PLANNING HORIZONS FOR THE DETERMINISTIC CAPACITY PROBLEM

ROBERT L. SMITH*

Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, U.S.A.

Abstract—We consider the problem of determining the optimal timing and sizing of new facility construction under a continuous technology model of possible sizes. The rate of convergence of short to long run optimal policies is established and planning horizon implications are explored. We also present an efficient algorithm for solving such problems for all horizon times by solving a single horizon time problem.

1. INTRODUCTION

The deterministic capacity problem [4, 5, 7, 14, 15, 19, 20] had its genesis in the now classic work of Manne [8]. He proposed a model for the augmentation of productive capacity in the face of a linearly increasing demand for service over an unlimited horizon. His solution struck an optimal tradeoff between the competing interests of deploying large capacity increments to exploit economy of scale effects as opposed to small increments to defer fixed cost expenditures. The model and its generalizations have been extensively applied in the analysis of capital investment problems in the process [3, 9], communications [4, 14, 15] and electric power [1] industries.

In practice, actual capacity expansion decisions are made under conditions that share few of the assumptions of the deterministic capacity problem [21]. The projected demand is rarely linear, or even smooth, since it is usually driven by solutions to a wider multilocation problem; the facility costs only roughly follow a power cost law; and so on. These practical considerations force the planner into use of a numerical algorithm and hence into at best a finite horizon solution to his problem†. One of the planner's greatest dilemmas is how long a horizon time to select so that end of horizon effects do not distort the first capacity deployment decision. It is of course this first decision which is the only one implemented and therefore the critical one. A *finite* horizon solution to the deterministic capacity problem can lend considerable insight into this kind of sensitivity. A finite horizon focus also comes into play in Jaskold–Gabszewicz and Vial [6] where a version of the deterministic capacity problem is considered wherein technological innovation can be introduced at some future time, leading to a direct consideration of finite horizon solutions as a step toward generating an infinite horizon solution.

In this paper, we present an efficient algorithm that solves the deterministic capacity problem for all horizons by solving a single finite horizon problem. We also provide formulas for upper and lower bounds for the first capacity expansion size as explicit functions of horizon time. These bounds establish planning horizon‡ sensitivity to forecast horizons for a range of possible demand profiles and facility costs.

Section 2 gives a formulation for the general horizon deterministic capacity problem and discusses its relationship to models considered in the literature. Section 3 presents structural results for the optimal policy including planning horizons. Section 4 details an efficient finite algorithm that solves any problem of the deterministic capacity type to within a prespecified maximal error for all horizon times T.

---

*Robert L. Smith is an Associate Professor in the Department of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. He received a Ph.D. in Operations Research from the Department of Industrial Engineering and Operations Research at the University of California at Berkeley. His research interests include probability modeling, statistics in optimization, and capacity expansion. His publications have appeared in *Operations Research*, *Journal of the American Statistical Association*, *Management Science*, and a number of other journals.

†In a case where projected demand is best fitted by a linear or exponential growth curve, an infinite horizon solution can be found under arbitrary facility fixed costs [13, 17].

‡A planning horizon [11] in our context refers to the time period over which the first (implementable) capacity expansion increment is exhausted. Since the planning horizon never truly stabilizes for any finite horizon time, and moreover its long run value depends on the entire demand profile, we are strictly speaking considering near (weak) planning horizons [11].

2. THE GENERAL HORIZON DETERMINISTIC CAPICITY PROBLEM

The infinite horizon version of the deterministic capacity problem is extensively discussed in Smith[15], including its formulation, applications, and solution. We will recast that formulation into a more general form which allows horizon time to be finite.

We assume a continuous technology of facility sizes allowing us to select a capacity increment of any size we choose. Defining the *fill time* of a facility as the time between installation and capacity exhaust of that facility as it satisfies demand for capacity, we may accordingly select a facility corresponding to any fill time $\tau > 0$. The fixed cost of providing capacity at time t that exhausts at time $t + \tau$ will be denoted by $f(\tau)$ and is incurred at installation time. Note that the model assumes that this undiscounted cost is independent of installation time t. Such an assumption is clearly satisfied under linear demand, and in fact as we will see below, is also satisfied under a range of non-linear demand profiles. Finally, all costs are discounted with a continuous factor $e^{-\nu t}$. The *Deterministic Capacity Problem* is:

(P)   To find a (finite or infinite) sequence of facilities with fill times $\langle \tau_i \rangle$ spanning the horizon time T so as to minimize present worth of total costs $\langle f(\tau_i) \rangle$ discounted at rate ν.

It should be noted that an optimal solution $\langle \tau_i^* \rangle$ may not exist, either for example, because the infimum of present worth total cost over an infinite horizon is infinite or alternatively the infimum is zero over a finite horizon. The problem for $T = \infty$ has been considered in various special cases by Manne[8], Srinivasan[19], Giglio[5] and others, and in the general case by Smith[15]. The finite horizon case $T < \infty$ has received considerably less attention although Freidenfelds[4], Skoog[14] and Jaskold–Gabszewicz and Vial[6] have analyzed some special cases of (P).

We will impose some general assumptions on the cost structure $f(\tau)$ in the next section that will allow for a general analysis of (P) for the finite horizon case. We will under these conditions establish existence of an optimal solution, present an algorithm for finding the optimal solution, and develop structural results yielding rates of convergence of finite horizon solutions to the infinite horizon solution as the horizon time is extended. Before doing this, we will first formulate three special cases of (P) that cover all versions of (P) analyzed in the literature, and furthermore satisfy the general assumptions we will impose on (P) in this paper.

In each case, we shall use the following concepts and terminology. The problem is to choose a sequence of facility sizes $\langle X_i \rangle$ so as to satisfy demand for capacity over a horizon T at minimum discounted cost. The growth in demand for capacity through time t is given by $D(t)$. The discounted cost of deploying a facility of capacity X at time t is given by $e^{-rt}F(X)$ where $r > 0$ is the force of interest and $F(X)$ is the undiscounted fixed cost of deploying a facility of capacity X. Table 1 gives specific forms for $D(t)$ and $F(X)$ for each of the three models together with the corresponding elements of (P). Proofs that these transformations are valid will not be given here, but may be found in Smith[15].

Model 1 has been considered among others by Srinivasan[19] and Valcoff[20], Model 2 by Manne[8] and McDowell[10] and Model 3 by Giglio[5]. The interested reader should see Smith[15] for telephone industry applications of these models.

3. STRUCTURAL RESULTS FOR THE FINITE HORIZON PROBLEM

We will now focus attention in the rest of this paper on the finite horizon version of (P) with $T < \infty$. As mentioned, the infinite horizon version is extensively analyzed in [15]. It is noted there that in Model 1, for sufficiently high growth rates, we may have the case that $\nu < 0$. In order to avoid notational confusion and in some cases for more substantive reasons which will be noted as we proceed, we will henceforth assume that ν is a true interest rate, i.e. $\nu \geq 0$.

A central tool of the analysis that will be used throughout this paper for stating and proving

Table 1.

	Model 1 (Increasing Marginal Demands)	Model 2 (Constant Marginal Demands)	Model 3 (Decreasing Marginal Demands)
Natural Model			
Demand $D(t)$	$D_0(e^{gt}-1)$, $D_0>0$ $g > 0$	dt, $d > 0$	$D_0(1-e^{-dt})$, $D_0>0$ $d > 0$
Facility Cost $F(X)$	$KX^{1-\alpha}$, $K>0$, $0 < \alpha < 1$	$A + BX$, $A > 0$ $B > 0$	$KX^{1-\alpha}$, $K > 0$, $0 < \alpha < 1$
Transformed Model (P)			
Facility Cost $f(\tau)$	$K(D_0(e^{g\tau}-1))^{1-\alpha}$	$A + Bd\tau$	$K(D_0(1-e^{-d\tau}))^{1-\alpha}$
"interest" rate ν	$\nu = r-(1-\alpha)g$	$\nu = r$	$\nu = r + (1-\alpha)d$

results is the notion of the equivalent cost rate $\gamma(\tau)$ of a facility filling in time τ.

Definition 3.1: $$\gamma(\tau) = \frac{\nu f(\tau)}{1 - e^{-\nu\tau}} * \quad \text{for } \tau > 0.$$

The analog of γ for a discrete problem was used in Shapiro and Wagner[13] where it is called a growth rate. Erlenkotter[2] introduced a somewhat more general form of γ in a different context. Smith [15] notes that γ is essentially a linearized version of the fixed cost $f(\tau)$, or more specifically $f(\tau) = \int_0^\tau \gamma(\tau) e^{-\nu t} dt$. $\gamma(\tau)$ may thus be regarded as a common yardstick for measuring the relative costliness of facilities. Moreover, it is shown in [2], [15] that for $\langle \tau_i^* \rangle$ optimal, $\gamma(\tau_i^*) \le \gamma(\tau_{i+1}^*)$ for all i for $\nu \ge 0$ (the ordering reverses for $\nu < 0$). We shall impose the following assumption on $\gamma(\tau)$ throughout.

Assumption 3.2: (a) $\gamma(\tau)$ is a continuous function over $\tau > 0$. (b) $\gamma(\tau)$ is monotone decreasing for sufficiently small values of τ and monotone increasing for sufficiently large values of τ. (c) $\gamma(\tau)$ is a strictly unimodal function over $\tau > 0$.

It is shown in [15] that under Assumptions 3.2(a, b) a minimum τ^* exists for $\gamma(\tau)$ and further that an equal cycle infinite horizon policy of period τ^* is optimal.

Lemma 3.3: Models 1 and 2 satisfy Assumption 3.2. Model 3 satisfies Assumption 3.2 if and only if $\nu > d$.

Proof: Assumptions 3.2(a,b) have been established for Models 1–3 in [15]. Assumption 3.2(c) is established in the appendix. ∎

See Fig. 1 for graphs of $\gamma(\tau)$ for Models 1 and 3.

Lemma 3.4:: For all $n \ge 1$, there exists a unique $\tau_{(n)}$ such that

$$\gamma(\tau_{(n)}) = \gamma\left(\frac{n+1}{n}\tau_{(n)}\right)$$

Proof: See appendix. ∎

See Fig. 2 for a graphic discovery of $\tau_{(n)}$. The implication of Lemma 3.4 may for now be viewed in the following way. Consider (P) for horizon $T = (n+1)\tau_{(n)}$. Then the cost of installing $n + 1$ facilities each filling in time $\tau_{(n)}$ is the same as the cost of installing n facilities each filling in time $((n + 1)/n)\tau_{(n)}$. Hence $\tau_{(n)}$ represents a breakpoint where we are indifferent between installing n vs $n + 1$ equal period facilities.

*Note that for $\nu = 0$, by L'Hospital's Rule, $\gamma(\tau) = \frac{f(\tau)}{\tau}$.
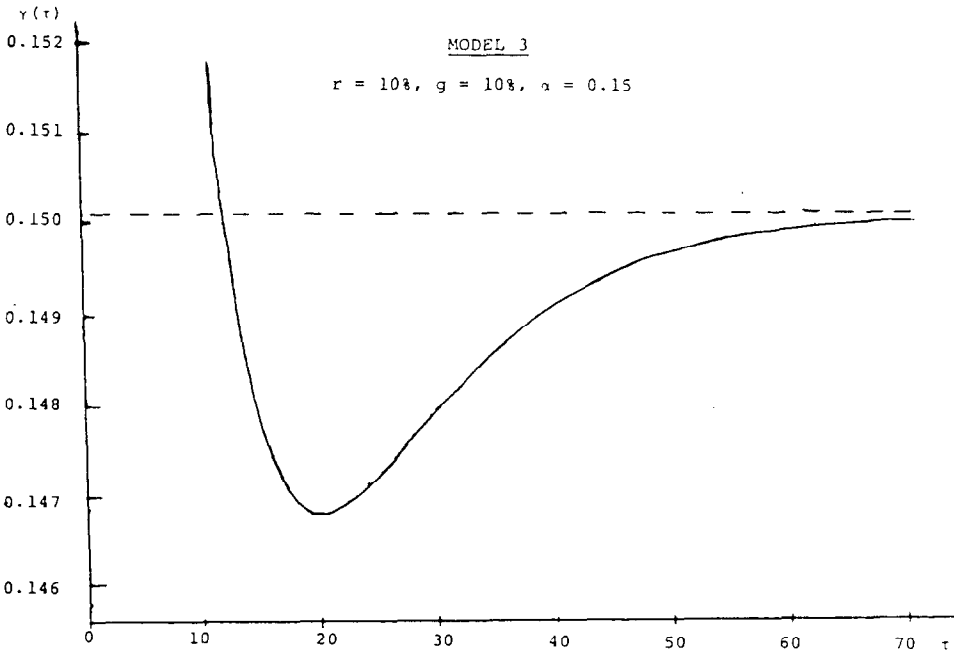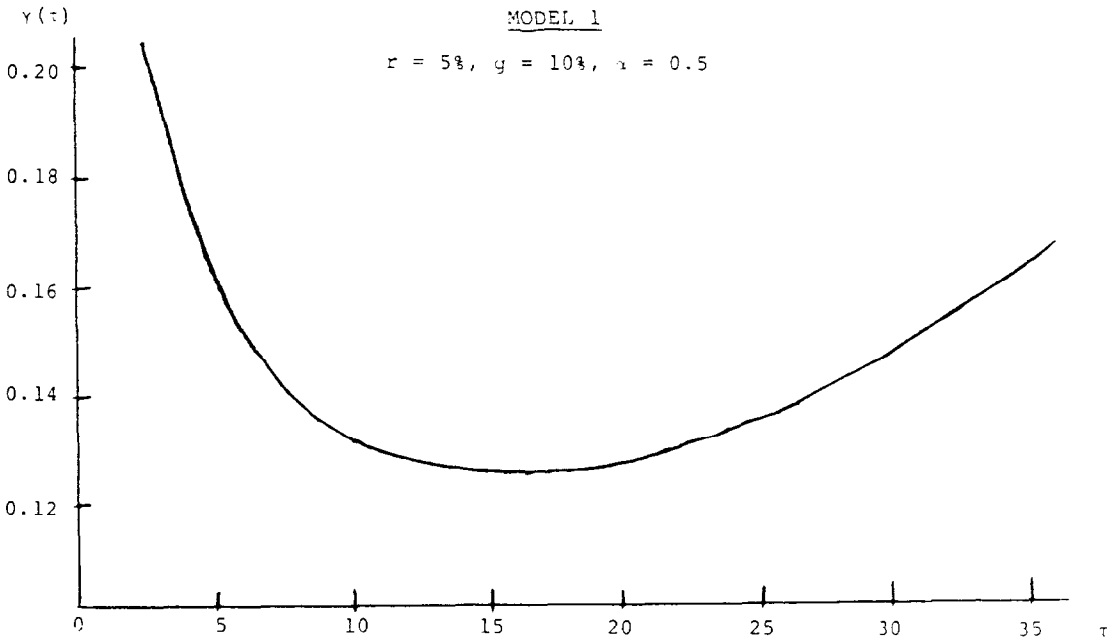
Fig. 1. Equivalent cost rate $\gamma(\tau)$ vs fill period τ for Models 1 and 3.
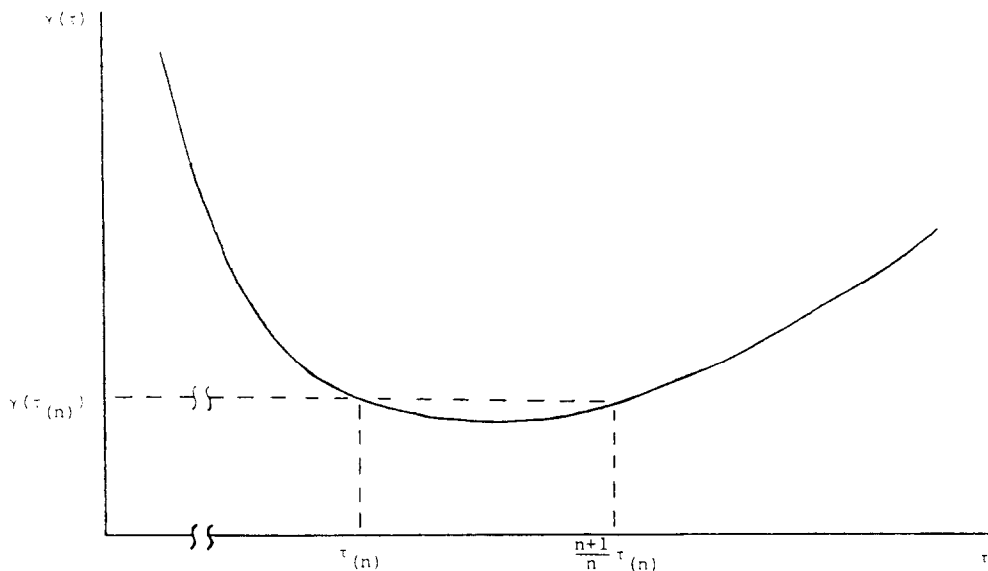
Fig. 2. Graphic discovery of $\tau_{(n)}$.

We are now in a position to state our first important structural result which not only establishes existence of an optimal solution, but gives us *a priori* bounds on the optimal facility sizes.

Theorem 3.5: An optimal finite sequence solution $\langle \tau_1^*, \tau_2^*, \ldots, \tau_n^* \rangle$ exists, and for all horizons T, $\tau_{(1)} \le \tau_i^* \le 2\tau_{(1)}$ for all $i \ne n$.

Proof: We will first establish that we may restrict our search for $\langle \tau_i^* \rangle$ (if it exists) to those fill periods $\tau_{(1)} \le \tau \le 2\tau_{(1)}$.

Consider any policy $\langle \tau_i \rangle$ with $\tau_k > 2\tau_{(1)}$ for some k. Then $\tau_k/2 > \tau_{(1)}$ so that $\gamma(\tau_k/2) < \gamma(\tau_k)$ since γ is strictly increasing for $\tau > 2\tau_{(1)}$. But then we can replace τ_k by a sequence of two facilities filling in time $\tau_k/2$ and achieve a strictly lower cost sequence. Hence we may restrict consideration to $\tau \le 2\tau_{(1)}$. Now suppose $\tau_k < \tau_{(1)}$ for some $k < n \le \infty$. We may without loss of generality assume that $\gamma(\tau_k) \le \gamma(\tau_{k+1})$ since otherwise we may permute the sequence $\langle \tau_i \rangle$ to establish such an ordering of $\gamma(\tau_i)$ and obtain a lower cost sequence. Hence $\tau_{k+1} < \tau_{(1)}$. But then $\tau_k + \tau_{k+1} < 2\tau_{(1)}$ and hence $\gamma(\tau_k + \tau_{k+1}) < \min\{\gamma(\tau_k), \gamma(\tau_{k+1})\}$. Hence we can get a strict decrease in cost by replacing the tandem sequence (τ_k, τ_{k+1}) by a single facility filling in time $\tau_k + \tau_{k+1}$. Hence we may restrict consideration to $\tau_{(1)} \le \tau \le 2\tau_{(1)}$ for all but the last facility (if there is a last). Let $c\langle \tau_i \rangle$ be the discounted cost of the policy $\langle \tau_i \rangle$. Then we have concluded that

$$\inf_{\langle \tau_i \rangle} c\langle \tau_i \rangle = \inf_{\langle \tau_1, \ldots, \tau_n \rangle} c\langle \tau_1, \ldots, \tau_n \rangle \text{ where } n \le N = [T/\tau_{(1)}] < \infty \text{ since } T < \infty.$$ Hence we may restrict our search for a minimum to those policies lying in the compact set $S = \{(\tau_1, \tau_2, \ldots, \tau_N): \sum_{i=1}^{N} \tau_i = T, \tau_i \ge 0\} \subseteq R^N$. Finally, $c\langle \tau_i \rangle$ is a continuous function and hence attains its minimum $\langle \tau_i^* \rangle$ over the compact set S. ∎

We will only remark that if $\nu < 0$ then the lower bound of Theorem 3.5 holds for all facilities i except possibly the *first*. The reason is that the exceptional facility, being smaller than $\tau_{(1)}$, has the largest γ value and therefore must come first by the optimal ordering properties of γ for $\nu < 0$. Skoog[14] has shown existence of an optimal finite horizon policy for a more general problem under an assumption that a minimum fill size exists.

It is clear, that since $\gamma(\tau)$ is minimized at τ^*, the optimal policy is equal period with period τ^* for horizons $T = n\tau^*$ for $n \ge 1$. Also since γ is strictly decreasing for $\tau < \tau^*$ the optimal policy for $T \le \tau^*$ is to deploy exactly one facility of size T. In all other cases, it is not possible to explicitly characterize the optimal policy $\langle \tau_i^* \rangle$. However, we will in the next few results characterize some of the special properties of $\langle \tau_i^* \rangle$ and moreover give insight into how rapidly the finite horizon solution converges to the equal period infinite horizon solution.

Theorem 3.6: $\langle \tau_i^* \rangle$ is monotone in i, i.e. either $\tau_1^* \ge \tau_2^* \ge \ldots \ge \tau_n^*$ or $\tau_1^* \le \tau_2^* \le \ldots \le \tau_n^*$.

Proof: Consider first the case $\tau_1^* < \tau^*$. Suppose $\tau_2^* > \tau_1^*$. Then $\tau_2^* > \tau^*$ since otherwise we would have $\gamma(\tau_2^*) < \gamma(\tau_1^*)$ contradicting the optimal ordering properties of γ. It follows that $\gamma(\tau_1^* + \delta) < \gamma(\tau_1^*)$ and $\gamma(\tau_2^* - \delta) < \gamma(\tau_2^*)$ for $\delta \le \min\{\tau_2^* - \tau^*, \tau^* - \tau_1^*\}$. But then we can achieve a strictly lower cost sequence by replacing τ_1^*, τ_2^* by $\tau_1^* + \delta, \tau_2^* - \delta$. Contradiction. Hence $\tau_2^* \le \tau_1^*$. We can repeat this argument recursively for facility 2 since we now have $\tau_2^* < \tau^*$ to conclude that $\tau_3^* \le \tau_2^*$ and so on. Hence $\tau_1^* \ge \tau_2^* \ge \ldots \ge \tau_n^*$. The argument proceeds similarly for the cases $\tau_1^* = \tau^*$ and $\tau_1^* > \tau^*$ respectively.   ■

The following is our first planning horizon result. It establishes upper and lower bounds for the first facility's fill period that converge to a common value as horizon time lengthens.

Theorem 3.7: $\tau_{(n)} \le \tau_1^* \le ((n+1)/n)\tau_{(n)}$ for all horizon times $T \ge (n+1)\tau_{(n)}$ where $n \ge 1$.

Proof: We shall first establish that $\tau_{(n)} \le \tau_1^* \le ((n+1)/n)\tau_{(n)}$ for $(n+1)\tau_{(n)} \le T \le (n+2)\tau_{(n+1)}$. Then since the bounds are telescoping, i.e. $\tau_{(n)} \le \tau_{(n+1)}$ and $((n+1)/n)\tau_{(n)} \le ((n+2)/(n+1))\tau_{(n+1)}$, the result will follow. So suppose that it is not the case that $\tau_{(n)} \le \tau_1^* \le ((n+1)/n)\tau_{(n)}$. Then $\gamma(\tau_1^*) > \gamma(\tau_{(n)}) = \gamma(((n+1)/n)\tau_{(n)})$. Further by the optimal ordering properties of γ, $\gamma(\tau_i^*) > \gamma(\tau_{(n)})$ for all i. But consider the equal period policy of deploying $n+1$ facilities each of period $T/(n+1)$. We have $\tau_{(n)} \le T/(n+1) \le ((n+2)/(n+1))\tau_{(n+1)} \le ((n+1)/n)\tau_{(n)}$, and hence $\gamma(T/(n+1)) \le \gamma(\tau_{(n)})$. But then this policy yields a strictly lower cost than $\langle\tau_i^*\rangle$. Contradiction.   ■

The horizon times $T = (n+1)\tau_{(n)}$ are the times at which the *a priori* difference between the upper and lower bounds on τ_1^* is greatest. This can be seen by noting the following. Let $\tau_1^{-1}(\gamma)$ and $\tau_2^{-1}(\gamma)$ represent the left and right inverses of the γ function. Then it is clear that for horizons that are integer multiples of $\tau_1^{-1}(\gamma)$ or $\tau_2^{-1}(\gamma)$, we have $\tau_1^{-1}(\gamma) \le \tau_1^* \le \tau_2^{-1}(\gamma)$. One can see from Fig. 3 that the bounds converge to τ^* as T increases from $(n+1)\tau_{(n)}$ to $(n+1)\tau^*$ after which the bounds again diverge until we reach $(n+2)t_{(n+1)}$. These observations are generalizations of those made by Skoog[14] for Model 1 under an assumption that the infinite horizon cost function for $\langle\tau\rangle$ is unimodal.

Numerically solving for the bounds given by Theorem 3.7 requires knowledge of $\tau_{(n)}$. $\tau_{(n)}$ may be approximately found graphically as in Fig. 2, or must be solved numerically for each special case. $\tau_{(n)}$ may not be expressed in closed form for most of the problems of the class (P). The following corollary gives bounds for the planning horizon τ_1^* that are easily computed and explicitly give the rate of convergence of τ_1^* to τ^* as the horizon T increases. Although the bounds are not as tight as those given in Theorem 3.7, they agree for large horizons T.

Corollary 3.8: Let $0 < \rho < 1$. Then $|((\tau_1^* - \tau^*)/\tau^*)| \le \rho$ for all $T \ge (1 + [1/\rho])\tau^*$.
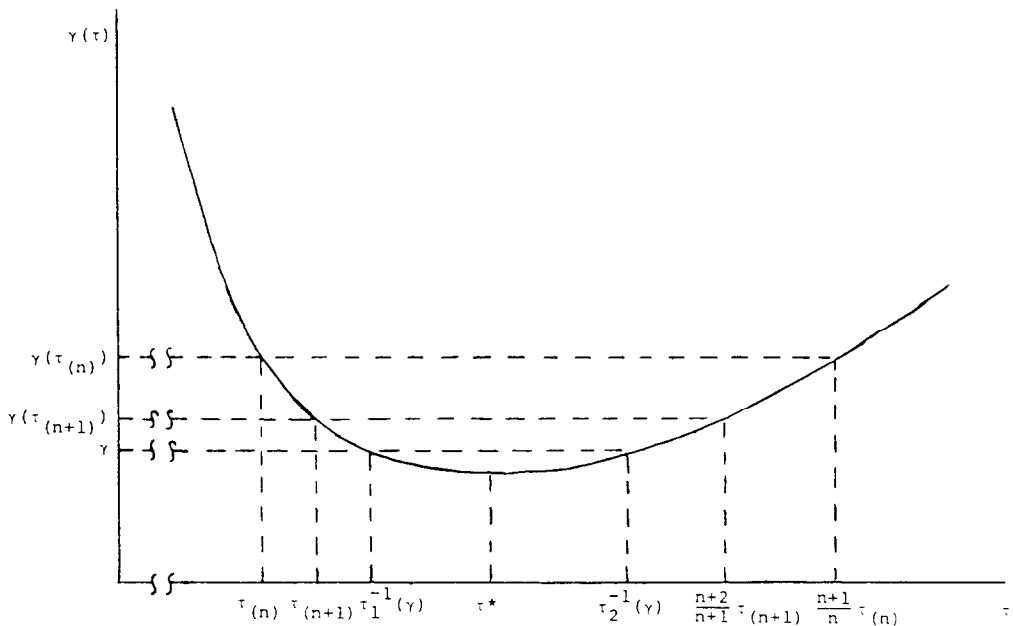

Fig. 3. Graphic discovery of bounds on τ_1^* for $(n+1)\tau_{(n)} \le T \le (n+2)\tau_{(n+1)}$.

Proof: Let $T \geq (1 + [1/\rho])\tau^*$. Then $T \geq (n_\rho + 1)\tau_{(n_\rho)}$ where n_ρ is the smallest $n \geq 1/\rho$ since $\tau_{(n)} \leq \tau^*$ for all n. By Theorem 3.7,

$$\tau_{(n_\rho)} \leq \tau_1^* \leq \frac{n_\rho + 1}{n_\rho} \tau_{(n_\rho)}.$$

But we also have

$$\tau_{(n_\rho)} \leq \tau^* \leq \frac{n_\rho + 1}{n_\rho} \tau_{(n_\rho)}.$$

Hence

$$|\tau_1^* - \tau^*| \leq \frac{1}{n_\rho} \tau_{(n_\rho)} \leq \frac{\tau^*}{n_\rho} \leq \rho\tau^*. \qquad \blacksquare$$

Corollary 3.8 indicates that the length of horizon necessary to yield a first decision within a certain per cent of the optimal long run policy is linear in the optimal infinite horizon planning period. Figure 4 gives per cent agreement between short and long run optimal first decisions for a series of horizon times. Note that the percentage figures are applicable to all problems of the form (P) regardless of parameter values. From Fig. 4 we note for example that a horizon time of at least 21 optimal infinite horizon planning periods is sufficient for *any* problem of the type (P) to yield a first decision within 5% of the long run optimal decision. Naturally, Theorem 3.7 can be used to provide better bounds for specific parameter values.

Although our focus in the last two results has been on the first decision τ_1^*, we may use these results to recursively derive bounds for all other fill periods τ_i^*. For example bounds on τ_2^* may be obtained for a given T by subtracting the upper bounds for τ_1^* from T yielding T'. τ_2^* now becomes the optimal first fill period for a new horizon time bounded from below by T'. Also it follows from Corollary 3.8 that $\lim_{T\to\infty} \tau_i^* = \tau^*$ for all i as we should expect.

4. A FINITE RENEWAL ALGORITHM FOR THE FINITE HORIZON PROBLEM

In the last section we presented results leading to bounds on the optimal fill periods τ_i^*. In this section, we consider an efficient algorithm for approximately solving (P) for all horizon times T within a specified maximal error by solving (P) for a single horizon time T^*.

We know from Theorem 3.5 that regardless of horizon time T, $\tau_{(1)} \leq \tau_i^* \leq 2\tau_{(1)}$ for all facilities i except possibly the last where however the upper bound still applies. Accordingly, we will partition the interval $[\tau_{(1)}, 2\tau_{(1)}]$ into N fill periods and use the resulting discrete model to approximate a solution to the continuous problem (P). Later results will give bounds on the resulting error in optimal cost which will in turn allow determination of N.

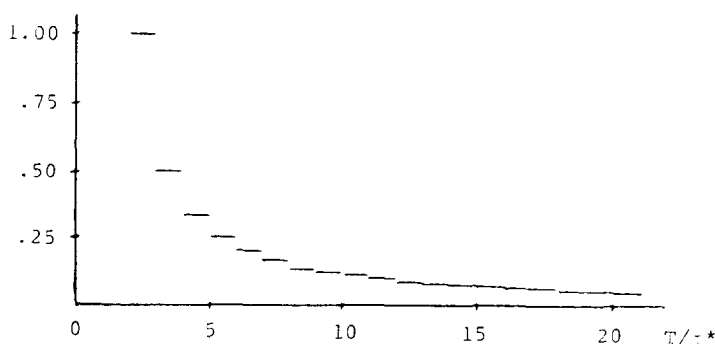*Definition 4.1*: Let

$$\delta = \frac{\tau_{(1)}}{N}$$

Fig. 4. Upper bounds on fractional agreement ρ between short and long run initial fill periods vs horizon time T.

and

$$\theta_l = \tau_{(1)} + l\delta \qquad l = 0, 1, \ldots, N$$

so that

$$\theta_0 = \tau_{(1)}$$

and

$$\theta_N = 2\tau_{(1)}.$$

Let

$$T_\delta = \left[\frac{T}{\delta}\right]\delta.$$

The following discrete time problem will be referred to as P_δ: find $C(T_\delta)$ where

$$(P_\delta) \quad C(t) = \begin{cases} \min\limits_{l=0,1,\ldots,N} \{f(\theta_l) + e^{-\nu\theta_l} C(t - \theta_l)\} & \text{for } t = \delta, 2\delta, \ldots, T_\delta \\ 0 & \text{for } t \leq 0. \end{cases}$$

The discrete time problem (P_δ) is of the same form as that solved by Shapiro and Wagner[13], and we will use a version of their renewal algorithm for its solution. We will later note that this algorithm allows us to solve *all* horizon times to a given accuracy by solving a single finite horizon problem.

A finite renewal algorithm

The problem of solving the dynamic programming recursion for (P_δ), as in all deterministic dynamic programs, is equivalent to finding the shortest path in a network whose nodes correspond to points in time and whose arcs correspond to decisions (facility choices in our case). We solve for the minimum path using a backward algorithm so that node 0 corresponds to time T_δ and node T_δ/δ corresponds to time 0. There are $T_\delta/\delta + 1$ nodes in all numbered 0 through T_δ/δ corresponding to the discrete time points of (P_δ). We find the minimum path to successive nodes following node 0 using a standard dynamic programming recursion with the following exception. Using the optimal ordering properties of the γ functions which also hold for (P_δ), we dynamically delete arcs from the network as necessarily non-optimal as the minimum path algorithm proceeds from node to node. More precisely, there is an arc in the dynamic network from node $k \neq 0$ to node $k' > k$ only if $\theta_l = (k' - k)\delta$ for some l (i.e. there is a facility whose fill period spans the nodes k and k') and $\gamma(\theta_l) \leq \gamma(\theta_{l(k)})$ where $l(k)$ is the index of the last fill period of the optimal path to node k (i.e. θ_l does not violate the optimal ordering dictated by γ). At the start of the algorithm at node $k = 0$, there are arcs from node 0 to node k' for all $k' = \theta_l/\delta$ for some facility l. The length of the arc from node k to node k' is

$$e^{-\nu(T_\delta - k'\delta)} f((k' - k)\delta),$$

i.e. the discounted cost of installing a facility of fill period $\theta_l = (k' - k)\delta$ at time $T_\delta - k'\delta$.

The following theorem gives an upper bound on the error in cost associated with the approximation (P_δ) to (P). We will restrict our attention to the case $\tau_{(1)} \leq \tau_i^* \leq 2\tau_{(1)}$ for all i

Fig. 5. Dynamic network for the discrete problem (P_δ).

because in our extensive numerical work we never encountered a $\theta_i^* \le \tau_{(1)}$ for T greater than τ^*. We also consider only horizon times T that are exact multiples of the (small) number δ.

Theorem 4.2: The fractional error in optimal cost for (P_δ) as compared with (P) is bounded by

$$\epsilon(\delta) = \max_{\tau_{(1)} \le \tau \le 2\tau_{(1)}} \left| \frac{\gamma(\tau + \delta) - \gamma(\tau)}{\gamma(\tau^*)} \right|$$

for all horizons T.

Proof: Let $\langle \tau_1^*, \tau_2^*, \ldots, \tau_n^* \rangle$ be optimal for problem (P) and let Y_i^* be the corresponding installation times.

Consider the following feasible policy for problem (P_δ) derived from $\langle \tau_i^* \rangle$:

$$\theta_i = ([Y_{i+1}^*/\delta] - [Y_i^*/\delta])\delta \text{ for } i = 1, 2, \ldots, n$$

where $Y_{n+1}^* \equiv T$ (see Fig. 6).

Let Θ_i be the corresponding installation times of the sequence $\langle \theta_i \rangle$. Then

$$C\langle \tau_i^* \rangle = \sum_{i=1}^{n} \int_{Y_i^*}^{Y_{i+1}^*} \gamma(\tau_i^*) e^{-\nu t} \, dt$$

and

$$C\langle \theta_i \rangle = \sum_{i=1}^{n} \int_{\Theta_i}^{\Theta_{i+1}} \gamma(\theta_i) e^{-\nu t} \, dt.$$

Hence

$$C\langle \theta_i \rangle - C\langle \tau_i^* \rangle = \sum_{i=1}^{n} \int_{\Theta_i}^{\Theta_{i+1}} \gamma(\theta_i) e^{-\nu t} \, dt - \sum_{i=1}^{n} \int_{\Theta_i}^{Y_{i+1}^*} \gamma(\tau_i^*) e^{-\nu t} \, dt$$

$$- \sum_{i=1}^{n-1} \int_{Y_{i+1}^*}^{\Theta_{i+1}} \gamma(\tau_{i+1}^*) e^{-\nu t} \, dt.$$

But $\gamma(\tau_{i+1}^*) \ge \gamma(\tau_i^*)$ by the optimal ordering property of γ, and thus

$$C\langle \theta_i \rangle - C\langle \tau_i^* \rangle \le \sum_{i=1}^{n} \int_{\Theta_i}^{\Theta_{i+1}} (\gamma(\theta_i)$$

$$- \gamma(\tau_i^*)) e^{-\nu t} dt \le \max_{\tau_{(1)} \le \tau \le 2\tau_{(1)}} |\gamma(\tau + \delta) - \gamma(\tau)| \frac{1 - e^{-\nu T}}{\nu}.$$

Further

$$C\langle \tau_i^* \rangle \le \int_0^T \gamma(\tau^*) e^{-\nu t} dt = \gamma(\tau^*) \frac{1 - e^{-\nu T}}{\nu}.$$

Hence

$$\frac{C\langle \theta_i^* \rangle - C\langle \tau_i^* \rangle}{C\langle \tau_i^* \rangle} \le \frac{C\langle \theta_i \rangle - C\langle \tau_i^* \rangle}{C\langle \tau_i^* \rangle} \le \epsilon(\delta).$$
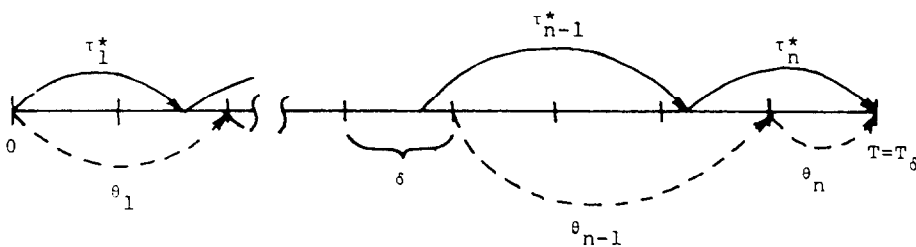

Fig. 6. Deriving a discrete policy close to the optimal continuous policy.

Theorem 4.2 allows us to find a spacing δ so that *all* horizon solutions to (P_δ) will be within an error of $\epsilon(\delta)$ of (P). An application of the turnpike theorem of Shapiro and Wagner[13] states that for large enough horizon times, the first optimal facility $\theta^*_{i_1}$ is always the so-called turnpike policy.

Definition 4.3: Let θ^* be a θ_i such that $\gamma(\theta^*) = \min_{i=1,2,\ldots,n} \gamma(\theta_i)$. If there is more than one, let θ^* be the smallest. θ^* is called the *turnpike policy*.

Theorem 4.4: (Shapiro and Wagner [13]): There exists a threshold horizon time T^*_δ such that $\theta^*_{i_1} = \theta^*$ for all $T_\delta \geq T^*_\delta$.

The finite renewal algorithm can be used to find T^*_δ. We construct the dynamic network successively finding optimal policies to larger horizon problems until the following condition is met. The condition is that the optimal last facility type taking us to node k is the turnpike policy for all k spanning a time interval exceeding the fill time of the largest facility. Then by the optimal ordering properties of γ, all facilities (and in particular the "first" facility) installed beyond the largest of these nodes k^* must be turnpike facilities. Hence we may set $T^*_\delta = k^*$. We know that such a k^* will always exist by Theorem 4.4. Hence in solving the $T = T^*_\delta$ problem, we have solved (P_δ) for *all* T. Furthermore the solution error is bounded by $\epsilon(\delta)$ for all horizons T.

We coded the finite renewal algorithm into a computer program that solved Model 1 for any choice of parameters K, D_0, α, g and r. The program solved the problem for successively longer horizon times until T^*_δ was discovered. We solved 125 different cases where δ was chosen to produce a maximal error of 1%. The resulting computation times averaged 3.8 sec to solve each problem for all horizon times T. Figure 7 gives $\gamma(\tau)$ for one of the cases considered where we set $K = 1$, $D_0 = 1$, $\alpha = 0.7$, $r = 9\%$, $g = 16\%$. As can be seen from the figure, $\tau^* = 13.7$ and $\tau_{(1)} = 9.5$. A 1% upper bound on cost error required a spacing $\delta = 1.35$. It took 4.8 sec to discover $T^*_\delta = 43$ and thus solve for all horizon times within a maximal error of 1%.
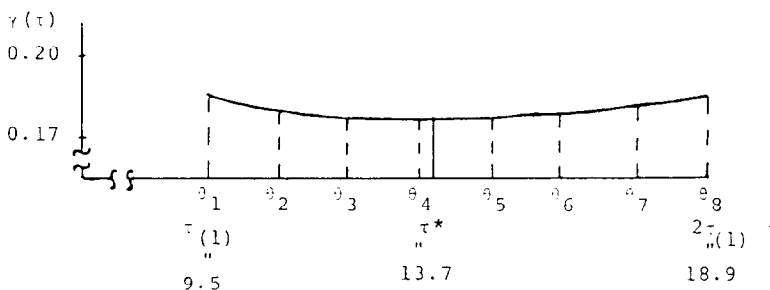

Fig. 7. Discrete fill periods necessary to bound approximation error by 1%.

5. CONCLUSIONS

We have presented planning horizon results that suggest fractional agreement between short run and long run initial policies is eventually inversely proportional to horizon time. Moreover we presented formulas for obtaining specific numerical bounds on this agreement. These can lend considerable insight to the problem of choosing a horizon time long enough to avoid end of study effects in a more realistic context. We also detailed a finite algorithm for approximately solving the deterministic capacity problem for all horizons by solving a single horizon time problem.

REFERENCES

1. D. Anderson, Models for determining least-cost investments in electricity supply, *Bell J. Economics and Management Sci.* 267–299 (1972).
2. D. Erlenkotter, Sequencing Expansion Projects, *Op. Res.* 21, 543–544 (1973).

3. D. Erlenkotter and A. S. Manne, Capacity expansion for India's nitrogenous fertilizer industry. *Management Sci.* 14, 553–572 (1968).
4. J. Freidenfelds, *Capacity Expansion: Analysis of Simple Models with Applications*, forthcoming.
5. R. J. Giglio. A note on the deterministic capacity problem, *Management Sci.* 19, 1096–1099 (1973).
6. J. Jaskold-Gabszewicz and J.-P. Vial, Optimal capacity expansion under growing demand and technological progress. (Edited by G. P. Szego and K. Shell), *Mathematical Methods in Investment and Finance*, pp. 159–189. North-Holland. Amsterdam (1972).
7. A. J. Kalotay. Two comments on the deterministic capacity problem. *Management Sci.* 20, 1313 (1973).
8. A. S. Manne. Capacity expansion and probabilistic growth. *Econometrics* 29 (1961).
9. A. S. Manne (Editor). *Investments for capacity expansion: Size. Location and Time-Phasing.* The MIT Press. Cambridge. Mass. (1967).
10. Ian McDowell, The economical planning period for engineering works, *Op. Res.* 8, 533–542 (1960).
11. T. E. Morton. Forward Algorithms for Forward Thinking Managers, WP 7-78-79. Carnegie-Mellon University, (1979).
12. F. W. Sinden, The replacement and expansion of durable equipment. *SIAM Appl. Math.* 8, 466–480 (1960).
13. J. F. Shapiro and H. M. Wagner. A finite renewal algorithm for the knapsack and turnpike models, *Op. Res.* 15, 318–341 (1967).
14. R. A. Skoog. General Results on Infinite Horizon Expansion Policies, Bell Telephone Laboratories Internal Memorandum, (1973).
15. R. L. Smith, Optimal expansion policies for the deterministic capacity problem, *The Engng Economist* (1980).
16. R. L. Smith. General horizon results for the deterministic capacity problem. *IEEE Int. Conf. Communications*, (1976).
17. R. L. Smith. Turnpike results for single location capacity expansion. *Management Sci.* 25, 474–484 (1979).
18. R. L. Smith. Deferral strategies for a dynamic communications network. *Networks* 9, 61–87 (1979).
19. T. N. Srinivasan, Geometric rate of growth of demand. (Edited by A. S. Manne), pp. 150–156. In *Investments for Capacity Expansion*. MIT Press, 1967.
20. N. Valcoff, Optimal size of transmission systems, *IFAC Symp. Optimal Systems Planning.* (1968).
21. B. Yaged, Minimum cost routing for dynamic network models, *Networks* 3, 193–224 (1973).

APPENDIX

Proof of Lemma 3.3: As noted in the text, we need only establish Assumption 3.2(c) here.

Model. 1: First, it is clear that $\gamma(\tau)$ is continuously differentiable over $(0, \infty)$, and we know from Assumption 3.2(b) that $\gamma'(\tau) < 0$ for small τ and $\gamma'(\tau) > 0$ for large τ. Further since $\gamma'(\tau)$ is continuous it suffices in showing strict unimodality to show that there is at most one (in fact exactly one) zero of $\gamma'(\tau)$ over $(0, \infty)$. Strict unimodality follows from the fact that the sign pattern of $\gamma'(\tau)$ is then $-, 0, +$, where the 0 occurs at exactly one point. We have

$$\gamma'(\tau) = \frac{X(\tau) - Y(\tau)}{Z(\tau)}$$

where

$$X(\tau) = vg(1 - \alpha)e^{g\tau}(e^{g\tau} - 1)^{-\alpha}(1 - e^{-v\tau}), \quad Y(\tau) = v^2 e^{-v\tau}(e^{g\tau} - 1)^{1-\alpha},$$

and

$$Z(\tau) = (1 - e^{-v\tau})^2.$$

Hence a necessary condition for $\gamma'(\tau) = 0$ is that

$$(1 - \alpha)\frac{e^{v\tau} - 1}{v} = \frac{1 - e^{-g\tau}}{g}.$$

Let $v' = -v$. Then we have

$$\lambda = \frac{1 - e^{-g\tau}}{1 - e^{-v'\tau}} \quad \text{where } \lambda = \frac{1}{1 - \dfrac{r}{(1 - \alpha)g}} > 1. \tag{1}$$

A useful lemma is that $\phi(\tau) = (1 - e^{-a\tau})/(1 - e^{-b\tau})$ is strictly increasing for $a < b$. Now $g = v'$ implies $r = -\alpha g < 0$ which contradicts $r > 0$. Hence $g \neq v'$. If $g < v'$ then the r.h.s. of (1) is strictly increasing. Otherwise, the r.h.s. is strictly decreasing. In either case, (1) has at most one solution. Hence $\gamma'(\tau)$ has at most one zero. ∎

Model 2: $\gamma'(\tau) < 0$ for small τ and $\gamma'(\tau) < 0$ for large τ again from Assumption 3.2(b). Again from continuity of $\gamma'(\tau)$, we must only establish that there is at most one zero of $\gamma'(\tau)$. We have $\gamma'(\tau) = 0$ implies

$$\frac{e^{r\tau} - 1}{\tau} = \lambda \quad \text{where } \lambda = \frac{r}{1 - \alpha}. \tag{2}$$

But the l.h.s. of (2) is strictly increasing for $r > 0$ and $0 < \tau < \infty$. Hence (2) has at most one solution. Hence $\gamma'(\tau)$ has at most one zero.

Model 3: Again, we must only establish that there is at most one zero of $\gamma'(\tau)$. We have $\gamma'(\tau) = 0$ implies

$$\frac{e^{d\tau} - 1}{e^{g\tau} - 1} = \lambda \tag{3}$$

where

$$\lambda = \frac{1}{1 + \frac{r}{(1 - \alpha)d}}.$$

But the l.h.s. of (3) is strictly decreasing for $v > d$. Hence (3) has at most one solution. Hence $\gamma'(\tau)$ has at most one zero.
■

Proof of Lemma 3.4: Since γ is continuous and strictly unimodal by Assumption 3.2. there exists "left" and "right" inverse functions τ_1^{-1} and τ_2^{-1} with domains $[\gamma(\tau^*), L_1]$ and $[\gamma(\tau^*), L_2]$ respectively. Let $L = \min\{L_1, L_2\}$ so that τ_1^{-1} and τ_2^{-1} are both defined over $[\gamma(\tau^*), L]$. Since γ is defined over $(0, \infty)$ and γ is unimodal, $\lim_{\gamma \to L_1} \tau_1^{-1}(\gamma) = 0$ and $\lim_{\gamma \to L_2} \tau_2^{-1}(\gamma) = \infty$.

Consider $\tau^{-1} = \tau_2^{-1} - ((n+1)/n)\tau_1^{-1}$ with domain $[\gamma(\tau^*), L]$. Clearly, $\lim_{\gamma \to \gamma(\tau^*)} \tau^{-1}(\gamma) = -\tau^*/n < 0$ and $\lim_{\gamma \to L} \tau^{-1}(\gamma) > 0$. Further τ^{-1} is continuous since $\gamma(\tau)$ is continuous, and $\gamma(\tau)$ can if necessary be restricted to a compact domain.

Hence there exists a $\gamma = \gamma_0$ such that $\tau^{-1}(\gamma_0) = 0$ and hence

$$\tau_2^{-1}(\gamma_0) = \left(\frac{n+1}{n}\right)\tau_1^{-1}(\gamma_0).$$

Let

$$\tau_{(n)} = \tau_1^{-1}(\gamma_0).$$

Finally, $\tau_{(n)}$ is unique since τ^{-1} is strictly increasing due to the strict monotonicity of γ.
■