear programming solution that calls for the production of a few parts in this physically absurd manner? From a purely abstract standpoint, such a solution is completely infeasible, and it is easy to construct numerical examples for which the linear programming solution could not be "patched up" without a large increeace in the total system costs. Despite this perfectly valid formal objection, it may seriously be doubted that this difficulty really detracts from the usefulness of the model. The detailed optimal solution to such a model is hardly intended as a literal forecast of production activities up to eighteen months in the future, but only as a guide to making a number of immediate decisions that will affect the future - overtime, recruiting and training of new personnel, and outside procurement of certain parts. For the purpose of choosing among these broad alternatives - although not for the detailed short-run scheduling problem - the few apparent infeasibilities should be of minor significance.

This same line of reasoning should do much to dispel another kind of objection that may be raised against the model presented here. The usefulness of this proposal depends upon the magnitude of the number of distinct time periods, T. Since the number of alternate production activities to be enumerated for a single part category is of the order of 2^T, the capacity of current computing equipment would not be taxed by a model with $T \leq 8$, but would clearly be swamped for $T \geq 15$.* Certainly there is no guarantee that it will always be satisfactory

---

* Even for large values of T, it would still be possible to enumerate just a small number of the more interesting alternative production programs for any one part. A solution based upon such an incomplete enumeration would still be a feasible one, and should be near-optimal - even though no a priori guarantees can be made as to its optimality.

---

to plan production over an 18-month period in time units as large as one to three months. Indeed a determined critic would be within his rights in pointing out that it might be necessary to plan a single year's operation ahead in terms of 365 individual time units - each one day in length. The answer to such a critic can only come from a study of the empirical problem to which the model is to be applied. Assuming that the purpose of the model is to aid in answering certain broad questions dealing with overtime, outside procurement, etc., it should not be a serious limitation upon the problem formulator for him to keep the value of T well within the limits of present-day computing feasibility.

7. Significance of the results

The production scheduling example discussed in this paper is by no means an isolated instance in which, starting with a problem that entailed indivisibilities in terms of one set of variables, it was nevertheless possible to redefine the variables so as to transform the original problem into a new one that could be studied from the computational viewpoint of linear programming. This same approach has already been illustrated in the newsprint trim problem [5], in the coat-and-pants problem [3], in Salveson's machine loading problem [6, pp. 234-245], and doubtless in others. There appears to be an entire class of optimization problems that involve indivisibilities in terms of one set of variables, but which can nevertheless be translated into the linear programming format. Some precise characterization of this class of problems seems to be needed, but is lacking at present.

Although the economist's primary interest is not in numerical analysis, but rather in the possibility of market analogue solutions to welfare maximization problems, the indivisibility of setup costs places him in an awkward position. As long as he regards the individual "activity" as one of determing the lot size for a given part in a particular time period, there need be no set of intra-firm shadow prices that is compatible with a cost-minimizing equilibrium, and hence no possibility of a market analogue solution. The curious aspect of the production problem outlined here is that it is possible to redefine activities and commodities so as to end up with a linear programming system -- i.e., one for which, in principle, a market analogue solution is possible. From the viewpoint of the theory of market decentralization, the chief feature of this alternative version is that the individual activities represent a greater degree of vertical integration than is assumed in the inititial statement of the problem. Paradoxically enough, successful decentralization requires that the manager of each activity have a longer "span of control" than the size of the individual lot in a particular time period. It is necessary for each such manager to be familiar with the entire program of labor inputs that is implied by his particular sequence of output for the individual part.

Appendix.

"Dominance" properties of the set of alternative production programs for a given item.

The machine shop is engaged in producing a number of items with a resource input that is homogeneous except for date. If an item is produced in the t th time period $(t = 1,2,\ldots T)$, resource inputs are required from the total available in that period, but from no other. The amount of resources used in the t th period by producing x_t units of a particular item is given by:

$$(1) \qquad a\,\delta_t + b\,x_t$$

where $\quad x_t \begin{pmatrix} > 0 \\ = 0 \end{pmatrix} \quad$ implies $\quad \delta_t \begin{pmatrix} = 1 \\ = 0 \end{pmatrix}$

The non-negative constant a is said to represent the "setup cost" for that item and the non-negative constant b the "incremental unit cost." Since $\delta_t = 0,1$, there are altogether 2^T column vectors of the following form:

$$(2) \qquad \Delta_j = \begin{pmatrix} \delta_{j1} \\ \delta_{j2} \\ \vdots \\ \delta_{jt} \\ \vdots \\ \delta_{jT} \end{pmatrix}$$

Now suppose that the firm is to deliver R_t units of the item in the t th period. Corresponding to each of the 2^T vectors Δ_j, the time phased production vector X_j may be written, where:

$$X_j = \begin{pmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jt} \\ \vdots \\ x_{jT} \end{pmatrix} \tag{3}$$

and where the output levels x_{jt} are determined according to either (4), (5), or (6). These conditions are equivalent to the rule that each delivery requirement be satisfied out of production during the nearest preceding period in which setup costs are being incurred:

(4) if $\delta_{jt} = 0$, then $x_{jt} = 0$

(5) if $\delta_{jt} = \delta_{j,t+1} = 1$, then $x_{jt} = R_t$

(6) if $\delta_{jt} = 1$, and $\bar{\tau}$ is the largest value of τ such

that $\delta_{j,t+\tau} = 0$, then $x_{jt} = \sum_{\tau=0}^{\bar{\tau}} R_{t+\tau}$

The setup plan Δ_j and the corresponding output plan X_j are said to be "feasible" from the viewpoint of delivery requirements if the components of X_j also satisfy:

$$(7a) \qquad \sum_{\tau=1}^{t} x_{j\tau} \geq \sum_{\tau=1}^{t} R_\tau \qquad (\underline{t} = 1,2,\ldots,T-1)$$

and

$$(7b) \qquad \sum_{\tau=1}^{T} x_{j\tau} = \sum_{\tau=1}^{T} R_\tau$$

For each of the "feasible" Δ_j and X_j vectors, the resource input column vector β_j may be defined as follows:

$$(8) \quad \beta_j = a\,\Delta_j + b\,X_j \qquad (j = 1,\ldots,J)$$

The $T \times J$ matrix B is composed of the vectors β_j:

$$(9) \quad B = (\beta_1,\beta_2,\ldots,\beta_j,\ldots,\beta_J)$$

Now let the "implicit value" or "shadow price" of any resources used in the t th time period be represented by u_t. ($u_t \leq 0$; $u_t \gtrless u_{t-1}$) . The column vector formed from these components is termed U :

$$(10) \qquad U = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_t \\ \vdots \\ u_T \end{pmatrix}$$

Suppose that a setup plan Δ_r and a production plan X_r satisfy conditions (11) - (13):

$$(11) \qquad x_{rt} \begin{pmatrix} > 0 \\ = 0 \end{pmatrix} \text{ if and only if } \delta_{rt} \begin{pmatrix} = 1 \\ = 0 \end{pmatrix}$$

$$\text{(12a)} \qquad \sum_{\tau=1}^{t} x_{r\tau} \geq \sum_{\tau=1}^{t} R_{\tau} \qquad (t = 1,2,\ldots,T\text{-}1)$$

$$\text{(12b)} \qquad \sum_{\tau=1}^{T} x_{r\tau} = \sum_{\tau=1}^{T} R_{\tau}$$

and $$\text{(13)} \qquad x_{rt} \geq 0 \qquad (\text{all } \underline{t})$$

Denote by β_r the vector of resource inputs that is required in order to carry out this production plan:

$$\text{(14)} \qquad \beta_r = a\,\Delta_r + b\,X_r$$

<u>"Dominance" theorem:</u> If the vector $U \leq 0$, there is no pair of vectors Δ_r and X_r satisfying conditions (11) - (13) for which it is also true that:

$$\text{(15)} \qquad U'\beta_j < U'\beta_r \leq 0 \qquad (\text{all } \beta_j \in B)$$

In words, this theorem says that if the production program X_r is feasible from the viewpoint of delivery requirements, then there will always be at least one program X_j within the previously enumerated set that has an implicit cost at least as low as that for program X_r. This is the sense in which the set of resource input vectors B is said to "dominate" all others.

<u>Proof:</u>*

---

* I am indebted to H. Houthakker for a general outline of this proof, and to M. Beckmann and S. Winter for their constructive criticisms.

In order to prove this theorem, we first observe that if the vectors Δ_r and X_r satisfy (11) - (13), there will be exactly one vector $\beta_s \in B$ such that $\Delta_r = \Delta_s$. (The symbol X_s will denote the particular output vector associated with β_s .) The T- component "error" vector η (1) will be defined by:

(16) $$\eta\ (1) = X_r - X_s$$

If η_t (1) is the t th component of η (1), and if X_r permits a feasible schedule of deliveries, then clearly

(17a) $$\sum_{\tau=1}^{t} \eta_\tau\ (1) \geq 0 \qquad (t = 1,2,\ldots,T-1)$$

and

(17b) $$\sum_{\tau=1}^{T} \eta_\tau\ (1) = 0$$

Lemma:

If $U'\ \beta_s < U'\ B_r \leq 0$, there exists a pair of vectors X_{r+1} and Δ_{r+1}, and also a pair of vectors X_{s+1} and Δ_{s+1} (with $\beta_{s+1} \in B$) such that:

(18) $$\Delta_{r+1} = \Delta_{s+1}$$

(19) $$\eta\ (2) = X_{r+1} - X_{s+1}$$

(20a) $$\sum_{\tau=1}^{t} \eta_\tau\ (2) \geq 0 \qquad (t = 1,2,\ldots,T-1)$$

(20b) $$\sum_{\tau=1}^{T} \eta_\tau\ (2) = 0$$

(21) $$U'\ \beta_r \leq U'\ \beta_{r+1}$$

and (22), η (2), with a total of T components, has at least one more zero component than η (1).

By repeated applications of this lemma, it follows that if the error vector $\eta\ (1)$ has $\underline{n}$ non-zero components, then $\eta\ (n+1) = 0$, and $X_{r+n} = X_{s+n}$, with $\beta_{s+n} \in B$. But by combining this result with (15) and (21):

$$U'\ \beta_j < U'\ \beta_r \leq U'\ \beta_{r+1} \leq \cdots \leq U'\ \beta_{r+n} = U'\ \beta_{s+n} \qquad (\text{all } \beta_j \in B) \tag{23}$$

At this point, we observe that $\beta_{s+n} \in B$, and so condition (23) cannot be satisfied. Therefore if the lemma is true, then the "dominance" theorem is also true.

<u>Proof of lemma</u>:

Let t^* represent the last time period for which $\eta_t\ (1) \neq 0$. By (17a) and (17b):

$$\eta_{t^*}\ (1) = - \sum_{\tau=1}^{t^*-1} \eta_\tau\ (1) < 0\ . \tag{24}$$

Also let t^{**} represent the last period for which $\eta_t\ (1) > 0$. (Note that $t^* > t^{**}$.) In forming the new vector X_{r+1} from the preceding one, calculate the individual components according to the following rules:

(25)

	value of $x_{r+1,t}$	
$\underline{t}$	If $(u_{t^{**}} - u_{t^*}) \geq 0$, then	If $(u_{t^{**}} - u_{t^*}) < 0$, then
t^{**}	$x_{rt^{**}} + x_{rt^*}$	$x_{rt^{**}} + \eta_{t^*}\ (1) \geq 0$
t^{*}	0	$x_{rt^*} - \eta_{t^*}\ (1) = x_{st^*}$
$t \neq t^{**}, t^{*}$	x_{rt}	

Corresponding to the new vector X_{r+1}, there will be exactly one vector $\beta_{s+1} \in B$ such that $\Delta_{r+1} = \Delta_{s+1}$, and the T- component "error" vector η (2) will be defined by:

$$(26) \qquad \eta\ (2) = X_{r+1} - X_{s+1}$$

Regardless of the sign of $(u_t^{**} - u_t^{*})$, it can be verified that $U'\ \beta_r \leq U'\ \beta_{r+1} \leq 0$. Also, in either case the error vector η (2) resulting from the construction outlined in (25) has at least one more zero component than η (1) . The new vectors X_{r+1}, X_{s+1}, and η (2) satisfy conditions (18) - (22), and so the lemma is proved.

References

1. Beckman, M., and Muth, R., "An Inventory Policy for a Case of Lagged Delivery," Management Science, Vol. 2, No. 2, January 1956, pp. 145-155

2. Churchman, C.W., Ackoff, R.L., and Arnoff, E.L., Introduction to Operations Research, John Wiley, 1957.

3. Dorfman, R., Solow, R., and Samuelson, P. A., Linear Programming and Economic Analysis, McGraw-Hill, 1957

4. Markowitz, H.M., and Manne, A.S., "On the Solution of Discrete Programming Problems," Econometrica, Vol. 25, No. 1, January 1957

5. Paull, A.E., and Walter, J.R., "The Trim Problem: An Application of Linear Programming to the Manufacture of Newsprint Paper," (abstract) Econometrica, Vol. 23, No. 3, July 1955, p. 336

6. Salveson, M.E., "A Problem in Optimal Machine Loading," Management Science, Vol. 2, No. 3, April 1956