The Effect of Technological Improvement on Capacity Expansion for Uncertain Exponential Demand with Lead Times

Dohyun Pak
Department of Industrial & Operations Engineering
1205 Beal Avenue
The University of Michigan
Ann Arbor, MI 48109

Nattapol Pornsalnuwat
Department of Industrial & Manufacturing Systems Engineering
2019 Black Engineering Building
Iowa State University
Ames, IA 50011-2164

Sarah M. Ryan*
Department of Industrial & Manufacturing Systems Engineering
2019 Black Engineering Building
Iowa State University
Ames, IA 50011-2164

March 2004

For publication in *The Engineering Economist*

*Corresponding Author:
email: smryan@iastate.edu
voice: 515-294-4347
fax: 515-294-3524

The Effect of Technological Improvement on Capacity Expansion for Uncertain Exponential Demand with Lead Times

Abstract

We formulate a model of capacity expansion that is relevant to a service provider for whom the cost of capacity shortages would be considerable but difficult to quantify exactly. Due to demand uncertainty and a lead time for adding capacity, not all shortages are avoidable. In addition, technological innovations will reduce the cost of adding capacity but may not be completely predictable. Analytical expressions for the infinite horizon expansion cost and shortages are optimized numerically. Sensitivity analyses allow us to determine the impact of technological change on the optimal timing and sizes of capacity expansions to account for economies of scale, the time value of money and penalties for insufficient capacity.

Introduction

Capacity expansion is the process of adding new facilities of similar types over time to meet a rising demand for their services. Planning for the expansion of capacity is of vital importance in many applications within the private and public sectors. Examples can be found in heavy process industries, communication networks, electrical power service, and water resource systems. Capacity expansion planning consists of determining future expansion times, sizes, locations and types of facilities in the face of uncertain demand forecasts, costs, and completion times.

In many new technology industries, demand for capacity grows according to an exponential trend. For example, recent work by Dumortier [5] predicted the exponential growth in the number of Internet users and the end-user multimedia applications. Rai et al. [24] used the number of service hosts as the measure of the Internet size and suggested that an exponential model provided the closest fit with the increasing number of service hosts. Kruger [12] predicted the significant growth of electric consumption due to substitution of hydrogen for fossil fuels in motor vehicles. The major concern is the magnitude of additional electric power capacity necessary to build a large-scale hydrogen fuel industry, especially in a state such as California with large number of vehicles.

The effect of a construction lead time for adding new capacity is also an important issue in capacity expansion problems. If a lead time for adding capacity exists, the capacity expansion problem is more complicated because uncertain demand creates the risk of potentially costly shortages during the construction period. If there were no lead times for adding new capacity, despite the uncertainty of demand there would be no risk of capacity shortage, since the manager could simply wait until demand equals current capacity and then install new capacity instantaneously.

Technological progress is an important factor to be considered in capacity expansion problems. Traditionally, improvements in technology are measured either in terms of increased revenue associated with the new technology or decreased costs of procurement and operation of the new technology. In many industries, such as commercial satellite communications or computer central processing units, introduction of the new improved technology causes increases in product efficiency and rapid decreases in unit costs, thus affecting the expansion planning decision. Technological change also enlarges markets indirectly through improved productivity.

Productivity improvements reduce production costs. Falling costs enable price reductions and expand the customer base and thus the market.

Several examples show how technological progress could influence the cost of expansion. Snow observed that the per-unit capacity cost of satellite communication INTELSAT was decreased significantly due solely to technological progress [30]. Newer, improved technology of the satellite component expanded voice channel capacity. During the 12 years of consideration, the voice channel capacity had been increased from 480 to 25000, which is more than 50 times, while the capital cost per satellite increased only 3.77 times from 16.5 to 62.3 million dollars. This ratio of channel increase to capital cost increase clearly shows how technological improvement could affect the cost of expansion. Moore's Law, which stated that computer CPU speed would be doubled every 18 months, is another distinctive example of how technological progress affects the cost of expansion in information technology industries. The improvement in technology implied by Moore's Law enhances productivity, while simultaneously causing older technology to become obsolete and prices to drop regularly. With this technological progress, a manager can choose to purchase the latest technology at the highest price, or purchase the older technology for a lower price.

Through all the examples mentioned above, we can see that a capacity expansion problem with uncertain demand growth could be more complicated under the technological progress environment. The total cost of expansion over a long horizon will be considerably different from the expansion problem with stationary technology. Some research in the past explored various capacity expansion problems and determined the optimal policies for those cases, but none of them have combined the consideration of random exponential demand growth and the uncertain technological change together with lead times for expanding capacity. This paper investigates optimal capacity expansion policies. The goal of this research is to determine the impact of technological change on the optimal timing and sizes of capacity expansions to account for economies of scale, the time value of money and penalties for insufficient capacity.

We formulate a model of capacity expansion that is relevant to a service provider for whom the cost of capacity shortages would be considerable but difficult to quantify exactly. The objective function for the optimal policy is the total cost of expansion over an infinite time horizon combined with a penalty for shortages. We determine the expansion policy in two dimensions. The first dimension concerns the timing of each expansion in terms of the relationship between demand and capacity when the expansion is initiated. The second dimension

concerns size, i.e., the amount of capacity to be added by each expansion in view of cost discounting and economies of scale. An analytical expression for the total cost allows numerical solution for the policy parameters that minimize a weighted combination of the total discounted expected expansion cost and the cost of shortage.

Relevant Capacity Expansion Studies

Since the late 1950s, many studies of capacity expansion problems have been conducted. Sinden [29] studied the capacity expansion problem of certain facilities providing service for a growing population, such as a power plant, a transportation system, or a telephone system. Sinden assumed the demand for services as a function of time is given. The facility must expand and replace its equipment from time to time in order to meet its demand. He showed that in certain cases, there is an optimal expansion policy with equal time intervals between successive expansions. Manne [16] studied the capacity expansion problem with probabilistic growth, including the penalties involved in accumulating backlogs of unsatisfied demand. The results showed that uncertainty in demand growth causes a larger size of capacity expansion. Another study by Manne of several heavy process industries in India is an example widely known for its application [17]. In these models the demand growth follows a linear trend. The total cost of expansion over an infinite time horizon can be calculated simply by summing the costs of each replenishment discounted back to time zero.

Srinivasan [31] extended Manne's work to the growth of heavy industries in India. He formulated a model in which demand grows at a constant geometric rate, and assumed that there are no demand backlogs (excess demand). When the economies of scale in construction are incorporated into the capacity expansion cost, it is optimal to expand capacity at each of a sequence of equally spaced time points. Therefore, the optimal expansion size would grow exponentially. Srinivasan also assumed that technology is static, the construction lead time for adding new capacity is zero, and demand growth is deterministic. A survey of Luss [15] can be consulted as an extensive literature review on capacity expansion. In his survey, Luss unified the existing literature, emphasizing modeling approaches, algorithmic solutions and relevant application.

Various models have been formulated for the capacity expansion problem with random demand. Freidenfelds [6] studied the effects of uncertainty in demand on capacity expansion decisions. He specified demand as a birth and death process for fixed expansion increment and showed that the effect of randomness is identical to the effect of a larger growth rate. Bean et al. [1] showed that the capacity expansion problem over an infinite horizon

with demand that follows either a nonlinear Brownian motion or a non-Markovian birth and death process could be transformed into an equivalent deterministic problem. This equivalent deterministic problem is formed by replacing the stochastic demand by its deterministic trend and discounting costs by a new deterministic equivalent interest rate, which is smaller than the original, in approximate proportion to the uncertainty in the demand. More details of this result will be discussed in Section 4.

Some studies of capacity expansion model include lead times for adding capacity. Nickell [20] formulated a model with an uncertain future change in demand and showed that the existence of a fixed lead time for adding new capacity would cause a firm to introduce a capacity increase earlier. He also showed that a longer lead time results in earlier anticipation of demand increases. Davis et al. [4] presented a more mathematical model of the capacity expansion process of large scale projects that incorporated a controllable non-zero lead time of constructing new capacity into uncertain future demand forecast model. Their demand model was a random point process that increases by discrete amounts at random times. The capacity expansion model also included a cost associated with failure to meet demand, and a cost of wasteful overcapacity. They studied the problem by methods of stochastic control and presented a numerical algorithm to determine the optimal policy. Chaouch and Buzacott [3] studied the effect of lead time on the timing of plant construction with the objective of minimizing the expected discounted costs of expansions and shortages over the infinite time horizon. Their model has a certain fixed lead time of construction that is independent of plant size. The demand grows alternately with a constant rate in some periods and stagnant growth in other periods. They suggested that it may be economically attractive to defer plant construction beyond the time when existing capacity becomes fully absorbed. Longer lead times increase the optimal capacity trigger levels and sizes of capacity additions. Ryan [28] studied the problem having correlated random demand with a linear trend. Also, Ryan [27] formulated a dynamic programming model of capacity expansion for uncertain exponential demand growth and deterministic expansion lead times, and used option pricing formulas to estimate the shortages to result from a capacity expansion policy. With the expected lead time shortage fixed, the discounted expansion cost could be minimized by expanding capacity by a constant multiple of existing capacity.

Several studies include the effect of technological progress on the capacity expansion models. Snow [30] reviewed the previous work of Manne and Srinivasan and included a technological progress parameter in the capacity expansion model of the communications satellite INTELSAT. This added parameter is the annual

exponential rate at which prices fall due solely to the effect of technological progress. Snow showed how technological change affects the capacity expansion model by adding a constant to interest rate, thus decreasing the discounted cost of each replenishment. Other previous studies that suggested the importance of technological change on capacity expansion or replacement decisions are listed as follows. Goldstein et al. [7] studied the effect of technological breakthroughs on the machine replacement problem. They presented a dynamic discounted cost model and a method for finding the optimal age for replacement of an existing machine in a technological development environment. In their research, they assumed that a new technological breakthrough is about to enter the market in the form of new machine, which has higher purchase cost but lower maintenance costs than the existing machine. Hopp and Nair [8] developed a procedure for computing the optimal replacement decision in an environment of technological change. Their model assumed that the costs associated with the present and future technologies are known, but the appearance times of the future technologies are uncertain. Nair [19] also studied uncertain sequential technological change, which affects the firm's strategic investment decisions. He suggested that the appearance of the future technologies are considered uncertain with probabilities that may vary with time, but the order in which they appear is assumed sequential, such as the different generations of microchips for personal computers. Finally, he developed an approach using nonunique terminal rewards to solve a dynamic programming model of the replacement problem. All the previous works discussed above demonstrate the importance of technological change in the capacity expansion problem. The prediction and forecasting of technological change itself was described by Porter et al. [23], who discussed the models of technology growth based on the previous work of Gompertz and Fischer-Pry. They suggested that the growth in capacity of many technologies is exponential over a considerable time period. Rajagopalan et al. [25] formulated a capacity expansion and replacement model with a sequence of technological breakthroughs. They modeled the stochastic technological change as a semi-Markov process by specifying the distribution of the time between two consecutive innovations and the matrix of transition probabilities for the levels of technology achieved. By proper choice of the time-to-discovery distribution, their model can accommodate the diverse characteristics of timing between innovations across different industries.

Problem Definition and Notation

Let $D(t)$ be the demand for service at time $t \geq 0$. We assume that this demand follows a geometric Brownian motion with drift $\mu > 0$ and variance σ^2. It follows that, given $D(t)$, the growth in the logarithm of

demand over an interval beginning at time t is normally distributed with mean and variance proportional to the length of the interval:

$$\ln\left(\frac{D(t+\Delta t)}{D(t)}\right) = \mu\Delta t + \sigma\sqrt{\Delta t}Z, \tag{1}$$

where Z is a standard normal random variable. Alternatively, we can characterize the demand at a future time point as lognormally distributed with conditional expected value:

$$E\left[D(t+\Delta t)\middle|D(t)\right] = D(t)e^{g\Delta t}, \tag{2}$$

where $g = \mu + \sigma^2/2$ is the expected rate of exponential growth in demand. Note that, although demand can increase or decrease over time, it can never become negative. This model is appropriate for demand patterns with the following characteristics:

- Although demand can increase and decrease over time, the long term expected trend is upward.
- The expected demand at the end of a period is best expressed as a constant percentage increase over the demand at the beginning of the period.
- The uncertainty in the logarithmic demand growth over an interval, as measured by its variance, is proportional to the length of the interval. This characteristic is consistent with the diminishing reliability of forecasts as they extend into the future.

Luenberger [14] explains how the parameters μ and σ can be estimated from historical data as the sample mean and standard deviation of $\ln\left(D(k+1)/D(k)\right)$, using data collected at equally spaced time points k = 1, 2, … in the past. The assumption of exponentially increasing demand over an infinite horizon is strong. However, the geometric Brownian motion assumption for demand is supported by statistical analysis of the actual usage of electric power and airline travel over recent decades [18]. Also, the expansion size policy assumed by Whitt [32] for geometric Brownian motion demand without lead times and with no particular assumption about expansion costs, which is the same policy used in this paper, was found to match the actual industry-wide expansions of chemical production capacity over two decades [13].

In order to meet the predicted growth in demand, capacity can be increased by any continuous quantity, but economies of scale imply that the expansions will occur at discrete time points rather than continuously over time. Specifically, we assume that $C(X,t)$, the cost of an expansion of size X at time t, satisfies

$C(X,0) = kX^a$, $0 < a < 1$, where k is a proportionality constant and a is an economy of scale parameter. Without loss of generality, assume costs are scaled so that k=1. We further assume for simplicity that the expansion cost is incurred all at once when the expansion is initiated. We study two models of the cost impact of technological innovation. The first, as in [30], assumes a deterministic exponential decrease in the cost of capacity due to technological change, so that $C(X, t+\Delta t) = e^{-p\Delta t} C(X,t)$ for all t, where $p > 0$. The second model assumes that technological innovations follow a Poisson process with rate λ, and that the average rate of cost reduction per innovation is $q > 0$. Under randomly timed technological change, the cost is $C(X, t+\Delta t) = e^{-qN(\Delta t)} C(X,t)$, where $N(t)$ is Poisson distributed with mean λt. There is a fixed lead time, L, required for any size expansion. Costs are continuously discounted at rate r. We assume $r > g$ to guarantee that the expected infinite horizon costs incurred by the expansion policy converge. Intuitively, even though expansion sizes keep pace with demand growth, incurring exponentially increasing costs, potential shortages also increase exponentially over time. Therefore, for optimization to be meaningful, both expansion and shortage costs must be discounted at a rate higher than the expected demand growth rate.

Let n be an index for the sequence of expansions. An expansion policy $\{(t_n, X_n), n \geq 1\}$ consists of a sequence of expansion time points and capacity increments. See Figure 1. For a given policy, let K_n be the installed capacity after n expansions have been completed, where $K_0 > D(0)$ is the initial capacity and, for $n \geq 1$, $K_n = K_{n-1} + X_n$. The installed capacity at time t is given by

$$\mathrm{K}(t) = \begin{cases} K_0, & 0 \leq t < t_1 + L \\ K_n, & t_n + L \leq t < t_{n+1} + L, \; n \geq 1, \end{cases} \tag{3}$$

while the capacity position is

$$\Pi(t) = \begin{cases} K_0, & 0 \leq t < t_1 \\ K_n, & t_n \leq t < t_{n+1}, \; n \geq 1. \end{cases} \tag{4}$$

The capacity position includes available capacity and any additional capacity that is under construction.

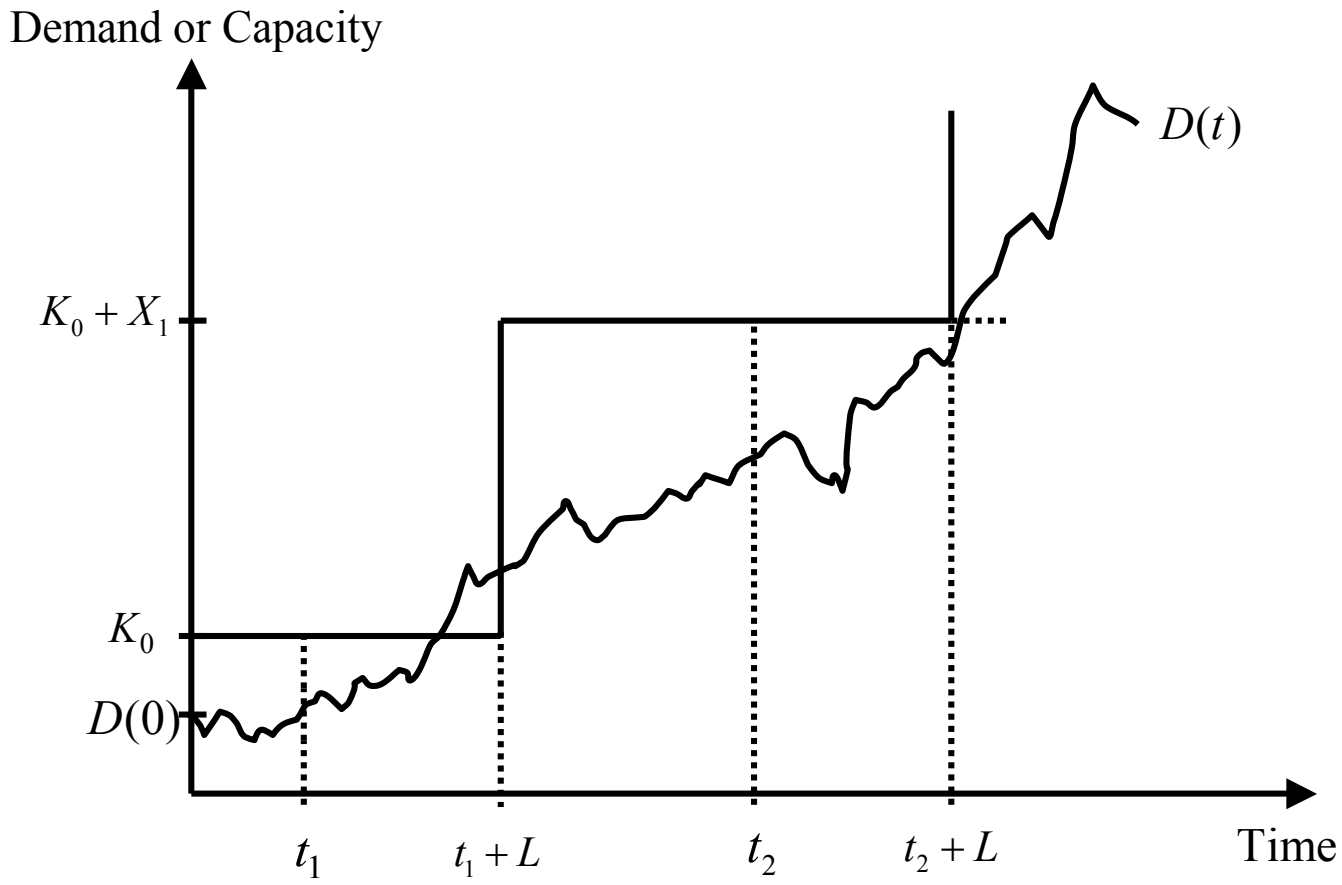

Figure 1. Illustration of capacity expansion problem and potential for shortages during lead times.

We assume that there is a significant penalty of *m* per unit of shortage per unit time. Therefore, in light of the lead times, each expansion should be initiated before a shortage occurs. Since in the worst case the detection of a shortage would automatically trigger an expansion, it follows that the risk of shortage is present only during lead times for expansion. On the other hand, if expansions occur far in advance of need, high opportunity costs result from the overinvestment. The problem is to find an expansion policy that minimizes the sum of expected infinite horizon discounted expansion cost and shortage penalties.

Expansion Policy

Ryan [27] showed that the shortage during any lead time, expressed as a proportion of installed capacity, depends on previous events and decisions only through the ratio of demand to capacity position at the beginning of the lead time, $D(t_n)/K_{n-1}$. Pak [21] showed how the total expected shortage throughout the lead time could be obtained by numerical integration of an analytical formula that depends on the demand-to-capacity ratio. In this

paper we assume a timing policy in which an expansion will be triggered whenever demand reaches a constant (over time) proportion of the capacity position, so that $D(t_n)/K_{n-1} = \gamma < 1$ for all n. If lead times do not overlap, then the capacity position at time t_n is equal to the installed capacity at that time. The specific proportion, γ, that is used is a decision variable. This policy is equivalent to the proportional reserve policy that has traditionally been used by many utilities [10].

Ryan [27] also showed that, under this timing policy, the infinite horizon expected discounted cost of expansions is minimized when each capacity increment equals the same proportion of the current capacity position, i.e., $X_n = xK_{n-1}$ for each n. That proof relied on the equivalent deterministic formulation of Bean et al., [1], in which the true interest rate, r, is adjusted down according to the demand uncertainty. The second decision variable, x, is found using this lowered interest rate, r^*, along with the deterministic demand function $D^*(t) = D(0)e^{\mu t}$. In this paper, we show that deterministic or random technological change are both equivalent to increases in the interest rate. Therefore, we use the form of the policy suggested by Ryan [27], though we note that it has not yet been proved optimal for minimizing the weighted combination of expansion cost and shortage.

Let $T(y) \equiv \min\{t \geq 0 : D(t) = y\}$. If an expansion takes place when demand reaches y for the first time, the expected discount factor for the expansion cost is $E\left[e^{-rT(y)}\right]$. The equivalent interest rate for the deterministic problem is found from the fact that $E\left[e^{-rT(y)}\right] = \exp\left(-r^* \ln\left(y/D(0)\right)/\mu\right)$ (see [11]). Note that, in the preceding expression, $\ln\left(y/D(0)\right)/\mu$ is the time at which the deterministic demand function, $D^*(t)$, reaches y. Under our policy, $t_n = T(\gamma K_{n-1})$, $X_n = x(x+1)^{n-1} K_0$ and $K_n = (x+1)^n K_0$.

Expected Discounted Cost of Expansions

We develop expressions for expected discounted cost in three cases: no technological change, deterministic technological change, and random technological change. Comparing Equations (8), (10) and (14) below, we will see that the effects of these different assumptions are captured by modifications to one parameter, ρ, used in the cost discounting factors [22].

If there were no cost impact of technological change, the infinite horizon expected discounted expansion cost would be given by:

$$u(\gamma, x) = E\left[\sum_{n=1}^{\infty} e^{-rt_n} (X_n)^a\right] = \sum_{n=1}^{\infty} E\left[e^{-rT\left(\gamma(x+1)^{n-1} K_0\right)}\right]\left(x(x+1)^{n-1} K_0\right)^a. \tag{5}$$

The expected discount factor for the *n*th expansion can be written as

$$E\left[e^{-rT\left(\gamma(x+1)^{n-1} K_0\right)}\right] = \left(\frac{D(0)}{\gamma(x+1)^{n-1} K_0}\right)^{\rho}, \tag{6}$$

where

$$\rho = \frac{r^*}{\mu} = \frac{\mu}{\sigma^2}\left(\sqrt{1 + 2r\left(\frac{\sigma}{\mu}\right)^2} - 1\right). \tag{7}$$

Bean et al. [1] pointed out that $r^* < r$. One can also verify by algebra that $r > g$ if and only if $\rho > 1$. Using these facts, the infinite horizon expected discounted cost may be evaluated in closed form as:

$$\begin{aligned} u(\gamma, x) &= \sum_{n=1}^{\infty}\left(\frac{D(0)}{\gamma(x+1)^{n-1} K_0}\right)^{\rho}\left(x(x+1)^{n-1} K_0\right)^a \\ &= \left(\frac{D(0)}{\gamma K_0}\right)^{\rho}(xK_0)^a \sum_{n=1}^{\infty}\left((x+1)^{n-1}\right)^{a-\rho} \\ &= \left(\frac{D(0)}{\gamma K_0}\right)^{\rho}\frac{(xK_0)^a}{1-(x+1)^{a-\rho}}, \end{aligned} \tag{8}$$

since $a - \rho < 0$.

Under our assumption concerning deterministic technological change, the corresponding cost is given by

$$u_D(\gamma, x) = E\left[\sum_{n=1}^{\infty} e^{-rt_n} e^{-pt_n} (X_n)^a\right] = \sum_{n=1}^{\infty} E\left[e^{-(r+p)t_n}\right](X_n)^a. \tag{9}$$

As noted by Snow [30], we see that in this model, the cost impact of technological improvement is equivalent to an increase in the interest rate. Technological change would have the qualitative effects of delaying investment and/or reducing the size of each expansion since it is anticipated that the same capacity will be available for less cost in the future. We can quantify this effect using the simpler expression for the cost:

$$u_D(\gamma, x) = \left(\frac{D(0)}{\gamma K_0}\right)^{\rho_D}\frac{(xK_0)^a}{1-(x+1)^{a-\rho_D}}, \tag{10}$$

where

$$\rho_D = \frac{\mu}{\sigma^2}\left(\sqrt{1+2(r+p)\left(\frac{\sigma}{\mu}\right)^2}-1\right). \tag{11}$$

If, as is more likely the case, technological innovations cannot be predicted with certainty, the closed form expression for the cost is slightly more complicated to derive. According to our model, the expected infinite horizon discounted cost of expansions under random technological change is:

$$u_R(\gamma,x) = E\left[\sum_{n=1}^{\infty} e^{-rt_n} e^{-qN(t_n)} (X_n)^a\right] = \sum_{n=1}^{\infty} E\left[e^{-rt_n} e^{-qN(t_n)}\right](X_n)^a. \tag{12}$$

For any t, since $N(t)$ is a Poisson random variable, we can use the Poisson moment generating function $E\left[e^{-qN(t)}\right] = e^{-\left(1-e^{-q}\right)\lambda t}$ [26, p.60] to obtain an equivalent deterministic cost decrease rate of $p = \left(1-e^{-q}\right)\lambda$. The discount factor for the cost of the nth expansion can then be found by conditional expectation.

$$\begin{aligned} E\left[e^{-rt_n} e^{-qN(t_n)}\right] &= E_{t_n}\left[E\left[e^{-rt_n} e^{-qN(t_n)} \middle| t_n\right]\right] \\ &= E_{t_n}\left[e^{-rt_n} E\left[e^{-qN(t_n)} \middle| t_n\right]\right] \\ &= E_{t_n}\left[e^{-rt_n} e^{\lambda t_n\left(e^{-q}-1\right)}\right] \\ &= E_{t_n}\left[e^{-\left(r+\lambda\left(1-e^{-q}\right)\right)t_n}\right]. \end{aligned} \tag{13}$$

Therefore, the closed form expression for the cost under random technological change is:

$$u_R(\gamma,x) = \left(\frac{D(0)}{\gamma K_0}\right)^{\rho_R} \frac{(xK_0)^a}{1-(x+1)^{a-\rho_R}}, \tag{14}$$

where

$$\rho_R = \frac{\mu}{\sigma^2}\left(\sqrt{1+2\left(r+\lambda\left(1-e^{-q}\right)\right)\left(\frac{\sigma}{\mu}\right)^2}-1\right). \tag{15}$$

Expected Discounted Cost of Shortages

Ideally, one would prefer to be able to adjust capacity continuously in order to exactly meet the demand as it occurs. However, when there are economies of scale and lead times for adding capacity, continuous capacity adjustment is not economical. There could be penalties associated with over- as well as undercapacity but, to many service providers, having insufficient capacity to meet the demand has far more serious consequences than having

The Version of record of this manuscript has been published and is available in Engineering Economist 2004, http://www.tandfonline.com/10.1080/00137910490453301. Posted with permission.

too much capacity. Further, by including the time value of money in our cost functions, we are already encouraging the postponement of capacity expansions until they are needed. Therefore, the second component of the total cost deals strictly with penalties for capacity shortage to balance against the expansion cost.

The shortage at a future point in time is a random variable that represents the difference between demand and installed capacity, if this difference is positive. Let $s(t) = \max\left(D(t) - \mathrm{K}(t), 0\right)$ be this random quantity. At time *t* in the interval $\left[t_{n-1} + L, t_n + L\right)$, the instantaneous shortage is $s(t) = s_n(t) \equiv \max\left(D(t) - K_{n-1}, 0\right)$. Under our policy, where $D(t_n) = \gamma(x+1)^{n-1} K_0$, it is possible for lead times to overlap, i.e., $t_n < t_{n-1} + L$ if demand during the (*n*-1)st lead time grows very quickly. Ryan [27] showed that the probability of this occurring is constant over *n* and can be made as small as desired by choosing *x* large enough. Therefore, in our analysis, we neglect the possibility of overlapping lead times and assume that $s(t) = s_n(t)$ throughout the *n*th lead time $\left[t_n, t_n + L\right)$.

Due to the Markovian character of the geometric Brownian motion model for demand, the shortage during a lead time depends only on the installed capacity and the demand at the beginning of the lead time. Specifically, for the *n*th lead time we are interested in the total expected shortage discounted to the beginning of the lead time:

$$S_n = E\left[\int_{t_n}^{t_n+L} e^{-r(t-t_n)} s_n(t)\, dt \,\middle|\, D(t_n)\right] \tag{16}$$

When measured as a proportion of the current capacity, this expected value can be evaluated using a formula developed for financial option pricing [9].

Lemma: *If V is a lognormal random variable and the standard deviation of* $\ln V$ *is s,*

then $E\left[\max(V - K, 0)\right] = E(V)\Phi(d_1) - K\Phi(d_2)$, *where* $d_1 = \left(\ln\left(E[V]/K\right) + s^2/2\right)/s$,

$d_2 = d_1 - s$ *and* $\Phi(\cdot)$ *is the standard normal cumulative distribution function*.

Theorem: *Under the timing policy where* $t_n = \min\left\{t : D(t_n) = \gamma K_{n-1}\right\}$, *the ratio* S_n / K_{n-1} *is independent of* n *and can be evaluated numerically as*:

$$\frac{S_n}{K_{n-1}} = f(\gamma) \equiv \int_0^L \left(\gamma e^{(g-r)\tau} \Phi\left(h(\gamma, \tau)\right) - e^{-r\tau} \Phi\left(h(\gamma, \tau) - \sigma\sqrt{\tau}\right)\right) d\tau\,, \tag{17}$$

where $h(\gamma, t) = \left(\ln\gamma + \left(\mu + \sigma^2\right)t\right) / \left(\sigma\sqrt{t}\right)$.

Proof: According to our demand model, for $t > t_n$, the ratio $D(t)/D(t_n)$ is lognormal and the standard deviation of its natural logarithm is $\sigma\sqrt{t-t_n}$. Therefore, given $D(t_n)$, $D(t)$ is also lognormal and the standard deviation of $\ln D(t)$ equals the standard deviation of $\ln D(t) - \ln D(t_n)$, which is also $\sigma\sqrt{t-t_n}$. The expected value of $D(t)$ given $D(t_n)$ is $D(t_n)e^{g(t-t_n)}$. Therefore, according to the Lemma,

$$E\left[e^{-r(t-t_n)}S_{n-1}(t)\middle|D(t_n)\right] = D(t_n)e^{(g-r)(t-t_n)}\Phi(h_1) - e^{-r(t-t_n)}K_{n-1}\Phi(h_2),$$

where

$$h_1 = \frac{\ln\left(D(t_n)e^{g(t-t_n)}/K_{n-1}\right) + \left(\sigma^2/2\right)(t-t_n)}{\sigma\sqrt{t-t_n}} = \frac{\ln\left(D(t_n)/K_{n-1}\right) + \left(\mu+\sigma^2\right)(t-t_n)}{\sigma\sqrt{t-t_n}} \text{ and}$$

$h_2 = h_1 - \sigma\sqrt{t-t_n}$. Since $D(t_n) = \gamma K_{n-1}$ under the timing policy,

$$E\left[s_{n-1}(t)\middle|D(t_n)\right]/K_{n-1} = \gamma e^{g(t-t_n)}\Phi\left(h(\gamma, t-t_n)\right) - \Phi\left(h(\gamma, t-t_n) - \sigma\sqrt{t-t_n}\right).$$

Substitute $\tau = t - t_n$ in the integral for S_n/K_{n-1} to obtain the result. ■

The Lemma can also be used to derive the Black-Scholes formula for the value of a call option on an asset. Birge [2] pointed out the correspondence between a limit on capacity and an option on any demand that exceeds the capacity limit. Specifically, in a competitive environment, having a limited capacity can be seen as equivalent to selling one's competitors the option to satisfy the excess demand.

To evaluate the expected discounted penalty due to shortages, we can multiply a penalty factor *m* per unit shortage per unit time by the function:

$$v(\gamma, x) = \sum_{n=1}^{\infty} E\left[e^{-rt_n}\right]S_n, \tag{18}$$

which represents the infinite horizon expected discounted shortage. Using the Theorem and $\rho < 1$, we can write $v(\gamma, x)$ in an analytical form as:

The Version of record of this manuscript has been published and is available in Engineering Economist 2004, http://www.tandfonline.com/10.1080/00137910490453301. Posted with permission.

$$v(\gamma,x)=\sum_{n=1}^{\infty}\left(\frac{D(0)}{\gamma K_{n-1}}\right)^{\rho}S_n=\left(\frac{D(0)}{\gamma}\right)^{\rho}f(\gamma)\sum_{n=1}^{\infty}\left((x+1)^{n-1}K_0\right)^{1-\rho}$$
$$=\left(\frac{D(0)}{\gamma}\right)^{\rho}\frac{f(\gamma)K_0^{\,1-\rho}}{1-(x+1)^{1-\rho}}. \tag{19}$$

Total Expected Discounted Cost

The overall total cost to be minimized is the sum of the expansion cost and the shortage penalty:

$$w(\gamma,x)=u(\gamma,x)+mv(\gamma,x). \tag{20}$$

Note that the ρ values used in the cost function $u(\gamma,x)$ will vary according to the existence and type of technological change, while the ρ values used in the shortage function $v(\gamma,x)$ will be unaffected by innovations. Under deterministic technological change, the total cost is specified as:

$$w_D(\gamma,x)=\left(\frac{D(0)}{K_0}\right)^{\rho_D}\frac{\gamma^{-\rho_D}x^a}{1-(x+1)^{a-\rho_D}}K_0^{\,a}+m\left(\frac{D(0)}{K_0}\right)^{\rho}\frac{\gamma^{-\rho}f(\gamma)}{1-(x+1)^{1-\rho}}K_0. \tag{21}$$

Under random technological change, the objective function is identical, with subscripts "*D*" replaced by "*R*." The factors $(D(0)/K_0)^{\rho}$, where ρ can be replaced by either ρ_D or ρ_R in the expansion cost term, can be interpreted as adjustments to the discount factors to account for the ratio of demand to capacity at time 0. The function $c(\gamma,x)\equiv\gamma^{-\rho_D}x^a/\left(1-(x+1)^{a-\rho_D}\right)$ is a dimensionless multiplier for the time 0 cost of providing an expansion of size K_0, while its counterpart $d(\gamma,x)\equiv\gamma^{-\rho}f(\gamma)/\left(1-(x+1)^{1-\rho}\right)$ is a dimensionless multiplier for a shortage of K_0 capacity units. We have not yet been able to prove that $w(\gamma,x)$, $w_D(\gamma,x)$ or $w_R(\gamma,x)$ are jointly convex functions of the decision variables γ and x. However

- The function $c(\gamma,x)=c_1(\gamma)c_2(x)$, where $c_1(\gamma)=\gamma^{-\rho}$ is strictly decreasing and convex. Since $a<1$ and $\rho>1$, one can show that there is a unique $x^*>0$ such that $c_2'(x^*)=0$, $c_2'(x)<0$ for $x<x^*$, and $c_2'(x)>0$ for $x>x^*$. Thus costs are lowered by delaying expansions and choosing the expansion size that optimally trades economies of scale against cost discounting.

- The function $d(\gamma, x) = d_1(\gamma)\, d_2(x)$, where $d_2(x) = \left(1 - (x+1)^{1-\rho}\right)^{-1}$ is convex decreasing. The behavior of $d_1(\gamma)$ is difficult to verify analytically, but numerical plots indicate it is convex increasing. We would expect that making larger expansions would reduce the size and likelihood of shortages, while postponing expansions increases the risk of shortage.

In the next section, we examine numerical solutions and their sensitivity to changes in the problem parameters. To be sure the solutions are optimal, we have verified convexity numerically by checking that the Hessian matrix for the total cost function is positive semidefinite on the domain $(\gamma, x) \in (0,1) \times (0,1)$ for all the combinations of parameter values used.

Numerical Results and Sensitivity Studies

As a baseline numerical case, we used the parameter values μ = 0.05 and σ = 0.2 in the demand model. The expected rate of exponential demand growth is g = 0.07, and we assumed an annual nominal interest rate r = 0.10 > g. The values for the lead time and the economy of scale parameter were L = 0.5 years and a = 0.7, respectively. We chose arbitrary values of $D(0)$ = 50 and K_0 = 100. For these parameter values with a shortage penalty factor of m = 5, by numerically minimizing $w(\gamma, x)$ we find optimal values of γ = 0.84 and x = 0.75. Therefore, each new expansion should be initiated when demand reaches 84% of capacity, and each expansion should increase capacity by 75%. Larger penalty factors reduce the optimal value of γ significantly, so that expansions are undertaken earlier, and also decrease the optimal x value slightly. Using this expansion size in Theorem 4 of [27], the probability that a lead time overlaps the previous one is just 0.00015, supporting our assumption that lead times do not overlap. Lead time overlap probabilities are similarly small for all the numerical cases except as noted below for long lead times.

Figure 2 and 3, respectively, show the effects of changes in the mean logarithmic growth rate, μ, and the volatility of demand, σ, for various penalty factors. The effect of increasing either the expected growth in demand or its uncertainty is to expand capacity earlier and in larger amounts. The relative magnitudes of these policy adjustments vary with the size of the penalty factor. The case where σ = 0 represents deterministic demand

The Version of record of this manuscript has been published and is available in Engineering Economist 2004, http://www.tandfonline.com/10.1080/00137910490453301. Posted with permission.

$D^*(t) = D(0)e^{\mu t}$. In this case, t_n satisfies $D(0)e^{\mu t_n} = \gamma K_{n-1}$ and shortages during the nth lead time commence at time t_n' such that $D(0)e^{\mu t_n'} = K_{n-1}$. The shortage ratio function can be evaluated in closed form as:

$$f_0(\gamma) = \frac{1}{K_{n-1}} \int_{t_n'}^{t_n+L} e^{-r(t-t_n)} \left(D(0)e^{\mu t} - K_{n-1} \right) dt = \frac{\mu\gamma^{r/\mu} + (r-\mu)e^{-rL} - \gamma r e^{-(r-\mu)L}}{r(r-\mu)}. \tag{22}$$

Also, since

$$E\left[e^{-rt_n} \right] = e^{-rt_n} = \left(\frac{D(0)}{\gamma (x+1)^{n-1} K_0} \right)^{r/\mu}, \tag{23}$$

we can replace ρ by r/μ throughout.

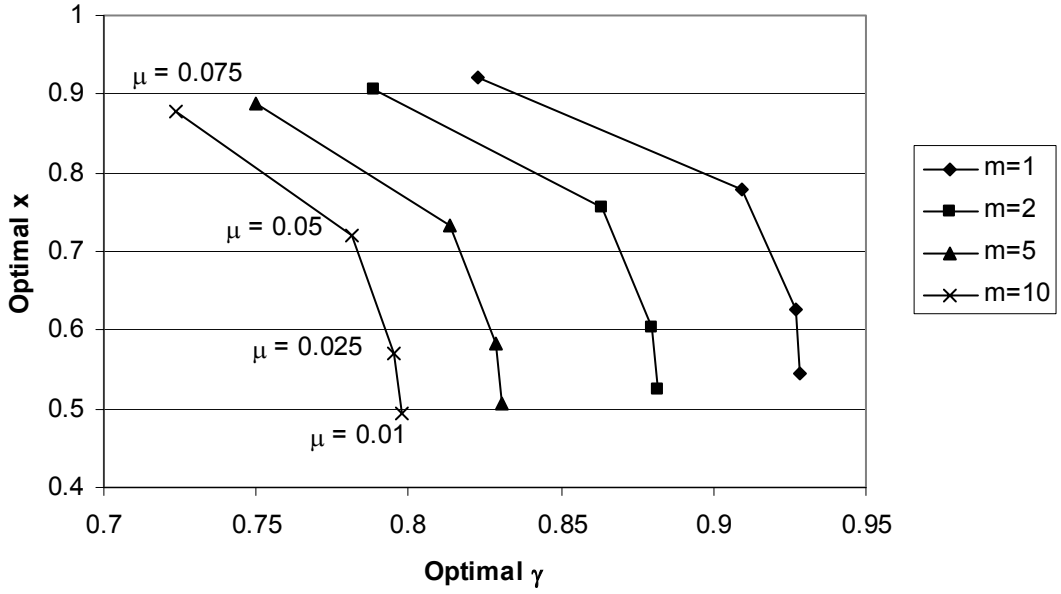

Figure 2. Effect of the mean logarithmic growth rate of demand on the optimal values of the policy parameters.

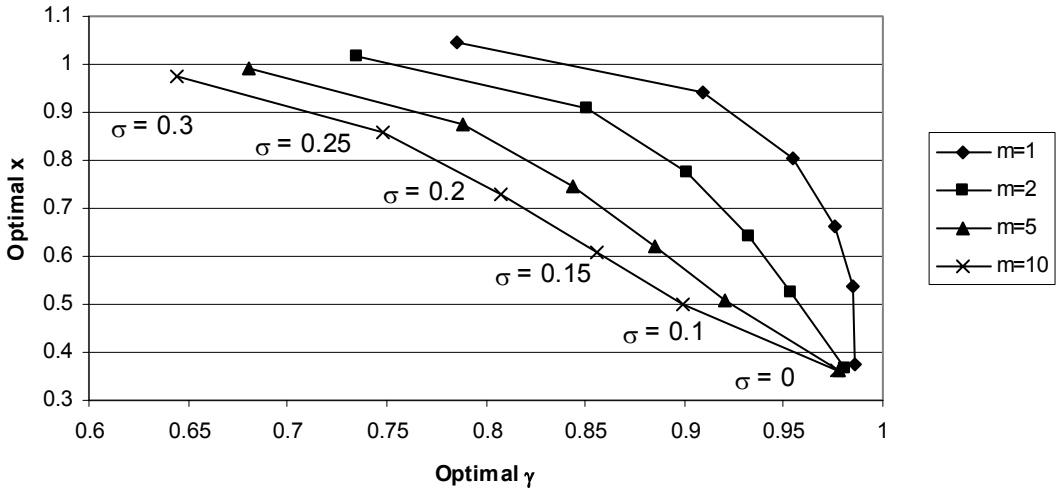

Figure 3. Effect of the demand uncertainty on the optimal policy parameters.

The effect of the lead time length is shown in Figure 4 to prompt earlier expansions in larger sizes. These larger expansions can also be interpreted as occurring less frequently. A long lead time therefore reduces the flexibility to wait and see what happens to demand and respond in small capacity increments. It should be mentioned that for $L = 2$, when applying the optimal expansion sizes, the probability of lead time overlap climbs as high as 0.05. Therefore, the impact of shortages is not fully captured in this case and it may be necessary to decrease γ and/or increase x. If $L = 0$, the expected lead time shortages would vanish, so that $w(\gamma, x) = u(\gamma, x)$. When considering only the expansion cost, since $\partial u / \partial \gamma < 0$, one would delay expansions indefinitely. Figure 5 shows that the primary effect of increasing economies of scale (decreasing values of a) is to increase the size of expansions, as would be expected. However, in the simultaneous optimization of both policy parameters, the expansions also occur somewhat earlier.

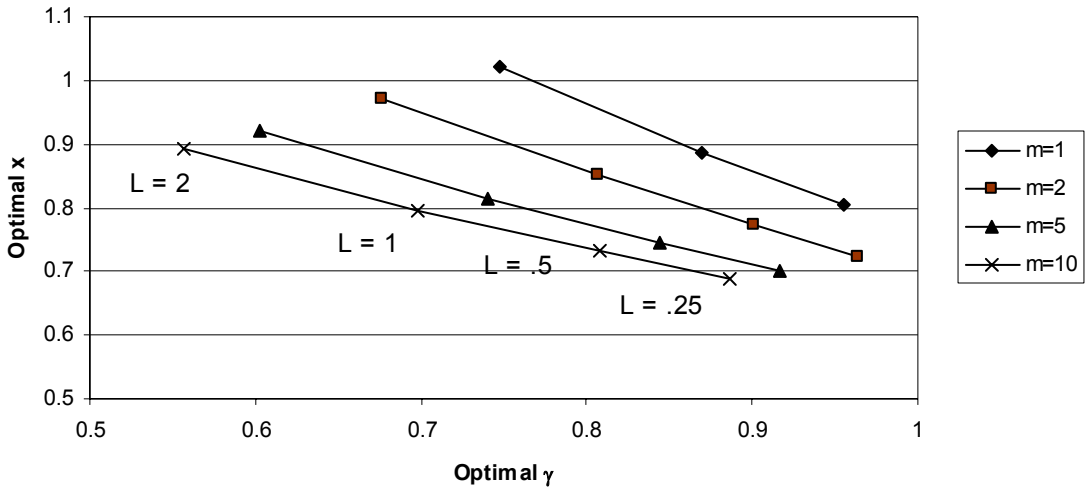

Figure 4. Effect of the lead time length on the optimal policy parameters.

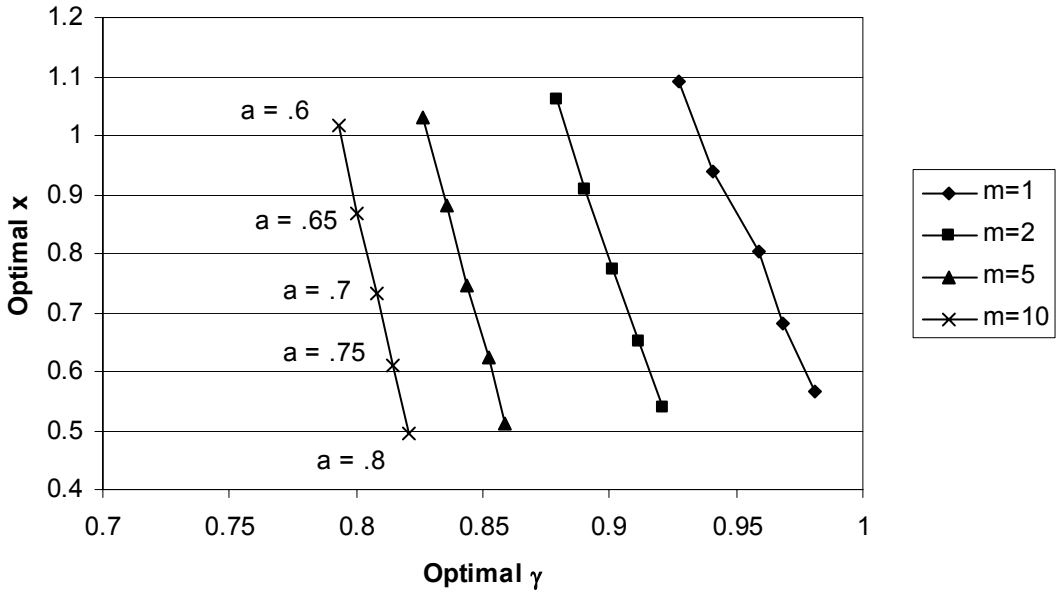

Figure 5. Effect of economies of scale on the optimal policy parameters.

Finally, we can observe the impacts of technological change on the optimal policy parameters. Figure 6 depicts the values of γ and x that minimize $w_D(\gamma, x)$ for various values of the deterministic technological change parameter, p. As expected, if capacity costs are expected to decrease in the future, the optimal expansions are

smaller. Not so intuitively, the expansions also should begin earlier, when demand reaches a relatively smaller proportion of capacity. Similar effects are seen in Figure 7 and 8, from minimizing $w_R(\gamma, x)$ for different values of q and λ. In Figure 7, the innovation rate, λ, is held constant at 0.5 per year while the rate of cost decrease per innovation (q) varies. In Figure 8, q is fixed at 0.25 and λ assumes the values shown in the chart.

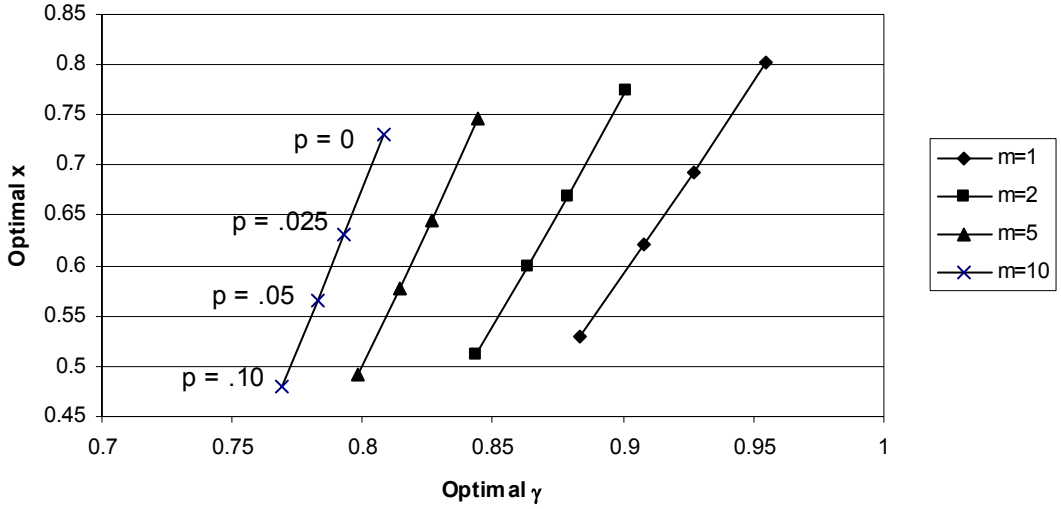

Figure 6. Effect of deterministic cost decrease due to technological change on the optimal policy parameters.

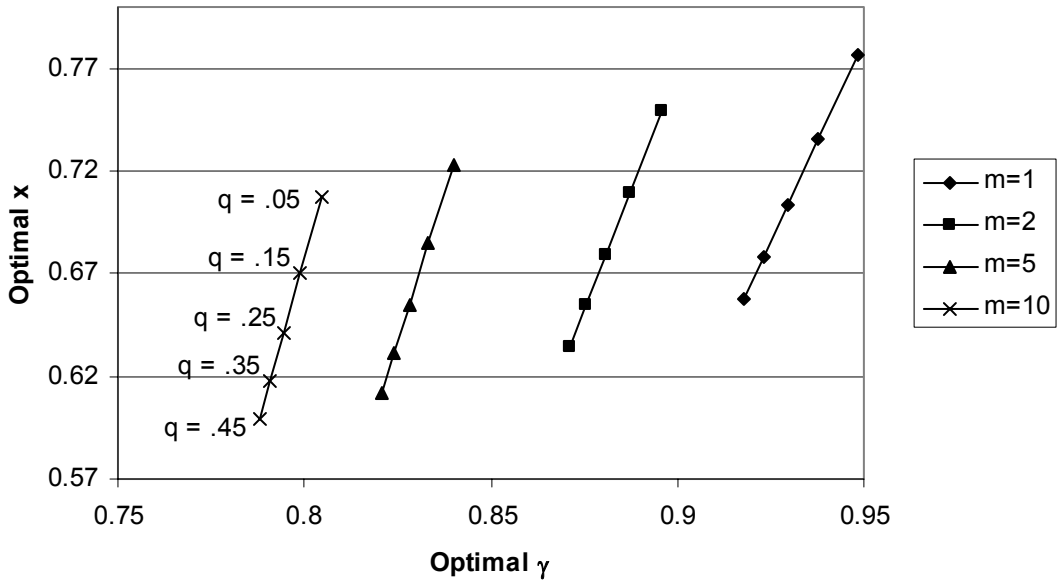

Figure 7. Effect of cost decrease per random innovation on the optimal policy parameters.

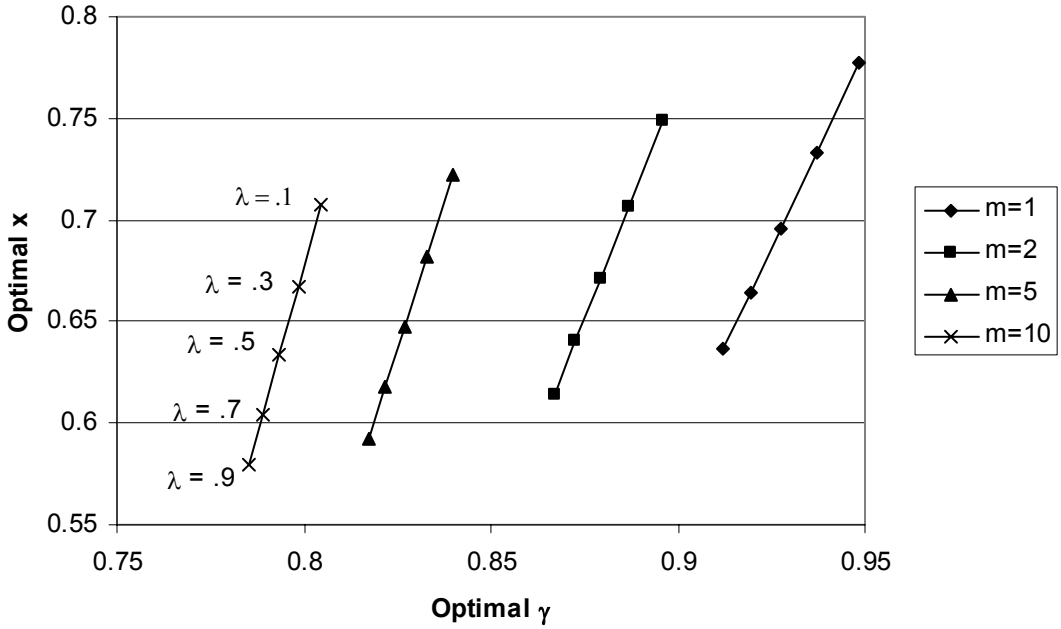

Figure 8. Effect of the rate of random technological innovations on the optimal policy parameters.

6. Discussion and Conclusions

Managers of service facilities faced with increasing demand must wrestle with the questions of when to expand capacity and by how much. In this paper we have considered a model that includes the interactions and, in some cases, conflicts among several problem characteristics. Uncertainty in the demand growth combined with a lead time for adding capacity creates the risk of capacity shortage even under the assumption that expansions are initiated while excess capacity remains. Economies of scale encourage larger and less frequent expansions, while cost decreases due to technological change motivate a wait-and-see attitude.

The mathematical model clarifies some of the relationships between problem characteristics and reveals some unexpected interactions between solution characteristics. Previous research revealed that uncertainty in the demand reduces the interest rate [1], but the cost impact of deterministic technological change is to increase the interest rate [30]. Our model for the impact of random technological change on the capacity cost also implies a higher interest rate. Further, it shows how demand uncertainty and technological change combine to affect movements in the interest rate that decision makers should use when solving a deterministic equivalent problem.

However, since the parameter of the Poisson process determines both the expected rate of technological change and its variance, further study is needed to isolate the effect of uncertainty in innovations. By deriving analytical expressions for the total cost as a function of both timing and sizing decisions, we have observed the interactions between these policy dimensions. For example, the economy of scale parameter might be expected to influence expansion decisions solely through the size parameter, x. On the contrary, our results show that it also affects the optimal degree of anticipation of future demand growth as expressed by the timing parameter, γ.

Our model and sensitivity studies treat each problem characteristic as independent of the others. More realistic models, which we defer to future research, could consider dependencies among them. For example, in some technology driven markets, the introduction of new innovations can affect the growth rate of demand. In addition, considering the lead time as a function of the capacity increment would be of interest for some industries such as energy generation. Other valuable extensions would be to model the lead time as a controllable variable, a random factor, or a function of the technology used to provide capacity.

Acknowledgment

This work was supported by the National Science Foundation under grant number DMI-9996373. Thanks to Rahul Marathe for his help with the numerical computations and convexity analysis.

References

[1] Bean, J.C., Higle, J. and Smith, R.L., Capacity expansion under stochastic demands, *Operations Research*, 40 (1992) S210-S216.

[2] Birge, J.R., Option methods for incorporating risk into linear planning models, *Manufacturing & Service Operations Management*, 2 (2000) 19-31.

[3] Chaouch, B.A. and Buzacott, J.A., The effects of lead time on plant timing and size, *Production and Operations Management*, 3 (1994) 38-54.

[4] Davis, M.H.A., Dempster, M.A.H., Sethi, S.P. and Vermes, D., Optimal capacity expansion under uncertainty, *Advances in Applied Probability*, 19 (1987) 156-176.

[5] Dumortier, P., Shortcut techniques to boost Internet throughput, *Alcatel Telecommunications Review* (1997) 300-306.

[6] Freidenfelds, J., *Capacity Expansion: Analysis of Simple Models with Applications*, North-Holland, New York, 1981.

[7] Goldstein, T., Ladany, S.P. and Mehrez, A., A discounted machine-replacement model with an expected future technological breakthrough, *Naval Research Logistics*, 35 (1988) 209-220.

[8] Hopp, W.J. and Nair, S.K., Timing replacement decisions under discontinuous technological change, *Naval Research Logistics*, 38 (1991) 203-220.

[9] Hull, J.C., *Options, Futures and Other Derivatives*, 4th edn., Prentice Hall, Upper Saddle River, NJ, 2000, 698 pp.

[10] Kahn, E., *Electric Utility Planning & Regulation*, American Council for an Energy-Efficient Economy, Washington, DC, 1988.

[11] Karlin, S. and Taylor, H.M., *A First Course in Stochastic Processes*, 2nd edn., Academic Press, New York, 1975.

[12] Kruger, P., Electric power requirement in California for large-scale production of hydrogen fuel, *International Journal of Hydrogen Energy*, 25 (2000) 395-405.

[13] Lieberman, M.B., Capacity utilization: theoretical models and empirical tests, *European Journal of Operational Research*, 40 (1989) 155-168.

[14] Luenberger, D.G., *Investment Science*, Oxford University Press, New York, 1998.

[15] Luss, H., Operations research and capacity expansion problems: a survey, *Operations Research*, 30 (1982) 907-947.

[16] Manne, A.S., Capacity expansion and probabilistic growth, *Econometrica*, 29 (1961) 632-649.

[17] Manne, A.S., Calculation for a single production area. In A.S. Manne (Ed.), *Investments for Capacity Expansion*, MIT Press, Cambridge, 1967, pp. 28-48.

[18] Marathe, R. and Ryan, S.M., On the Validity of the Geometric Brownian Motion Assumption.,, Iowa State University, Ames, IA, 2003.

[19] Nair, S.K., Modeling strategic investment decisions under sequential technological change, *Management Science*, 41 (1995) 282-297.

[20] Nickell, S., Uncertainty and lags in the investment decisions of firms, *Review of Economic Studies*, 44 (1977) 249-263.

[21] Pak, D., Option pricing methods for estimating capacity shortages, M. S. Thesis, Industrial & Manufacturing Systems Engineering, Iowa State University, Ames, IA, 2002.

[22] Pornsalnuwat, N., Capacity Expansion with Technological Change, M. S. Thesis, Industrial & Manufacturing Systems Engineering, Iowa State University, Ames, IA, 2002.

[23] Porter, A., Roper, A.T., Mason, T., Rossini, F. and Banks, J., *Forecasting and Management of Technology*, Wiley-Interscience, New York, 1991.

[24] Rai, A., Ravichandran, T. and Samaddar, S., How to anticipate the Internet's global diffusion, *Communications of the ACM*, 41 (1998) 97-106.

[25] Rajagopalan, S., Singh, M.R. and Morton, T.E., Capacity expansion and replacement in growing markets with uncertain technological breakthroughs, *Management Science*, 44 (1998) 12-30.

[26] Ross, S.M., *Introduction to Probability Models*, Third edn., Academic Press, Orlando, 1985.

[27] Ryan, S.M., Capacity expansion for random exponential demand growth with lead times.,, Industrial & Manufacturing Systems Engineering, Iowa State University, Ames, IA, 2003.

[28] Ryan, S.M., Capacity expansion with lead times and correlated random demand, *Naval Research Logistics*, 50 (2003) 167-183.

[29] Sinden, F.X., The replacement and expansion of durable equipment, *Journal of the Society of Industrial & Applied Mathematics*, 8 (1960) 466-480.

[30] Snow, M.S., Investment cost minimization for communications satellite capacity: refinement and application of the Chenery-Manne-Srinivasan model, *Bell Journal of Economics*, 6 (1975) 621-643.

[31] Srinivasan, T.N., Geometric rate of growth of demand. In A.S. Manne (Ed.), *Investments for Capacity Expansion: Size, Location, and Time-Phasing*, MIT Press, Cambridge, 1967, pp. 150-156.

[32] Whitt, W., The stationary distribution of a clearing process, *Operations Research*, 29 (1981) 294-308.

Biographical Sketches

Dohyun Pak is a doctoral student in Industrial & Operations Engineering at The University of Michigan. He completed a M.S. in Industrial & Manufacturing Systems Engineering at Iowa State University. This paper is based in part on his master's thesis entitled, "Option Pricing Methods for Estimating Capacity Shortages." As a graduate assistant, he is involved in teaching financial engineering courses. His research interests are in asset pricing, modeling of telecommunication networks and production optimization, optimal investment under uncertainty, and risk management.

Nattapol Pornsalnuwat earned an M.S. in Industrial & Manufacturing Systems Engineering at Iowa State University. This paper is based in part on his master's thesis entitled, "Capacity Expansion with Technological Change." He received his B.S. in Mechanical Engineering from Chulalongkorn University in Thailand. He previously worked for an engineering consultants company.

Sarah M. Ryan is an associate professor of Industrial & Manufacturing Systems Engineering at Iowa State University. She teaches courses in optimization, stochastic modeling and engineering economic analysis. Her research uses stochastic models to study long term investment decision problems as well as resource allocation problems in manufacturing. She is the recipient of a Faculty Early Career Development (CAREER) Award from the National Science Foundation, which funded the research leading to this article. She serves on the editorial board of *IIE Transactions* and as an Area Editor for *The Engineering Economist*.