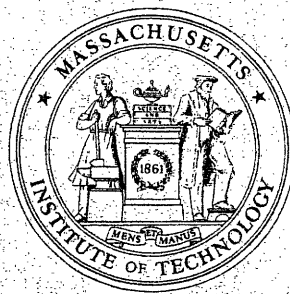


OPERATIONS RESEARCH CENTER

working paper



**MASSACHUSETTS INSTITUTE
OF TECHNOLOGY**

Optimal Capacity Expansion Planning
When There Are Learning Effects*

by

Randall S. Hiller and Jeremy F. Shapiro

OR 121-83

September, 1983

*Research supported by the National Science Foundation under
Grant No. ECS-8117105.

Optimal Capacity Expansion Planning When There Are Learning Effects

by

Randall S. Hiller and Jeremy F. Shapiro

1. Introduction

The learning or experience curve is one of the earliest management science tools, its development dating back to World War II when it was used to forecast aircraft production rates ([13]). Starting in the late 1960's, several consulting companies, notably the Boston Consulting Group [2], expanded the scope of the curve by applying it to the output of an entire industry in order to analyze competition among companies in that industry. Strategic plans for a particular company were related to its position relative to other companies on the industry learning curve. See [4] for a recent survey of these applications and [5] for discussions of the learning curve in the more general context of strategic planning.

More recently, articles have appeared in the economics literature ([6], [10], [11], [12]), containing mathematical models that formalize the insights of the consulting studies. The models are simplified descriptions of markets, industries and companies within the markets. Their purpose is to expose fundamental relationships that the authors believe exist among a small number of key variables which a company can adjust to control output, market position, and their competitors' behavior.

The purpose of this paper is to show how mathematical programming models can be used for analyzing a company's capacity expansion plans in the face of significant learning effects in production, and in the market. Mathematical programming models differ from those appearing in the literature cited above by the degree to which they are data driven, their explicit incorporation of resource constraints, and their ability to capture discontinuities and economies of scale which are intrinsic to a company's expansion plans.

Learning curves provide the decision maker with options regarding returns to scale that are spread out over time. Mixed integer programming (MIP) modeling techniques can readily capture this phenomenon, but the resulting models may be difficult to optimize. Thus, more than the usual amount of attention must be given to the interplay between model formulation and optimization. For this reason, we present in this paper a simple MIP model intended to illustrate the important modeling and analytic concepts. This model, which we call the study model, is presented in sections 2 and 3. Although models based on linear programming incorporating learning effects have appeared in the management science literature (e.g., [3], [7]), the development here appears to be the first to use MIP. Section 4 contains the details of an implemented study model and results of several optimization runs. In the final section, we discuss extensions of the study model encompassing additional details that would be required by an actual application. We also discuss briefly in that section future modeling research directed at other aspects of long range planning related to learning such as uncertainties in production and the market, and competitors' behavior.

2. Study Model and Optimization Conditions

The model that we propose to study in detail addresses the multi-period planning problems faced by a company that manufactures and sells a single product for which there are significant learning effects. These effects are manifested by unit manufacturing costs that decrease as a function of cumulative production. We also assume that market prices for the product decrease over time reflecting industry wide learning and competition among companies in the industry. The company must simultaneously decide

how quickly it wants to get out on the learning curve, and how it wants to invest over time in expanding manufacturing capacity in order to implement this plan.

Let $t = 1, \dots, T$ denote the time periods of the planning horizon. For each period, the company must decide how much to produce, denoted by x_t , and how much to sell, denoted by d_t . The values of these variables determine in the usual manner the change in inventory level from the end of the previous period to the end of the current period; namely, if i_t denotes ending inventory in period t , we have $i_t - i_{t-1} = x_t - d_t$. Holding costs are h_t per unit of ending inventory. A salvage credit v_T is given for each unit of inventory at the end of the planning horizon.

Marginal manufacturing costs decline with cumulative production y as depicted in Figure 1. Letting $C(y)$ denote the total manufacturing cost to produce y units, the cost of producing x_t units in period t is given by

$$C(x_1 + \dots + x_t) - C(x_1 + \dots + x_{t-1}).$$

A specific functional form for $C(y)$ is given in section 4 where we discuss an implemented version of the study model. Throughout the paper, we assume that C is a concave function.

The company's capacity expansion options over the T periods are described as follows. Analysis reveals K discrete alternatives from which at most one can be selected. Associated with alternative k are the T scalars b_{k1}, \dots, b_{kT} , where b_{kt} denotes the capacity addition in period t under option k . Also associated with alternative k is a net present value I_k equal to the sum of discounted costs required to achieve the associated capacity additions minus a discounted salvage credit for equipment in place at the end

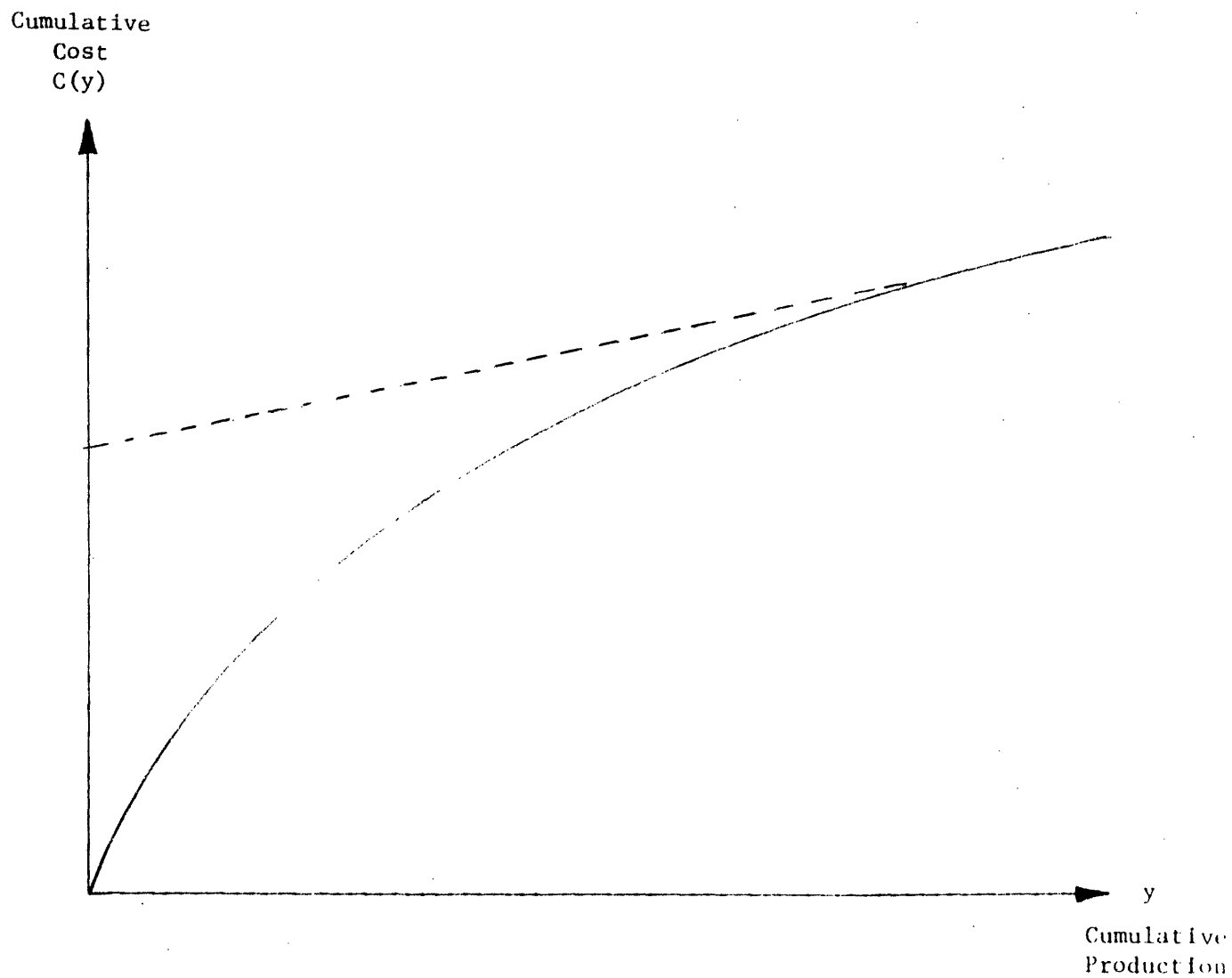


Figure 1

of the horizon.

The quantities I_k may reflect returns to scale in construction in the following sense. Suppose capacity expansion alternatives k and j refer to the same physical construction with the same timing, but differ in scale by the factor $\theta > 1$; that is, $b_{jt} = \theta b_{kt}$ for $t = 1, \dots, T$. In this case, we would have $f_j < \theta f_k$ due to economies of scale associated with larger construction.

In the study model, we assume the company is relatively small and not an industry leader. According to the Stackelberg-Nash-Cournot theory, this implies that the market price for the product is independent of the company's actions. In other words, the company cannot set the market price nor can it affect the price in any way through its production strategies. Moreover, the sales of the company in any period are constrained by the maximum market share it can expect to capture. Under these assumptions, it is appropriate to define an exogenous revenue function for each period in our planning horizon. Specifically, revenues in period t from selling d_t units of the product are given by $f_t(d_t)$, where f_t is a concave function. Numerical realizations of these functions should capture the market phenomenon of declining prices over time. One such function is given in section 4.

With this background, we can give a mathematical statement of our study model

$$\max_{\alpha} \sum_{t=1}^T \alpha^{t-1} \left\{ f_t(d_t) - [C(x_1 + \dots + x_t) - C(x_1 + \dots + x_{t-1})] - h_t i_t \right\} + \alpha^T v_T h_T - \sum_{k=1}^K I_k \rho_k \quad (1.a)$$

$$\text{s.t.} \quad \left. \begin{aligned} x_t - \sum_{k=1}^K \sum_{\tau=1}^t b_{k\tau} \rho_k &\leq b_0 \\ i_t - i_{t-1} - x_t + d_t &= 0 \end{aligned} \right\} \quad t = 1, \dots, T \quad (1.b)$$

$$(1.c)$$

$$\sum_{k=1}^K \rho_k \leq 1 \quad (1.d)$$

$$x_t \geq 0, i_t \geq 0, d_t \geq 0, \rho_k = 0 \text{ or } 1$$

$$i_0 = 0, x_0 = 0 \quad (1.e)$$

The constraints (1.b) link the production decisions to the capacity expansion decisions. The constraints (1.c) are inventory balance equations. The multiple choice constraint (1.d) states that at most one of the K capacity expansion plans may be selected.

Problem (1) as stated is nonlinear and nonconvex. It is difficult to optimize primarily due to the concave production cost functions to be minimized. In particular, if we let

$$G(x_1, \dots, x_T) = \text{total discounted cost of the production vector } (x_1, \dots, x_T)$$

we see from (1) that

$$G(x_1, \dots, x_T) = (1-\alpha) \sum_{t=1}^{T-1} \alpha^{t-1} C(x_1 + \dots + x_t) + \alpha^{T-1} C(x_1 + \dots + x_T). \quad (2)$$

It follows immediately that G is a concave function since it is the positive sum of the concave functions C(•).

The special form of G, however, given by (2) permits us to approximate it to an arbitrary degree of accuracy using MIP modeling techniques. This transformation is given in the following section.

Before proceeding to the MIP formulation of problem (1), we should like to discuss briefly two related properties of the production cost function G. First, we note that

$$\begin{aligned} \frac{\partial G(x_1, \dots, x_T)}{\partial x_t} &= \alpha^{t-1} \frac{dC(x_1 + \dots + x_t)}{dy} \\ &+ \alpha^t \left(\frac{dC(x_1 + \dots + x_{t+1})}{dy} - \frac{dC(x_1 + \dots + x_t)}{dy} \right) \\ &+ \alpha^{T-1} \left(\frac{dC(x_1 + \dots + x_T)}{dy} - \frac{dC(x_1 + \dots + x_{T-1})}{dy} \right) \end{aligned}$$

The first term in this expression is the immediate marginal production cost in period t that results from producing x_t units. The remainder of the terms are all non-positive and represent discounted rents paid back to production in period t by production in later periods which profit from the learning that took place in period t .

Additional interpretation is possible if we assume that the investment variables ρ_k in problem (1) have been exogenously set. Problem (1) then becomes a continuous optimization problem, allowing us to apply the Kuhn-Tucker necessary optimality conditions. To this end, consider an optimal solution for (1) with the fixed investment decision, denoted by $\tilde{x}_t, \tilde{d}_t, \tilde{i}_t$. Consider further a period t in which both sales \tilde{d}_t and production \tilde{x}_t are positive. In this case, the necessary optimality conditions imply the existence of $\tilde{\pi}_t$, which measures the marginal value of another unit of production capacity in period t , that satisfies

$$\tilde{\pi}_t = \alpha^{t-1} \left[\frac{df(\tilde{d}_t)}{d\tilde{d}_t} - \frac{\partial G(\tilde{x}_1, \dots, \tilde{x}_T)}{\partial x_t} \right]$$

This expression can be compared to a similar expression for measuring the marginal value of another unit of production capacity in period t when there are no learning effects; namely,

$$\alpha^{t-1} \left[\frac{df(\tilde{d}_t)}{d\tilde{d}_t} - \frac{dC(\tilde{x}_1 + \dots + \tilde{x}_t)}{dy} \right]$$

As we have noted above, this last expression will be less than $\tilde{\pi}_t$ by exactly the rents paid by future periods to period t from the learning caused by \tilde{x}_t .

3. Mixed Integer Programming Formulation of the Study Model

The MIP formulation of the study model (1) is achieved by adapting classical MIP modeling techniques. Suppose we approximate the concave cumulative cost function shown in Figure 1 by a piecewise linear function as shown in Figure 2. In general, we will have R linear intervals with c_r equal to the slope in interval r , and M_r equal to the length of interval r . Let x_{rt} denote the decision variable corresponding to production in interval r in period t . For our problem, the x_{rt} must obey the logical condition

$$x_{rt} > 0 \Rightarrow \sum_{\tau=1}^t x_{r-1, \tau} = M_{r-1}$$

In words, this condition says that x_{rt} can be positive, with its lower marginal production cost c_r , only if, in the interval immediately to the left of interval r , cumulative production over the first t periods has reached its upper bound. It is easy to see that this condition implies x_{rt} can be positive only if cumulative production over the first t periods equals at least $M_1 + \dots + M_{r-1}$, which is the lower bound on the r^{th} interval. The logical condition can be mathematically expressed by the inequalities

$$\left. \begin{array}{l} x_{rt} - \Delta_{rt} M_r \leq 0 \\ \Delta_{rt} M_{r-1} - \sum_{\tau=1}^t x_{r-1, \tau} \leq 0 \end{array} \right\} \begin{array}{l} \text{for } r=2, \dots, R; \\ t=1, \dots, T \end{array} \quad (3)$$

where the Δ_{rt} are zero-one variables.

Finally, we approximate the concave revenue function $f_t(d_t)$ by a piecewise linear function. In particular, let f_j , $j=1, \dots, J$ denote the slopes in successive intervals of length D_j ; since $f_t(\cdot)$ is assumed to be concave, we have $f_j > f_{j+1}$ for all j . Let d_{jt} denote the decision variable corresponding to sales in the j^{th} interval in period t .

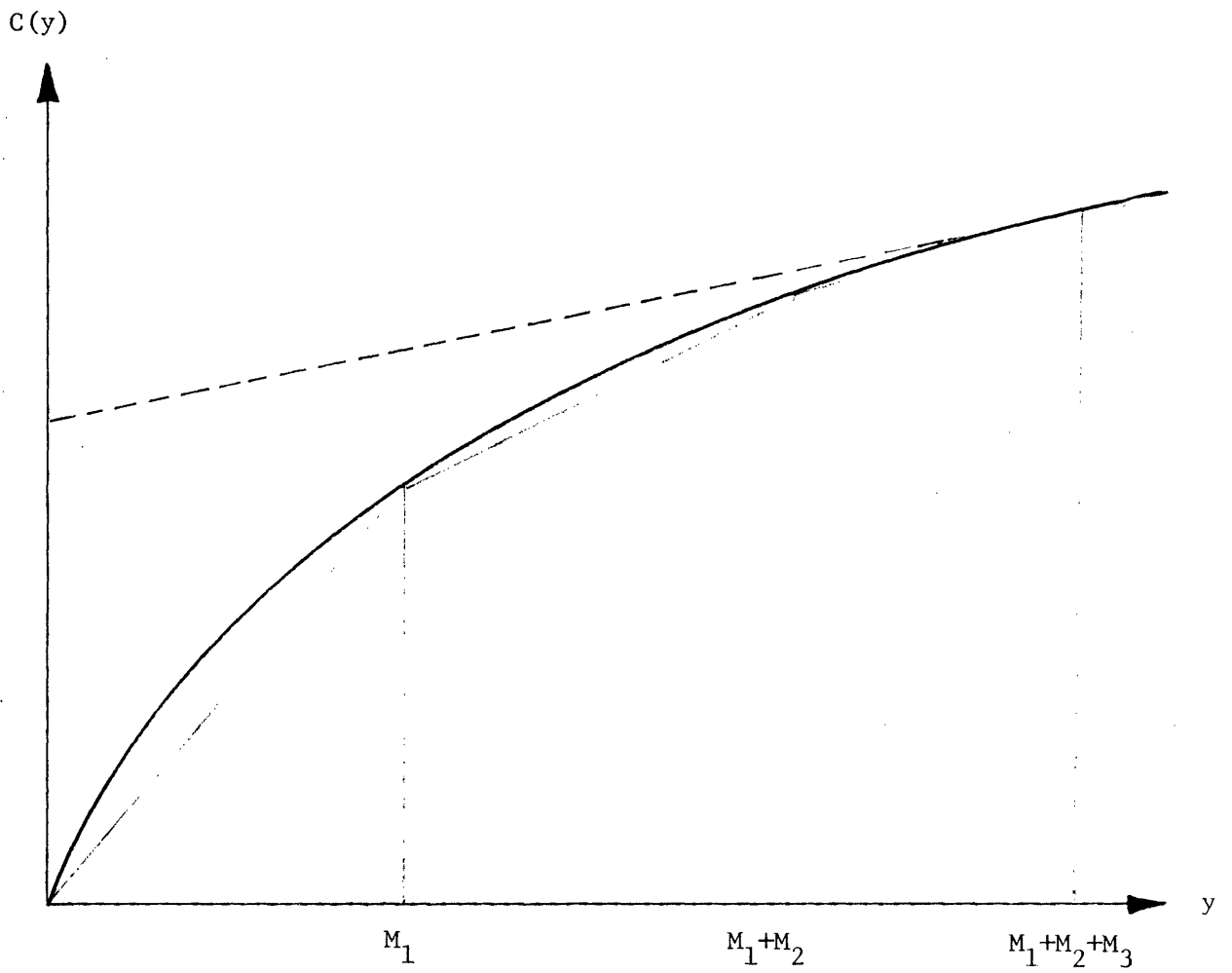


Figure 2

Substituting

$$d_t = d_{1t} + d_{2t} + \dots + d_{Jt}$$

$$x_t = x_{1t} + x_{2t} + \dots + x_{Rt}$$

and the corresponding objective function approximations in the study model (1), plus the logical conditions (3), we obtain

$$\begin{aligned} \max \quad & \sum_{t=1}^T \alpha^{t-1} \left\{ \sum_{j=1}^J f_{jt} d_{jt} - \sum_{r=1}^R c_r x_{rt} - h_t i_t \right\} \\ & + \alpha^T v_T h_T - \sum_{k=1}^K I_k \rho_k \\ \text{s.t.} \quad & \left. \begin{aligned} \sum_{r=1}^R x_{rt} - \sum_{k=1}^K \sum_{\tau=1}^t b_{k\tau} \rho_k \leq b_0 \\ i_t - i_{t-1} - \sum_{r=1}^R x_{rt} + \sum_{j=1}^J d_{jt} = 0 \end{aligned} \right\} t=1, \dots, T \\ & \left. \begin{aligned} x_{rt} - \Delta_{rt} M_r \leq 0 \\ \Delta_{rt} M_{r-1} - \sum_{\tau=1}^t x_{r-1, \tau} \leq 0 \end{aligned} \right\} \begin{array}{l} \text{for } r=2, \dots, R \\ t=1, \dots, T \end{array} \\ & \sum_{k=1}^K \rho_k \leq 1 \end{aligned} \tag{4}$$

$$x_{rt} \geq 0, D_j \geq d_{jt} \geq 0, i_t \geq 0, \rho_k = 0 \text{ or } 1, \Delta_{rt} = 0 \text{ or } 1$$

$$i_0 = 0, x_0 = 0$$

Problem (4) is the MIP model approximating the original study model (1). Although we omit any details here, further analysis could be performed to bound apriori any errors resulting from the objective function approximation. Of course, the closer we make the approximation of the cumulative production bounds cost function, the greater the number of integer variables we will have in problem (4).

We observe in passing that there are many valid inequalities involving the integer variables that could be added to problem (4).

For example

$$\Delta_{r,t} - \Delta_{r,t+1} \leq 0$$

is a valid inequality for any r and t since $\Delta_{r,t} = 1$ implies saturation of the $(r-1)$ st interval by period t which in turn implies saturation of the same interval by period $t+1$. Similarly,

$$\Delta_{r,t} - \Delta_{q,t} \leq 0$$

is a valid inequality for $q < r$.

Stronger inequalities than these can be written by relating the zero-one variables determining capacity expansion to those controlling learning.

Theorem 1: For all t and $r=2, \dots, R$,

$$\Delta_{r,t} (M_1 + \dots + M_{r-1}) - \sum_{k=1}^K \left(\sum_{\tau=1}^t (t-\tau+1) b_{k\tau} \right) \rho_k \leq t b_0 \quad (5)$$

is a valid inequality for problem (4).

Proof: From (4) we have for any s and any $r=2, \dots, R$

$$\sum_{r=1}^R x_{rs} \leq b_0 + \sum_{k=1}^K \sum_{\tau=1}^s b_{k\tau} \rho_k$$

Summing both sides from $s=1$ to $s=t$, we obtain

$$\sum_{s=1}^t \sum_{r=1}^R x_{rs} \leq tb_0 + \sum_{k=1}^K \sum_{\tau=1}^t (t-\tau+1) b_{k\tau} \rho_k \quad (6)$$

where we have rearranged the triple sum on the right to collect similar terms. The sum on the left in (6) equals cumulative production through period t .

Suppose the sum on the right in (6) is less than $M_1 + \dots + M_{r-1}$. This implies that cumulative production through period t must be insufficient to allow production in the r th interval of the discretized learning curve. From (5), we have $\Delta_{rt} = 0$, a valid condition in this case. On the other hand, if the sum on the right in (6) is greater than or equal to $M_1 + \dots + M_{r-1}$, the constraint (5) is not binding and Δ_{rt} may equal 1, which from a production capacity viewpoint is correct. The variable Δ_{rt} may still be constrained to zero due to other constraints in problem (4).

4. Numerical Example of the Study Model

In the numerical example, a model is implemented and optimized to determine the production strategy for a single product produced by a company over a five year planning horizon. The objective is to maximize the net present value of sales less production and capacity expansion costs. The industry producing this product is characterized by a learning curve

with moderate learning effects. Thus, as we discussed in section 2, early entrants to the industry have a competitive advantage over later entrants.

In particular, the learning curve for the company's product is given by

$$C(y) = m_0 y + c_0 (1 - e^{-\lambda y})$$

where y is cumulative production and $C(y)$ is the cumulative production costs. Here we take $m_0 = 800$, $c_0 = (2400)(23000)$ and $\lambda = 1/23000$, with an upper bound on y over the five year horizon equal to 50,000. The production costs therefore will range between 3200 dollars per unit and 1073 dollars per unit during this period.

An interesting observation by Spence [10] is that moderate learning presents the greatest planning dilemma. If λ is large, learning occurs rapidly and the learning effect approximates a fixed cost to entry equal to c_0 . The company must then decide whether or not to incur this fixed cost, a choice that should be relatively easy to make. On the other hand, if λ is small, learning effects are not pronounced, and the company can proceed as if there is a constant marginal cost approximately equal to $m_0 + \lambda c_0$.

The model tracks the average annual inventory of the company for which holding costs are convex piecewise linear functions of average inventory. Convex functions are used to model the increasing burden on the firm of holding large amounts of inventory. The inventory carrying costs are 350 dollars per unit for the first 500 units and 45 dollars per unit for the next 2,500 units.

The capacity expansion decisions are modeled by five capacity expansion scenarios as shown in Table 1. Table 2 gives the capacity expansion costs used to calculate the discounted costs in Table 1. The company can initially build plants of capacity 20 or 40 thousand machine hours per year at either

CAPACITY EXPANSION OPTION

	1	2	3	4	5
1	*(BI):20	(BI):40	(BI):40	-	(BI):20
2	(BII):20	-	-	(BI):20	-
YEAR 3	(EI):10	-	(BII):20	-	-
4	(EII):10	(BII):40	(EII):20	-	(BII):20
5	(EI):10	-	-	-	-
TOTAL DISCOUNTED COST	939,793	846,562	937,172	322,207	488,413

TABLE 1. Capacity expansion options (in thousand machine hours per year)

*Note: BI = build at site I, EII = expand at site II.

<u>BUILD/EXP.</u>	<u>SITE</u>	<u>SIZE</u>	<u>COST</u>
Build	I	20	417.8
Build	I	40	612.8
Build	II	20	487.4
Build	II	40	668.5
Expand	I	10	139.3
Expand	I	20	236.8
Expand	II	10	146.2
Expand	II	20	250.7

TABLE 2. Capacity expansion costs (1,000's \$)

site. Furthermore, if the firm opts to build a smaller plant initially, it can expand the plant at a later time. However, the maximum total capacity of 40 thousand machine hours per year cannot be exceeded at either site. The capacity expansion costs shown in Table 1 reflect the economies of scale normally encountered in highly volatile and rapidly expanding industries. The model selects the best capacity expansion option, identifying the tradeoffs between the economies of scale and the cost of maintaining excess capacity.

We do not constrain the cash flow of the company. Therefore, it is feasible for the firm to take an early loss in overcoming the barrier to entry. The company may be seen as an established firm attempting to break into a new high-growth industry. It must decide how to deal simultaneously with the learning curve effects, which serve to lower production costs, and the increase in competition which serves to lower the market price.

The market prices in our example are determined from the equation

$$P(t,d) = a_0 e^{.2t} d^{-2/3}$$

where t is time and d is sales. The exponential term is derived from the assumption that the market is growing at a rate of 20% per year. In any given time period τ , the function $P(\tau, \cdot)$ is monotonically decreasing. The interpretation of this function is more natural if one examines the inverse function

$$d(P, \tau) = K(\tau) p^{-3/2}$$

This function represents the market the firm can capture in period τ if it charges a price p . Note that the form of the equation is identical to that for a constant elasticity demand curve. In our numerical example, we set $a_0 = 900$.

The results of optimizing the example are shown in Table 3. The selected capacity expansion option takes full advantage of economies of scale by building the largest plants possible at the two sites in years 1 and 4. This capacity expansion decision resulted in lost sales and a corresponding drop in market share penetration in year 3.

Capacity utilization is maintained at 100% over the 5 year horizon by allowing inventory to build up during the years when production capacity exceeded demand. The inventory is subsequently drawn down when the demand exceeds production capacity. The costs associated with these fluctuations in average inventory are justified in part by the rapid growth of the market and in part by the relatively high cost of production and capacity expansion compared to the cost of holding inventory. If the economies of scale on the construction of new production facilities had been less pronounced, we would expect production to follow demand and inventories would go to zero.

The negative cashflow in the first year is the result of two factors. First, the production costs are high since the company is low on the learning curve. As more units are produced and the variable production costs decline, each unit becomes more profitable, at which point the company starts to make money. The second cause of the initial capital requirements is the capacity expansion expense. Such expenses also account for the relatively small net cashflow in year 4 compared to that of year 5. These periodic lumpy investments are found in all industries and are a source of significant barriers to entry for small undercapitalized firms.

Period	1	2	3	4	5
Production	5,500	5,500	5,500	10,500	10,500
Sales	4,000	7,000	5,500	10,000	11,000
Inventory	1,500	0	0	500	0
Capacity (Machine hours)	44,000	44,000	44,000	84,000	84,000
Cashflow	-492,030	1,524,760	668,700	716,000	1,646,770
Change in Market share		+ 43%	- 36%	+ 51%	- 8%

TABLE 3: Numerical Example Results

5. Directions of Future Research

The study model that was proposed and implemented for this paper illustrates how mixed integer programming can be applied to capacity expansion planning when there are learning effects. The model and the modeling concepts can be easily extended to capture a wider range of planning phenomena that we would expect to encounter in practice. For example, planning problems where there are products with multiple components, each with its own learning curve, can be readily described by the types of constraints we used in the study model. The issue for these problems is the selection of balanced expansion plans among the components.

Another desirable model extension would be to more explicitly describe the allowable ranges of sizes of new facilities, and their associated returns to scale. We chose to model these options implicitly in the study model in order to simplify it. Capital financing and cash flow constraints could also be added to the study model with no conceptual difficulty.

A model extension that is less obvious and requires further investigation is one permitting a closer reconciliation between discounting and learning, especially when the two occur at commensurate rates. Since both occur continuously over the relatively long time periods in our models, each period's production costs would be more appropriately calculated by integration, rather than taking differences in cumulative cost functions. The manner in which this can be done will be given in a subsequent paper.

Explicit treatment of uncertainty is an important extension of the study model. In particular, we envision the application of stochastic programming with recourse models to these capacity expansion problems.

Benders decomposition method can be useful in dealing with dimensionality difficulties arising from multiple recourse scenarios (see [8]). The method is also well known for its application to mixed integer programming. In fact, these two uses of Benders' method have been integrated and applied successfully to related capacity expansion models (see [1]). For capacity expansion with learning, we anticipate that significant experimentation with model representation and specific algorithmic procedures might be required to stabilize the method. Inequalities such as (5) added to the Benders' master problems should be useful in providing this stability.

Finally, an important area of future research is the explicit modeling and analysis of competitors behavior for manufacturing and marketing new products with learning effects. Here we envision using the study model and its extensions to describe each company in the competitive market. Assuming that there is a single company which leads the market, and the other companies follow, recent research indicates that the individual models can be integrated into a market model (see[9]).

References

1. Bienstock, D. and Shapiro, J.F., "Stochastic Programming Models for Capacity Expansion Planning: An Application to Electric Utilities," (in preparation).
2. Boston Consulting Group, "Perspectives on Experience." Boston, 1972.
3. Ebert, R.J., "Aggregate Planning with Learning Curve Productivity," Management Sci., Vol. 23, No.2 (October, 1976), pp 171-182.
4. Hax, A.C. and Majluf, N.S., "Competitive Cost Dynamics: The Experience Curve," Interfaces, Vol. 12, No. 5 (October 1982).
5. Porter, M.E., Competitive Strategy, New York: The Free Press, 1980.
6. Rao, R. and Rutenberg, D., "Preempting an Alert Rival: Strategic Timing of the First Plant by Analysis of Sophisticated Rivalry," Bell J. of Economics, Vol. 10, No. 2, Autumn, 1979.
7. Reeves, G.R. and Sweigart, J.R., "Multiperiod Resource Allocation with Variable Technology," Management Sci., Vol. 28, No. 12, December 1982, pp 1441-1449.
8. Shapiro, J.F., Mathematical Programming: Structures and Algorithms, John Wiley and Sons, 1979.
9. Sherali, H.D., Soyster, A.L., and Murphy, F.H., "Stackelberg-Nash-Cournot Equilibria: Characterizations and Computations," Operations Research, Vol. 31, No. 2, March-April 1983, pp 253-276.
10. Spence, A.M., "Entry, Capacity, Investment and Oligopolistic Pricing," Bell J. of Economics, Vol. 8, No. 2, Autumn, 1977.
11. Spence, A.M., "Investment Strategy and Growth in a New Market," Bell J. of Economics, Vol. 10, No. 1, Spring 1979.

12. Spence, A.M., "The Learning Curve and Competition," Bell J. of Economics, Vol. 12, No. 1, Spring 1981.
13. Yelle, L.E., "The Learning Curve: Historical Review and Comprehensive Survey," Decision Sci., Vol. 10, No. 2, April 1979, pp 302-328.