CAPACITY EXPANSION AND PROBABILISTIC GROWTH

BY ALAN S. MANNE

This paper is concerned with the optimal degree of excess capacity to be built into a new facility such as a pipeline, a steel plant, or a superhighway. The problem is complicated by: (1) the presence of substantial economies of scale in plant construction; (2) the penalties involved in accumulating backlogs of unsatisfied demand; and (3) the use of a random-walk pattern rather than a deterministic upward trend in demand.

1. INTRODUCTION

This study stems from an optimizing model originally suggested by Hollis Chenery for predicting investment behavior [4]. Like Chenery's paper, this one is concerned with the interplay between economies of scale and an anticipated persistent growth in demand for capacity. The generalizations discussed here are of two types: (a) the use of probabilities in place of a constant rate of growth in demand; and (b) a study of the economics and the penalties involved in accumulating backlogs of unsatisfied demand. The possibility of accumulating such backlogs raises considerable doubt with respect to Chenery's "excess capacity hypothesis."

Surprisingly enough, generalization (b) leads to greater difficulties in analysis than (a). The use of probabilities to describe the growth process does little—if anything—to complicate matters. A probabilistic version of Chenery's model turns out to be closely related to the classical problem of gambler's ruin, and a powerful tool can be borrowed from that area—the Laplace transform for the duration of the game. Thanks to this transform, the zero-backlog probabilistic model becomes no more difficult to study than the corresponding deterministic one. A direct implication is that a probabilistic growth course makes it necessary to incur higher expected costs, and also makes it desirable to install plant capacity of a somewhat larger size than would be optimal if demand were growing at a steady rate equal to the expected value of the probabilistic increments. Uncertainty, in this sense, has a stimulating effect upon the magnitude of individual investments.

Going beyond Chenery's model to include the possibility of backlogs, it

---

1 The author is indebted for suggestions made by Martin Beckmann, Gerard Debreu, Tjalling Koopmans, Jacob Marschak, Richard Rosett, and Herbert Scarf of the Cowles Foundation; by Gerd Reuter of the University of Manchester; by Ashley Wright of Standard Oil Co. (New Jersey); and by Robert Donovan, Donald MacArthur, and Gifford Symonds—all of Esso Standard Oil Co.

An early version of this paper was presented at the American summer meeting of the Econometric Society in Cambridge, Massachusetts, August 26-28, 1958.

Research undertaken by the Cowles Foundation for Research in Economics under Task NR 047-006 with the Office of Naval Research.
turns out that there is a curious ambiguity in the effects of an increase in the variance of demand. Once the possibility of backlogs is admitted, an increase in variance can even lead to a decrease in the optimal level of costs.

2. THE DETERMINISTIC MODEL—NO BACKLOGS IN DEMAND

In order to provide a reference point for discussion of the more difficult cases, Chenery's deterministic model itself will first be reviewed. Following this will come the modifications involving (a) probabilistic growth and (b) the possibility of accumulating backlogs in demand. Chenery's model grew out of his studies of the natural gas transmission industry, a sector characterized by rapid growth and by substantial economies of scale in pipeline construction and operation. Much the same situation seems to prevail in the case of oil pipelines [7], the telephone industry [6], highway construction, electric power generation, petroleum refining, and chemicals processing [5].

Figure 1 charts the course of demand and of capacity over time under the

![Diagram](image)

**Figure 1.**—Growth of demand and of capacity over time.
following simplifying assumptions: (1) that demand grows linearly over time; (2) that the equipment has an infinite economic life; and (3) that whenever demand catches up with the existing capacity, $x$ units of new capacity are installed.\(^2\) (The demand at $t_0$ is denoted by $D_0$.) Unlike Chenery, we shall assume that the planning horizon is infinite, rather than being truncated after an arbitrary finite number of years. Excess capacity, when plotted on Figure 2, then displays a sawtooth pattern typical of the closely related

![Figure 2.—Evolution of excess capacity over time.](image)

Wilson-type inventory model [1, pp. 252-255]. If, for convenience, the physical unit of capacity and of demand is set equal to one year's growth in demand this sawtooth cycle repeats itself every $x$ years.

The installation costs that result from a single capacity increment of size $x$ are assumed to be given by a cost relationship in the form of a power function:

\[
(2.1) \quad kx^a \quad (k > 0; \quad 0 < a < 1).
\]

If, for example, $a = 1/2$, this cost function says that a pipeline capable of handling 16 years' worth of growth in demand is only twice as expensive as one that can accommodate four years' worth. The existence of such substantial economies of scale implies the desirability of building new capacity consider-

\(^2\) Chenery [4] and Cookenboo [7] both point out that the concept of installed capacity is a slippery one, even when dealing with such a homogeneous facility as a gas or an oil pipeline. Once a line of given diameter has been laid, new pumping equipment can be added—enough to raise the ultimate installed capabilities to a level of perhaps two or three times the initial amount. From the viewpoint of our model, it seems best to regard the decision variable $x$ as a measure of the ultimate rather than the immediate amount of pumping capacity installed. In defense of this shortcut, it should be noted that on an optimum-diameter line for constant throughput, all pumping station equipment, according to Cookenboo's figures, generally comprises no more than 10 per cent of the total initial pipeline costs [7, pp. 65, 82, 106].
ably in advance of demand. But how much in advance? Here the discounting of future costs becomes crucial.

Without discounting, it would be perfectly sensible to spend a dollar now in order to save a dollar's worth of costs either next year or ten years from now, or 100 years hence. There is no limit to the size of line which it pays to build. With discounting, on the other hand, this paradox can be sidestepped. Throughout, we shall adopt the expression $e^{-rt}$ as the present value of a dollar due $t$ years in the future ($r > 0$). The quantity $r$ will be referred to as the "discount rate."

As a time origin for subsequent calculations, it will be convenient to take any such point as $t_0$ or $t_0 + x$ or $t_0 + 2x$ on Figure 2, a time at which the previously existing excess capacity has just been wiped out. Such a point will be known hereafter as a "point of regeneration." (For a similar application of the idea of regeneration points, see Karlin [9, pp. 280–85].) Note that when we have reached $t_0 + x$, the future looks identical with the way it appeared $x$ units of time previously. Then if we say that $C(x)$ is a function of $x$ that represents the sum of all discounted future costs looking forward from a point of regeneration, we may write down the following recursive equation:

$$
C(x) = kx^a + e^{-rx}C(x).
$$

The first term on the right hand side indicates the installation costs incurred directly at the beginning of the current cycle. (See equation (2.1).) The second term measures the sum of all installation costs incurred in subsequent cycles, and discounts these from the next point of regeneration back to the present one, a difference of $x$ years. From (2.2), it follows directly that:

$$
\frac{C(x)}{k^x} = \frac{x^a}{1 - e^{-rx}}.
$$

Differentiating $\log C(x)$ with respect to $x$, and setting the result equal to zero:

$$
\frac{d \log C(x)}{dx} = \frac{a}{x} - \frac{re^{-rx}}{1 - e^{-rx}} = 0
$$

or

$$
(2.4)
\frac{r\dot{e}}{e^{\ddot{r}x}} = 1
$$

where $\dot{e}$ denotes the optimal size of installation.

---

\* Gifford Symonds has suggested an additional reason for the discounting of future costs: the expectation of continuing progress in pipeline technology. If the general price level remains constant, it is reasonable to suppose that in, say 10 years' time the cost of building a line with a capacity of $x$ units will be significantly cheaper than the cost of such a line today. The proviso about constancy of the general price level is important. If one is a believer in the inevitability of creeping inflation, the one factor would tend to cancel out the other.
The reader can verify for himself that (2.4) is not only a necessary condition, but also a sufficient one to ensure a unique minimum-cost solution. With this equation, the optimal capacity increment $\Delta$ may be determined for any combination of the two parameters $a$ and $r$—a crossplot being provided in Figure 3.

![Figure 3](image)

This deterministic model lends itself readily to sensitivity testing. To find out how the optimal level $\Delta$ is affected by changes in $r$, one need only observe that (2.4) is written as a function of the product $r\Delta$, and that therefore, for a constant value of $a$:

$$r\Delta + \Delta dr = 0$$
or

\[
\frac{d\hat{a}}{dr} = \frac{-\hat{a}}{r} < 0.
\]

The derivative \(d\hat{a}/dr\) is clearly negative for positive values of \(\hat{a}\) and of \(r\). The higher the discount rate (i.e., the higher the cost of capital), the smaller will become the optimal size of each installation.

Both to an economist and to an operations researcher, it is likely that the general shape of the cost function \(C(x)/k\) will be of even greater interest than the optimal value \(\hat{a}\) itself. Figure 4 contains a plot for a fairly typical set of

![Figure 4](image)

parameter values: \(a = .50\) and \(r = .15\). (Incidentally, these numerical values of \(a\) and \(r\) will be the ones employed in all subsequent illustrations.) The optimal point indicated by this figure leads to a cost of 4.046 at \(x = 8.4\) year's worth of demand growth. The figure also gives an indication of how little these costs change within a fairly wide range of values assigned to the decision variable \(x\).

An \(x\)-value as high as 11.0 or as low as 6.0 will increase costs by less than 2 per cent. From the viewpoint of the operations researcher and the business forecaster, this insensitivity is fortunate indeed. Even a substantial error in
forecasting will not lead to an egregiously bad choice for the capacity increment.

What is fortunate from the viewpoint of the business executive may be disastrous, however, from the viewpoint of an economist trying to forecast investment choices on the basis of an optimizing model. Even if the economist happens to hit upon the identical values for \( a \) and \( r \) that are in the mind of the executive, the latter will suffer no great penalty for deviating from the optimal path predicted by the economist for his behavior.

3. The Probabilistic Model—No Backlogs

With this background, we are in a position to discuss a case of probabilistic growth and minimization of expected costs, still ruling out the possibility of deliberate backlogs in demand. This model represents, of course, just one of many possibilities for describing a growth process in probabilistic terms. The particular structure is one that has been postulated not only for analytical convenience, but also because of its close relationship to the normal distribution.

The model employed here is the Bachelier-Wiener diffusion process in continuous time. Except for notation, our exposition is virtually identical with Feller’s [8, pp. 323–27]. Feller begins by considering the case in which a discontinuous random change in demand occurs every \( \Delta t \) units of time, and then examines the limiting form of this process. With probability \( p \), the discrete change constitutes an increase of \( \Delta D \) units, and with probability \( q = (1-p) \), a decrease of \( \Delta D \) units. In Markov process terms:

\[
D(t) = D(t - \Delta t) + \varepsilon(t)
\]

where \( D(t) \) represents the demand at time \( t \), and where \( \varepsilon(t) \) is a random variable taking on the values of \( + \Delta D \) and \( -\Delta D \), with respective probabilities \( p \) and \( q \).\(^4\) (It is assumed that each of the \( \varepsilon(t) \) increments is distributed identically and independently.) With a change in demand occurring every \( \Delta t \) units of time, this means that over a fixed period of, say, \( t \) years in length, approximately \( t/\Delta t \) changes will have occurred. Quoting Feller directly now, with minor changes in notation:

\(^4\) An immediate objection arises to this formulation. According to (3.1), it is perfectly possible for the absolute rate of demand \( D(t) \) to become negative—an economic and a physical absurdity unless we suppose that customers are permitted to return products previously purchased. In order to avoid this absurdity, Michael Lovell has suggested defining a truncated demand path, \( D^*(t) \) such that \( D^*(t) = \max\{D(t), 0\} \). Any construction and backlog policy that is optimal for the time series \( D(t) \) must also be optimal for \( D^*(t) \). Why? Because whenever \( D(t) < 0 \), the firm will incur neither construction nor backlog costs, the only two cost elements considered in this entire analysis. Hence, although our optimal construction and backlog policy is based upon the untruncated path \( D(t) \), it will also be an optimal policy for the more realistic case of the truncated path, \( D^*(t) \).
Only multiples of $AD$ and $\Delta t$ represent meaningful coordinates, but in the limit $\Delta D \to 0$, $\Delta t \to 0$; every displacement and all times become possible.

We must not expect sensible results if $\Delta D$ and $\Delta t$ approach zero in an arbitrary manner. . . . Physically speaking, we must keep the $D$- and $t$- scales in an appropriate ratio or the process will degenerate in the limit, the variances tending to zero or infinity. To find the proper ratio note that the total displacement during time $t$ is the sum of about $t/\Delta t$ mutually independent random variables each having the mean $(p - q)\Delta D$ and variance $4pq(\Delta D)^2$. The mean and variance of the total displacement in time $t$ are therefore about $t(p - q)\Delta D/\Delta t$ and $4pq(t\Delta D)^2/\Delta t$, respectively. To obtain reasonable results we must let $\Delta D$ and $\Delta t$ approach zero in such a way that they remain finite for all $t$. The finiteness of the variance requires that $(\Delta D)^2/\Delta t$ should remain bounded; the finiteness of the mean implies that $(p - q)$ must be of the order of magnitude of $\Delta D$. This suggests putting

$$
\frac{(\Delta D)^2}{\Delta t} = \sigma^2, \quad \rho = \frac{1}{2} + \frac{\mu\Delta D}{2\sigma^2}, \quad \sigma = \frac{1}{2} - \frac{\mu\Delta D}{2\sigma^2}.
$$

. . . We use the norm [3.2] to pass to the limit $\Delta D \to 0$, $\Delta t \to 0$. The total displacement at time $t \approx \Delta t$ is determined by $n$ Bernoulli trials, and therefore the limiting form of $V_n$ [the probability that after $n$ trials, the demand will have grown by a total of exactly $D$ units] is given by the normal distribution, $\nu(D; t)$. For a fixed $\Delta D$ the displacement is the sum of finitely many independent variables, and its mean is $t(p - q)\Delta D/\Delta t = \mu t$; its variance $4pq(t\Delta D)^2/\Delta t = \sigma^2t$ [8, pp. 324-5].

To sum up: The parameters $\rho$, $\sigma$, $\Delta D$, and $\Delta t$ enable us to study a discrete stochastic process in which the total growth $D$ over a fixed period of $t$ years is a random variable $D(t)$. Furthermore, in the limit, for the case of continuous time and a continuous growth path, this process describes the demand increment as a random variable which is normally distributed with mean $\mu t$ and with a variance of $\sigma^2 t$.

Now in order to make use of this process for the capacity optimization problem, it is going to be necessary to work with a certain probability density function $u(t; x)\Delta t$: the probability with which $t$ time units elapse before the point at which demand first exceeds the initial level by $x$ units. In other words, $u(t; x)\Delta t$ represents the probability with which $t$ units of time elapse between one installation of capacity and the next one. In gambler’s ruin terminology, this is the probability with which exactly $t/\Delta t$ “trials” are needed in order for a gambler to go broke—a gambler whose initial capital is $x$, and who is playing against an adversary with infinite wealth. At each stage of such a game, the gambler would lose one unit with a probability of $\rho$, and gain one unit with a probability of $q = 1 - \rho$ [8, pp. 311-21]. The following relationship may therefore be written:

$$
(u(t + \Delta t; x) = \frac{\rho}{1 + \sigma} u(t; x - \Delta x) + \frac{q}{1 - \sigma} u(t; x + \Delta x) \quad (x > 0; 0 \leqslant t \leqslant \infty).
$$

Equation (3.3) says that whatever be the probability of ruin in exactly
\((t + \Delta t)/\Delta t\) steps for a gambler with an initial capital of \(x\), this quantity must equal the weighted sum of the ruin time probabilities in just \(t/\Delta t\) steps for a gambler with a capital of \((x - \Delta x)\) and \((x + \Delta x)\), the respective weights being the transition probabilities \(p\) and \(q\).

Expanding according to Taylor's theorem up to terms of second order:

\[
\Delta t \frac{\delta u(t; x)}{\delta t} = (q - p) \Delta x \frac{\delta u(t; x)}{\delta x} + \frac{(\Delta x)^2}{2} \frac{\delta^2 u(t; x)}{\delta x^2}.
\]

Equating the random variable \(D(t)\) to the capacity increment \(x\), substituting from (3.2), and taking the limit:

\[
\frac{\delta u(t; x)}{\delta t} = -\mu \frac{\delta u(t; x)}{\delta x} + \frac{\sigma^2}{2} \frac{\delta^2 u(t; x)}{\delta x^2}.
\]

The Laplace transform of \(u(t; x)\) will be indicated by \(\hat{u}(r; x)\), and in economic terms is defined as the discounted value of the probabilities \(u(t; x)dt\):

\[
\hat{u}(r; x) = \int_0^\infty u(t; x)e^{-rt}dt.
\]

Taking the Laplace transform of each side of (3.5), and recalling the boundary condition that \(u(0; x) = 0\), we obtain a second-order linear differential equation with respect to \(x\):

\[
r\hat{u}(r; x) = -\mu \frac{\delta \hat{u}(r; x)}{\delta x} + \frac{\sigma^2}{2} \frac{\delta^2 \hat{u}(r; x)}{\delta x^2}.
\]

The characteristic equation has two real roots:

\[
\lambda_1 = \frac{\mu}{\sigma^2} \left[ 1 + \sqrt{1 + \frac{2\sigma^2}{\mu^2}} \right],
\]

\[
\lambda_2 = \frac{\mu}{\sigma^2} \left[ 1 - \sqrt{1 + \frac{2\sigma^2}{\mu^2}} \right].
\]

The general solution for the Laplace transform is consequently of the form

\[
\hat{u}(r; x) = A(r)e^{\lambda_1 x} + B(r)e^{\lambda_2 x}
\]

where \(A(r)\) and \(B(r)\) are constants whose values depend upon \(r\) and also upon the boundary conditions for \(\hat{u}(r; x)\). These boundary conditions are twofold: first, that \(\hat{u}(r; x)\) lie between zero and unity, and second, that \(\hat{u}(r; 0) = 1\). Since \(\lambda_1 > 0\), and since \(\lambda_2 < 0\), the bounds upon \(\hat{u}(r; x)\) can be ensured only by setting the constant \(A(r)\) equal to zero. And to have \(\hat{u}(r; 0)\)

\footnote{For a rather different economic application of the Laplace transform, see Blyth [3].}
FRAME

When \( \sigma^2 = 0 \), we have the case of complete certainty, a steady annual increase in demand consisting of \( \mu \) units. Both the numerator and denominator of the expression for \( \lambda_2 \) vanish when \( \sigma^2 = 0 \). It is easy, however, to show that as \( \sigma^2 \) approaches zero, the expression for \( \lambda_2 \) approaches the value of \(-\sigma^2/\mu\).
variance of these changes is altered. First, it may be shown that \( d\lambda_2/d\sigma^2 > 0 \). If this be granted, then assertion (1) is proved directly: the greater the variance, the greater will be the level of expected discounted costs, regardless of the value of \( x \). (Note that \( \lambda_2 \) is a negative quantity, and that an increase in variance makes \( \lambda_2 \) less negative.)

In order to prove assertion (2), we return to the sensitivity analysis at the end of the preceding section. According to (2.5), the optimal size of installation increases as the discount rate \( r \) is lowered. In our probabilistic model, we have already shown that \( \lambda_2 \) may be viewed as nothing but an “adjusted” discount rate. Hence assertion (2): the greater the variance, the lower will be the absolute value of \( \lambda_2 \), and the higher will be the optimal value \( \delta \). This completes the proof.

### Table I

**Variance of Demand Versus Optimal Capacity Increments**

\( (x = .50, r = .15) \)

<table>
<thead>
<tr>
<th>( \sigma^2 )</th>
<th>0</th>
<th>1</th>
<th>4</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_2 = 1/\sigma^2 \left[ 1 - \sqrt{1 + 2\sigma^2} \right] )</td>
<td>-1.500</td>
<td>-1.402</td>
<td>-1.208</td>
<td>-0.9890</td>
</tr>
<tr>
<td>Optimal capacity increment = ( \delta )</td>
<td>8.4</td>
<td>9.0</td>
<td>10.4</td>
<td>14.3</td>
</tr>
<tr>
<td>Minimum expected discounted costs = ( C(\delta) = \delta^a )</td>
<td>4.046</td>
<td>4.185</td>
<td>4.508</td>
<td>5.282</td>
</tr>
<tr>
<td>( \frac{C(\delta)}{k} = \frac{\delta^a}{1 - \sigma^2 \delta \beta} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To illustrate these results, Table I provides a few calculations for several alternative values of \( \sigma^2 \). In each of the calculations presented in this table, the expected rate of annual growth in demand is, of course, identical—namely unity. Note that as the variance increases, so does the optimal size of installation, and the minimum value of expected discounted costs.

Other things equal, our model indicates that the riskier the growth in demand, the larger ought to be the amount invested in each installation. To some, this result will seem to fly in the face of common sense. However, to those familiar with models of inventory stockage under conditions of probabilistic demand (e.g., [1, pp. 256–259]), this should come as no paradox. In both this capacity model and in many cases of inventory control, the greater the risk of running out of capacity or out of inventory in a specified period of time, the greater the amount which it pays to invest in order to avert this contingency.

### 4. THE DETERMINISTIC MODEL—BACKLOGS CONSIDERED

It is now time to examine the zero backlog assumption in a more critical way, and to explore the implications that result from discarding it. The zero
backlog assumption seemed especially appropriate for the industry described by Chenery, natural gas transmission, an industry whose residential customers cannot readily be backlogged. Any attempt to do so would only result in a permanent switch in their allegiance to an alternative fuel.

From the zero backlog assumption, together with the economies of scale phenomenon, Chenery derives his "permanent" excess capacity hypothesis: "... excess capacity will occur even with perfect forecasting; this may be called 'optimum' overcapacity" [4, p. 2]. It would be a mistake, however, to suppose that the zero backlog assumption is a universally valid one. The economist who is accustomed to work with a downward sloping price-demand curve will certainly find it just as resonable to believe that backlogs are admissible, and that they are accompanied by some kind of penalty cost to the firm. The zero backlog model then turns out to be a special case—one in which backlog costs are infinite.

To a petroleum transporter, for example, these penalties are far less than infinite. If he is unable to ship crude or refined products via a pipeline, there is in almost all cases a transportation alternative available: tankship, barge, railroad tank-car, or tank truck. The penalty for failing to have enough pipeline capacity is simply the difference between the short-run marginal operating cost of the pipeline and the marginal cost of using the alternative mode of transport.

Similarly, to an underdeveloped country building up a steel industry, the penalty for being short of domestic capacity will consist of the difference between the marginal cost of steel imports and of domestic production. The shortage penalty will be incurred only intermittently, and will be confined to the phase of full-capacity operations.

In graphical terms, the analogy with Figure 2 is shown on Figure 5. Just as in the earlier case, we assume that demand grows linearly at the rate of one physical unit per year. Again, x units will denote the size of each new in-

\[ \text{Excess capacity} \]

\[ t_0 \]

\[ t_0 + x \]

\[ t_0 + 2x \]

\[ \text{Time} \]

\[ x-y \]

\[ y \]

\[ 0 \]

\[ \text{Figure 5.—Evolution of excess capacity over time.} \]
stallation and the points \( t_0, t_0 + x, t_0 + 2x \ldots \) still mark the points of regeneration: the points at which excess capacity has just been wiped out. The entire difference between this and the earlier case is that we allow excess capacity to become negative here, in other words, permit backlogs of demand. Once such backlogs become admissible, there is no longer any a priori reason to believe in the necessity of Chenery’s excess capacity hypothesis. With sufficiently low penalty costs, it is even conceivable that excess capacity will, on the average, be negative.

Figure 5 has been drawn on the assumption that whenever the backlog in demand grows to \( y \) units (that is, whenever excess capacity equals minus \( y \)), a new facility is built—one of size \( x \). We now have two decision variables: \( x \), the size of each installation, and \( y \), the “trigger” level for backlogs in demand.\(^7\) The rate of incurring penalty costs will be assumed strictly proportional to the quantity \( z \), hereafter employed to denote the size of the backlog. The total dollar magnitude of penalty costs within a cycle will be proportional to one of the triangular-shaped, shaded areas in Figure 5.

Looking forward into the future from a point of regeneration, total discounted costs are a function of both \( x \) and \( y \). If we denote these discounted costs by \( C(x,y) \), the expression that corresponds to (2.2) is as follows:

\[
(4.1) \quad C(x,y) = \int_{z=0}^{y} cze^{-rz}dz + e^{-ry}(hx^y) + e^{-rx}C(x,y)
\]

where \( c \) represents the annual rate of penalty costs per unit of backlog.

It is easy to see that when demand is growing steadily at the rate of one unit per year, a backlog of size \( z \) occurs exactly \( z \) years after a point of regeneration. The first term on the right hand side of (4.1) therefore measures the discounted sum of all penalty costs incurred during the course of a single construction cycle. The second term measures the installation costs, and discounts them \( y \) years back to the beginning of the cycle. Finally, the last term indicates the future value of all costs incurred in subsequent cycles, and discounts this value back over a period of \( x \) years. From (4.1), we readily obtain:

\[
C(x,y) = \frac{1}{1-e^{-rx}} \left[ c \int_{z=0}^{y} z e^{-rz}dz + e^{-ry}(hx^y) \right].
\]

Dividing through by \( h \) in order to eliminate one parameter, and christening

\(^7\) Any reader will note the striking similarity between this and the \((S,c)\) theory of optimal inventory policy. One important aspect tends to be concealed in the deterministic form of the two models. A replenishment lag is characteristic of the inventory studies, \([1, 2]\). In the interests of simplicity, however, the corresponding feature—a construction lag—is ignored in the present paper.
the ratio \( c/h \) with the name \( b \), we finally have the cost expression to be minimized:

\[
\frac{C(x,y)}{h} = \frac{1}{1 - e^{-vx}} \left[ b \int_{0}^{y} z e^{-vx} dz + e^{-vy}x^a \right]
\]

\[
= \frac{1}{1 - e^{-vx}} \left[ \frac{b}{y^a} \left[ 1 - e^{-vy} (1 + ry) \right] + e^{-vy}x^a \right].
\]

Expression (4.2) involves three parameters: \( a \), the economies-of-scale factor; \( b \), the penalty factor; and \( r \), the discount rate. Minimization of (4.2) with respect to both \( x \) and \( y \) could conceivably have been accomplished by calculus methods as in the earlier one-variable case, but this approach seemed rather clumsy.\(^8\) Instead, refuge was taken in numerical methods. An electronic computer\(^9\) evaluated \( C(x,y) \) for a large number of combinations of \( x \) and \( y \), and reported the minimum for each specified set of values of \( a \), \( b \), and \( r \). The results of four such calculations are shown in Table II. As in Table I,

| TABLE II |
| SHORTAGE PENALTY COSTS, OPTIMAL BACKLOG TRIGGER LEVELS AND CAPACITY INCREMENTS |
| \( \alpha = .50 \) and \( r = .15 \) |
| Shortage penalty costs = \( b \) | \( \infty \) | .50 | .10 | .05 |
| Optimal capacity increment = \( \dot{x} \) | 8 | \( \dot{y} \) | 12 | 21 |
| Optimal backlog trigger level = \( \dot{y} \) | 0 | 1 | 5 | 14 |
| Minimum discounted costs = \( C(\dot{x}, \dot{y})/h \) (see (4.2)) | 4.048 | 3.791 | 2.883 | 2.027 |

the parameters \( a \) and \( r \) were set at .50 and at .15, respectively. One word of caution about the numerical construction of this table: the decision variables \( x \) and \( y \) were restricted to integer values.

5. THE PROBABILISTIC MODEL—BACKLOGS CONSIDERED

The final stage of this investigation will consist of fitting together the two kinds of generalizations of Chenery's model: (a) probabilistic growth, and (b)

\(^8\) If we differentiate (4.2) with respect to \( y \), we obtain as a necessary condition:

\[
\dot{y} = \frac{r \dot{x}^a}{b}
\]

From this it is clear that \( \dot{y} \) vanishes only when \( b \), the unit penalty cost, becomes infinite.

\(^9\) The machine was the I.B.M. 650 located in the Yale University Computing Center. E. Uren performed the numerical analysis, with help from M. Davis, Director of the Center, and also from D. Ciusek.
backlogs in demand. Just as in the zero backlog case, we again assume the operation of a diffusion process such that \( D(t) \), the growth in demand that takes place in \( t \) years, is a normally distributed random variable: one with a mean of \( \mu t \) and a variance of \( \sigma^2 t \). The particular asymptotic process leading to this result consists of the cumulation of successive independent changes \( \varepsilon(t) \). (Refer back to (3.1).)

As before, we shall let \( u(t; x)dt \) represent the probability with which \( t \) time units have elapsed at the point when total demand first exceeds the initial level by \( x \) units. A similar definition holds for \( u(t; y)dt \). We already know the Laplace transforms for these two probability distributions:

\[
\begin{align*}
\tilde{u}(r; x) &= \int_{t=0}^{\infty} u(t; x)e^{-rt}dt = e^{rx}, \\
\tilde{u}(r; y) &= \int_{t=0}^{\infty} u(t; y)e^{-rt}dt = e^{ry}.
\end{align*}
\]

(Refer back to (3.6), (3.8), and (3.10).)

Now in order to deal with the backlog question, we shall have to introduce one additional piece of notation: \( w(z; t, y)dz \). This symbol will denote the probability with which the backlog level equals \( z \) given that \( t \) time units have elapsed since the most recent point of regeneration. Why does the decision variable \( y \) enter into the definition of this probability? Because the process of building up a backlog will come to an end as soon as demand has increased by a total of \( y \) units, that is to say, by an amount large enough to trigger off the construction of a new facility.

In Brownian motion language, \( w(z; t, y)dz \) represents the probability with which a particle, starting \( y \) units above the origin, will at time \( t \) be \( z \) units beneath its initial position, without having previously touched the absorbing barrier at the origin. Feller has already provided the generating function for the corresponding probability distribution in the case of discrete time and one-unit movements of the particle [8, problem 16, p. 336]. The analogous result for the Laplace transform in the continuous case is as follows:10

10 In order to derive this result, we recall the following definitions:

- \( u(z; t)dz \) is the probability that demand will change by exactly \( z \) units, given that \( t \) units of time have elapsed: a normal density function with mean \( \mu t \) and variance \( \sigma^2 t \) (unconstrained random walk with the particle initially at the origin).
- \( u(t; y)dt \) is the probability that exactly \( t \) time units have elapsed at the time when demand first exceeds its initial level by \( y \) units (absorbing barrier at the origin, with the particle initially located \( y \) units above the origin).

From these definitions,

\[
w(z; t, y)dz = u(z; t)dz - \int_{t=0}^{t} u(r; y)u(z-y; t-r)dzdr.
\]

Denote the Laplace transforms of \( u, v, \) and \( w \) by \( \tilde{u}, \tilde{v}, \) and \( \tilde{w} \), respectively. Then

\[
\tilde{w}(z; r; y) = \tilde{v}(z; r; y) - \tilde{u}(r; y)\tilde{u}(z-y; r).
\]

In order to derive \( \tilde{v}(z; r; y) \) and \( \tilde{u}(z-y; r) \), one follows the same line of reasoning as in
CAPACITY EXPANSION

(5.3) \[ \tilde{w}(z; r, y) = \int_{t=0}^{\infty} w(z; t, y) e^{-rt} dt = K [e^{\lambda_1 z} - e^{\lambda_2 y + \lambda_1 (y - z)}] \]

where \( \lambda_1 \) and \( \lambda_2 \) are as determined earlier by (3.8) and where the parameter \( K \) is given by:

(5.4) \[ K = \frac{\lambda_1 \lambda_2}{r(\lambda_2 - \lambda_1)}. \]

The cost equation for the new model may be written down by direct analogy with the deterministic one (4.1):

(5.5) \[ C(x,y) = \int_{z=0}^{y} c(z) \tilde{w}(z; r, y) dz + e^{\lambda_2 y} F(x, y) + e^{\lambda_2 y} C(x, y). \]

Total expected discounted costs, \( C(x, y) \), will, as in the preceding cases, be measured from a point of regeneration, a point at which excess capacity equals zero. Now the first term on the right-hand side of (5.5) measures the expected discounted sum of all backlog penalties incurred between this point of regeneration and the point at which the backlog reaches the critical level, \( y \). (See Figure 5; also equation (5.3).) The penalty cost integration extends over all possible backlog levels between 0 and \( y \). Note that it is quite conceivable that the backlog will become negative at any time after a point of regeneration. Our cost expression simply says that whenever this happens

deriving \( \bar{v}(r; x) \). One starts with the Fokker-Planck partial differential equation (see [8, pp. 325–6]):

\[ \frac{\partial v(z; t)}{\partial t} = -\mu \frac{\partial v(z; t)}{\partial z} + \frac{\sigma^2}{2} \frac{\partial^2 v(z; t)}{\partial z^2}. \]

Since the coefficients of this equation are identical with (3.5),

\[ \bar{v}(z; r) = K_1(r) e^{\lambda_1 z} + K_2(r) e^{\lambda_2 z}. \]

Here we have as our boundary conditions:

\[ \int_{-\infty}^{0} v(z; t) dz = 1, \]

\[ \int_{-\infty}^{\infty} \bar{v}(z; r) dz = \int_{t=0}^{\infty} \int_{r=0}^{\infty} v(z; t) e^{-rt} dz dt = \frac{1}{r}. \]

In order to satisfy these boundary conditions for all values of \( r \), and also in order to preserve continuity in the function \( \bar{v}(z; r) \):

\[ \bar{v}(z; r) = \begin{cases} K_1 e^{\lambda_1 z} & \text{according as } z \gg 0 \\ K_2 e^{\lambda_2 z} & \text{according as } z \ll 0 \end{cases} \]

where

\[ K = \frac{\lambda_1 \lambda_2}{r(\lambda_2 - \lambda_1)}. \]

Noting that \( z \gg 0 \), but that \( z - y \ll 0 \), this completes the derivation of \( \bar{v}(z; r) \), \( \bar{v}(z - y; r) \), and \( \bar{v}(r; y) \), and from these in turn, equations (5.3) and (5.4).

I am much indebted to Gerd Reuter for working out these expressions.
(that is, whenever demand drops off enough to create some excess capacity), no additional outlays are incurred beyond those that were previously committed.

The second term on the right hand side of (5.5) is the one having to do with construction costs during the current cycle. These costs are all incurred at the time of reaching the level $y$, and so the appropriate Laplace transform is (5.2).

Finally, the third term (that measuring the discounted sum of all costs incurred in subsequent cycles) refers to a cost that is dated as of the beginning of the following cycle. This cycle will begin whenever the total demand first increases by $x$ units over the current level, i.e., whenever the $x$ units of new capacity are, for the first time, fully utilized. In the third term, therefore, the appropriate Laplace transform is (5.1).

For purposes of numerical analysis, the cost function (5.5) may be rewritten:

$$ C(x,y) \frac{C(x,y)}{h} = \frac{1}{bK} \left[ \frac{\sigma^2}{\lambda^2} \left( 1 - e^{\lambda x} (1 - \lambda y) \right) - \frac{\sigma^2}{\lambda^2} \left( e^{\lambda x} (1 - \lambda y) \right) \right] + e^{\lambda y} x^2, $$

where $b$ again equals the ratio $c/h$, as in the deterministic calculations of the preceding section.

The numerical analysis of (5.6) is only slightly more complex than that of (4.2).\footnote{In fact, the same I.B.M. 650 program written to solve (5.6) also handled (4.2).}

| TABLE III |
| VARIANCE OF DEMAND VERSUS OPTIMAL BACKLOG TRIGGER LEVELS AND CAPACITY INCREMENTS |
| $(a = .50, b = .10, r = .15$ and $\mu = 1)$ |

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>0</th>
<th>1</th>
<th>4</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>$\frac{1}{\sigma^2} \left[ 1 + \sqrt{1 + 2\sigma^2} \right]$</td>
<td>$\infty$</td>
<td>2.1402</td>
<td>.6208</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$\frac{1}{\sigma^2} \left[ 1 - \sqrt{1 + 2\sigma^2} \right]$</td>
<td>$-1.500$</td>
<td>$-1.402$</td>
<td>$-.1208$</td>
</tr>
<tr>
<td>$K = \frac{\lambda_1 \lambda_2}{r(\lambda_2 - \lambda_1)}$</td>
<td>$-$</td>
<td>.8771</td>
<td>.6742</td>
<td>.4152</td>
</tr>
<tr>
<td>Optimal capacity increment = $\hat{x}$</td>
<td>12</td>
<td>14</td>
<td>16</td>
<td>21</td>
</tr>
<tr>
<td>Optimal backlog trigger level = $\hat{y}$</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Minimum expected discounted costs = $C(\hat{x}, \hat{y})/h$ (see (5.6))</td>
<td>2.883</td>
<td>2.834</td>
<td>2.831</td>
<td>3.079</td>
</tr>
</tbody>
</table>
parameters \(a\), \(b\), and \(r\). The only additional feature is that in the present case we must also take account of the Laplace transform parameters \(\lambda_1\), \(\lambda_2\), and \(K\). As can be seen from the row headings of Table III, these last-mentioned parameters all depend directly upon the variance \(\sigma^2\). Just as in the zero backlog case, we now examine the effects of increasing the variance while holding the expected increment in demand constant at unity. Also held constant in Table III are the parameters \(a\), \(b\), and \(r\).

Table III would be of little interest if it merely confirmed the backlog case; what we already knew about the zero backlog model: that an increase in variance is inevitably accompanied by an increase in \(C(\hat{x}, \hat{y})/h\), the minimum level of expected discounted costs. Instead, it provides an immediate counter-example to this conjecture. Minimum costs keep dropping as \(\sigma^2\) increases from zero to four times the expected annual increment in demand. Only for the case of \(\sigma^2 = 16\) does the level of expected costs increase again.

Stanford University

REFERENCES