

Self-Financing of Congestible Facilities in a Growing Economy

by

Richard Arnott* and Marvin Kraus*

June 1995

*Department of Economics, Boston College, Chestnut Hill, MA 02167, USA.

Self-Financing of Congestible Facilities in a Growing Economy^{*, **}

1. Introduction

The basic static *self-financing result* for congestible facilities, due to Mohring¹ and Harwitz (1962) and Strotz (1965), is that if the facility exhibits constant long-run average costs, then the revenue from the optimal toll exactly covers the cost of constructing and operating the optimal capacity. More generally, if the user cost function is homogeneous of degree zero in capacity and usage, then the degree of self-financing at the optimum equals the degree of homogeneity of the capacity cost function.²

These self-financing results are useful since they indicate for a first-best environment, the proportion of a congestible facility's capacity costs that should be covered from user fees/tolls and the proportion that should be financed out of general revenues. The results are receiving increased attention in policy circles³ (especially transportation policy – highways, public transit, airports). In the current fiscal climate, policy makers find the results attractive since they typically justify an increase in cost recovery over current levels. Policy economists too find them attractive since they provide a politically appealing argument in favor of congestion pricing.

The self-financing results were derived and promulgated on the basis of static models. Remarkably, no one seems to have enquired whether they carry over to intertemporal settings.⁴ The aim of this paper is to address this oversight. We examine the form of the self-financing results in a variety of intertemporal economic environments – continuous additions to capacity with and without adjustment costs, operating costs, depreciation, and irreversibility; and intermittent additions to capacity, where the intermittence may be due to fixed costs of capacity additions, increasing returns to the scale of capacity additions, natural discreteness of capacity increments (e.g., airport runways), or the technology of the planning process.

We find, with one exception which is readily explainable, that the self-financing results *do* extend to intertemporal environments *in present value terms*. We also explain why the results extend.

Section 2 briefly reviews previous work on the self-financing of congestible facilities in a

static environment. Section 3 examines self-financing results in intertemporal environments with *continuous* additions to capacity, and Section 4 treats *intermittent* capacity additions. Section 5 provides a brief *ex post* literature review. And Section 6 concludes.

2. Review of Static Self-Financing Results

To set the stage, we provide brief geometric and algebraic derivations of the self-financing results in a static context. To simplify notation, we assume throughout the paper that individuals are identical. The results can be extended to treat heterogeneous users by combining the analysis of this paper with that of Arnott and Kraus (1995).

Let p denote the price of using the congestible facility (swimming pool, freeway, etc.), which includes both the user cost and any toll, $N(p)$ the demand function, $c(N, K)$ the user cost function ($c_N > 0$, $c_K < 0$, with subscripts denoting partial derivatives) where K is capacity and $c_N > 0$ captures congestion, and $F(K)$ the construction (and operating) cost function. The social surplus associated with the congestible facility equals consumers' surplus plus toll revenue less construction costs, i.e.

$$S = \int_p^a N(p') dp' + (p - c(N(p), K))N(p) - F(K) \quad (1)$$

Maximizing with respect to p and K gives⁵

$$p: (p - (c + Nc_N))N_p = 0 \quad (2a)$$

$$K: -c_K N - F' = 0 \quad (2b)$$

Equation (2a) states that price should equal marginal (social) cost, the sum of the user cost plus the marginal congestion externality (Nc_N); equation (2b) that optimal capacity is such that the marginal benefit of capacity in terms of reduced congestion equals marginal cost.

Now, assume that $\phi(N, K) = c(N, K)N + F(K)$, the total cost function, is homogeneous of degree h in N and K . Then, by Euler's Theorem,

$$cN + c_N N^2 + c_K KN + F'K = h\phi \quad (3)$$

Multiply equations (2a) and (2b) by N and K , respectively, add the resulting equations, and

substitute (3). This yields

$$R^* = \tau^* N^* = (p^* - c^*) N^* = h\phi(N^*, K^*) - c^* N^* \quad (4)$$

where $*$'s denote evaluation at the optimum, τ^* is the optimal toll (equal to the marginal congestion externality at the optimum), and R^* is the revenue raised. When $h = 1$, there are constant long-run average costs and $R^* = F(K^*)$.

We illustrate the result graphically in Figure 1 for the case of decreasing long-run average cost. The optimum occurs where the demand curve intersects the long-run marginal cost curve, (N^*, p^*) . Associated with this is optimal capacity, K^* , and a corresponding set of short-run cost curves (UC denotes user cost). Toll revenue equals $N^* \tau^* = N^*(SRMC(N^*, K^*) - UC(N^*, K^*)) = ABFG$, while construction costs equal total costs less total user costs $= N^*(SRAC(N^*, K^*) - UC(N^*, K^*)) = ACEG$.

In many contexts, it is reasonable to assume that the user cost function is homogeneous of degree zero in N and K . Under this assumption, (4) reduces to

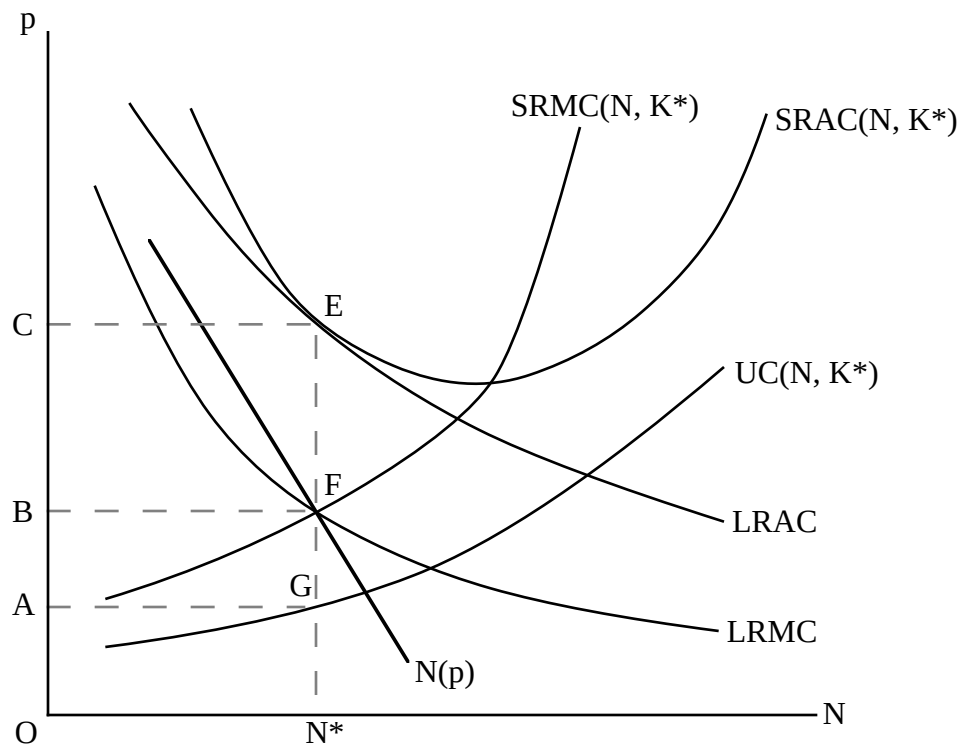
$$R^* = \hat{h} F(K^*), \quad (4')$$

where $\hat{h}$ is the degree of homogeneity of the capacity cost function; this is the result cited in the introduction.

Throughout the rest of the paper, we assume that the user cost function is homogeneous of degree zero in N and K and that the capacity cost function or its analog is homogeneous of degree one, and investigate under what circumstances the self-financing result holds in intertemporal environments. More specifically, we ask: Under these assumptions, with first-best pricing and the optimal program of additions to capacity, when will the present value of toll revenues equal the present value of facility construction, operation and planning costs? In subsequent work, we hope to present a theorem which provides a general answer to this question. This paper makes a more modest contribution by addressing this question in the context of a series of specific models.

The paper focuses on exact cost recovery in the first-best optimum because it is the central case of theoretical interest.⁶ The analysis can be extended to determine the extent of cost recovery in other circumstances.

Figure 1. The self-financing result with decreasing long-run average cost



3. Continuous Additions to Capacity

This section investigates a series of models in which the capacity expansion technology is specified such that capacity is adjusted continuously over time. The next section considers an alternative series of models in which the capacity expansion technology is such that capacity is increased intermittently or in increments of fixed size. By proceeding in this manner, we are sidestepping an important aspect of the problem – whether capacity is added continuously or intermittently should be derived rather than assumed.⁷

3.1 The simplest case

We assume that the economy is continually growing over time, that there are constant costs to adding capacity (no adjustment costs), and that the interest rate, the cost of capital, and the user cost function are fixed over time.⁸ We also ignore depreciation, maintenance, and operating costs. Subsequently, we shall relax these simplifying assumptions.

We employ the following notation: $p(t)$ – price at time t , $N(p, t)$ – demand at time t , $K(t)$ – capacity at time t , $c(N, K)$ – user cost, $I(t)$ – real investment (= addition to capacity at time t), r – interest rate, ρ – cost of a unit of investment.

Discounted social surplus is

$$S = \int_0^{\infty} \int_{p(t)}^{\infty} N(p', t) e^{-rt} dp' dt - \int_0^{\infty} \rho I(t) e^{-rt} dt - \rho K(0) + \int_0^{\infty} (p(t) - c(N(p(t), t), K(t))) N(p(t), t) e^{-rt} dt \quad (5)$$

We assume here and in subsequent models that the growth of the economy is such that S is bounded from above. The planner chooses an initial capacity $K(0)$ and investment and price trajectories $I(\cdot)$ and $p(\cdot)$ so as to maximize (5) subject to the state equation $\dot{K}(t) = I(t)$. The associated current-valued Hamiltonian is

$$H = \int_p^{\infty} N dp' - \rho I + (p - c)N + fI \quad (6)$$

where time variables have been dropped to simplify notation, and $f(t)$, the costate variable, is interpreted as the social value of an increment to capacity at time t . The corresponding first-order conditions are:

$$p: (p - (c + Nc_N))N_p = 0 \quad (7a)$$

$$I: -\rho + f = 0 \quad (7b)$$

$$f: \dot{f} - rf - Nc_K = 0 \quad (7c)$$

$$K(0): -\rho + f(0) = 0, \quad (7d)$$

and the infinite-horizon transversality condition⁹

$$\lim_{t \rightarrow \infty} f(t)e^{-\rho t} = 0 \quad (7e)$$

Combining (7a)-(7c):

$$\begin{aligned} R^* &= \tau^* N^* = (N^*)^2 c_N^* && (\text{since } \tau^* = N^* c_N^*) \\ &= -N^* K^* c_K^* && (\text{using homogeneity of degree zero of } c(\cdot)) \\ &= (-\dot{f}^* + rf^*)K^* && (\text{using (7c)}) \\ R^*(t) &= \rho K^*(t) && (\text{using (7b)}); \end{aligned} \quad (8)$$

at each point in time toll revenues equal amortized construction costs. Integration of (8) by parts gives

$$\begin{aligned} \int_0^{\infty} R^* e^{-\rho t} dt &= \int_0^{\infty} \rho K^* e^{-\rho t} dt \\ &= -\rho K^* e^{-\rho t} \Big|_0^{\infty} + \int_0^{\infty} \rho I^* e^{-\rho t} dt \\ &= \rho K^*(0) + \int_0^{\infty} \rho I^* e^{-\rho t} dt \quad (\text{see footnote 10}) \end{aligned} \quad (9)$$

which indicates that discounted toll revenues equal discounted investment expenditures plus the initial cost of capacity – *the self-financing result holds*.

We now provide an explanation of this self-financing result, which suggests its generality. One can regard the planner as producing a continuum of dated products – users or uses of the

facility at time t . These outputs are produced employing a continuum of dated inputs – investment at time t and users of the facility at time t . It may seem strange that "users at time t " are both an output and an input, but that is the nature of congestion. Associated with this production function is a cost function. A proportional change in inputs results in the same proportional change in outputs and in costs; construction costs increase proportionally by assumption, and aggregate user costs at t increase proportionally since they are homogeneous of degree one in $N(t)$ and $K(t)$. Since the proportional change in inputs is continuously variable, the multi-product cost function exhibits constant long-run ray average costs. And it is well-known (Baumol, Panzar and Willig (1982)) that marginal-cost pricing with constant long-run ray average costs results in break-even operation, which in this context is equivalent to self-financing.

It is sometimes argued that one of the virtues of optimal congestion pricing under constant returns to scale is that if toll revenues are employed to finance additional capacity, optimal capacity is automatically ensured. The argument is typically made for the situation where demand is uncertain, which is beyond the scope of the current paper, but we can check its validity here for the certainty case. The argument has two interpretations. The first is that at each point in time, *investment* should equal toll revenues; the second is that *amortized construction costs* should equal toll revenues. It follows from (8) that the latter interpretation is correct, and not the former. The incorrectness of the former rule is easily illustrated. Suppose that demand increases, and then remains flat for a period. Over the period that demand remains flat, toll revenue would be collected but capacity should not be expanded. Thus, the argument that, with optimal congestion pricing under constant returns to scale, investment of toll revenues to finance additional capacity ensures optimal capacity is correct *if construction costs are amortized*, but not otherwise. Thus, an efficient decentralization mechanism (ignoring incentive considerations) for a congestible facility entails instructing the facility operator to congestion price optimally, to borrow to finance capacity expansion, and to expand at that rate for which zero surplus is obtained at each point in time. The planner does not need to know the capacity cost function or the evolution of demand for this decentralization mechanism to work.

The intuition provided above for the self-financing result centered on the constancy of long-run ray average costs. This suggests that the self-financing result should hold for extensions which do not violate this condition. The rest of the section examines this intuition for several extensions.

3.2 Depreciation and maintenance

We relax some of the simplifying assumptions of the previous subsection's model to incorporate depreciation and maintenance, and to allow for time variation in the interest rate, etc.¹¹ Time variation is straightforward: Let $\Gamma(t) = \int_0^t r(t')dt'$ be the cumulative interest rate, $\rho(t)$ be the price of a unit of investment, and $c(N, K, t)$ be the user cost function. The treatment of depreciation and maintenance is less straightforward. There might be vintage effects; depreciation might depend on the age-profile of the capital stock; there might be a quality dimension to capital, etc. Rather than treat all these complications, we consider only a simple specification in which

$$\dot{K}(t) = g(K(t), m(t), t) + I(t) \quad (10)$$

where $g(\cdot)$ is the depreciation function (with $g_m > 0$ and $g_{mm} < 0$) and $m(t)$ is maintenance (operating expenses at time t).

Discounted social surplus is

$$\begin{aligned} S = & \int_0^a \int_{p(t)}^a N(p', t) e^{-\Gamma(t)} dp' dt - \int_0^a (\rho(t)I(t) + m(t)) e^{-\Gamma(t)} dt - \rho K(0) \\ & + \int_0^a (p(t) - c(N(p(t), t), K(t), t)) N(p(t), t) e^{-\Gamma(t)} dt \end{aligned} \quad (11)$$

The planner chooses an initial capacity $K(0)$ and investment, maintenance and price trajectories $I(\cdot)$, $m(\cdot)$ and $p(\cdot)$ so as to maximize (11) subject to the state equation (10). The associated current-valued Hamiltonian is

$$H = \int_p^a N dp' - \rho I - m + (p - c)N + f(g + I) \quad (12)$$

The corresponding first-order conditions are:

$$p: (p - (c + Nc_N))N_p = 0 \quad (13a)$$

$$I: -\rho + f = 0 \quad (13b)$$

$$f: \dot{f} - rf + fg_K - Nc_K = 0 \quad (13c)$$

$$m: fg_m - 1 = 0 \quad (13d)$$

$$K(0): -\rho + f(0) = 0 \quad (13e)$$

And

$$\begin{aligned} R^* &= \tau^* N^* = (N^*)^2 c_N^* \\ &= -N^* K^* c_K^* \\ &= -\dot{\rho} K^* + r \rho K^* - \rho g_K^* K^* \quad (\text{using (13b) and (13c)}) \end{aligned} \quad (14a)$$

If, furthermore, $g(\cdot)$ is homogeneous of degree one in K and m , i.e. the maintenance/depreciation technology exhibits constant returns, so that $\rho g_K K = \rho g - \rho g_m m = \rho g - m$ (using (13b) and (13d)), then (14a) reduces to

$$R^* = -\dot{\rho} K^* + r \rho K^* - \rho g^* + m^* \quad (14b)$$

Toll revenues plus capital gains equal the interest cost on the capital plus depreciation costs plus maintenance expenditures. Integration of (14b) by parts gives

$$\begin{aligned} \int_0^a R^* e^{-\Gamma} dt &= -\rho K^* e^{-\Gamma} \Big|_0^a + \int_0^a \rho (\dot{K}^* e^{-\Gamma} - r K^* e^{-\Gamma}) dt \\ &\quad + \int_0^a (r \rho K^* - \rho g^*) e^{-\Gamma} dt + \int_0^a m^* e^{-\Gamma} dt \\ &= \rho K^*(0) + \int_0^a \rho I^* e^{-\Gamma} dt + \int_0^a m^* e^{-\Gamma} dt \end{aligned} \quad (15)$$

Discounted toll revenues equal the initial cost of capacity plus discounted investment and maintenance expenditures; again, the self-financing result holds.

The assumption that the maintenance/depreciation technology exhibits constant returns was crucial in the above derivation. This is consistent with our argument that the essential condition for

the self-financing result to hold is that there be constant long-run ray average costs. Accordingly, we conjecture that the self-financing result of this subsection holds for any maintenance/depreciation technology which exhibits constant returns.

In the previous subsection, we discussed a decentralization mechanism which entailed the facility operator using optimal toll revenues to pay interest on capital costs, and showed that capacity expansion will proceed optimally if the operator expands at a rate such that she has zero cash flow surplus. The same decentralization mechanism works here, except that revenues include user fees and capital gains, and costs include debt charges, depreciation and maintenance/operating costs.

3.3 Irreversibility and adjustment costs

We now extend the simplest case of Section 3.1 to treat irreversibility¹² and adjustment costs. Irreversibility is an extreme case of adjustment costs. Even though it is empirically uninteresting since almost all capital equipment has some salvage value, we treat it since it has been the subject of attention in the literature.

We continue to measure investment in physical terms so that

$$\dot{K}(t) = I(t) \quad (16a)$$

And we capture adjustment costs via the function

$$Z(I, K), \quad Z_I > 0, Z_{II} > 0 \quad (16b)$$

This specification incorporates convex adjustment costs, as well as the higher costs of disinvestment than of investment¹³ (viz. \$x obtained through disinvestment causes K to fall by more than \$x of investment costs causes K to rise). To allow for the possibility that investment is irreversible, we append the constraint $I(t) \geq 0$, and consider both the situation where it applies and where it does not. We also assume that initial capacity is installed at constant cost.¹⁴

The corresponding social surplus function is

$$S = \int_0^a \int_0^a N e^{-rt} dp' dt - \int_0^a Z(I, K) e^{-rt} dt - \rho K(0) + \int_0^a (p - c) N e^{-rt} dt \quad (17)$$

The planner maximizes (17) subject to (16a) and the irreversibility constraint, $I(t) \geq 0$. The associated current-valued Hamiltonian is

$$H = \int_0^{\infty} e^{-\rho t} [Ndp' - Z(I, K) + (p - c)N + (f + \mu)I] dt \quad (18)$$

where μ is the shadow price on the irreversibility constraint. The first-order conditions are:

$$p: (p - (c + Nc_N))N_p = 0 \quad (19a)$$

$$I: -Z_I + f + \mu = 0 \quad (19b)$$

$$f: \dot{f} - rf - Z_K - Nc_K = 0 \quad (19c)$$

$$K(0): -\rho + f(0) = 0 \quad (19d)$$

Thus,

$$\begin{aligned} R^* &= \tau^* N^* = (N^*)^2 c_N^* \\ &= -N^* K^* c_K^* \\ &= -\dot{f}^* K^* + rf^* K^* + Z_K^* K^* \quad (\text{using (19c)}) \end{aligned} \quad (20a)$$

Now assume that the adjustment cost technology exhibits constant returns in K and I , so that a proportional increase in both results in the same proportional increase in Z . Then $Z_K K = Z - Z_I I$, and (20a) becomes

$$\begin{aligned} R^* &= -\dot{f}^* K^* + rf^* K^* + Z^* - Z_I^* I^* \\ &= -\dot{f}^* K^* + rf^* K^* + Z^* - (f^* + \mu^*) I^* \quad (\text{using (19b)}) \\ &= -\dot{f}^* K^* + rf^* K^* + Z^* - f^* I^* \quad (\text{since } \mu^* I^* = 0), \end{aligned} \quad (20b)$$

which states that shadow profits from operating the facility optimally are zero; shadow revenues equal toll revenues plus capital gains plus the shadow profit from investment ($fI - Z$), while shadow costs equal the opportunity cost of capital. Note that (20b) holds with or without irreversibility.

Integrating (20b) by parts yields

$$\begin{aligned}
\int_0^a R^* e^{-rt} dt &= -f^* K^* e^{-rt} \Big|_0^a + \int_0^a (f^* \dot{K}^* - r f^* K^*) e^{-rt} dt \\
&\quad + \int_0^a (r f^* K^* + Z^* - f^* I^*) e^{-rt} dt \\
&= \rho K^*(0) + \int_0^a Z^* e^{-rt} dt
\end{aligned} \tag{21}$$

Thus, again, discounted toll revenues equal discounted investment expenditures plus the cost of initial capacity.

The assumption that the adjustment cost technology exhibits constant returns in I and K was crucial to the self-financing result, which reinforces our earlier argument that exact cost recovery hinges on constant long-run ray average costs. The decentralization procedure with adjustment costs is similar to the decentralization procedures derived earlier, but is complicated by the fact that the facility operator has to continually revalue capital at its shadow price, f , which presumably would be quoted to her by the planner.

4. *Intermittent Capacity Additions*

When capacity additions are intermittent, it is important to incorporate into the model the cause of intermittence. One possibility is fixed costs in construction and planning, another is technological or planning increasing returns to capacity additions, another is costs to haste in planning, another is natural discreteness to capacity increments such as airport runways, and yet another is a planning cycle. We begin with the simplest case to illustrate basic ideas. We then consider various extensions.

4.1 *The simplest case*

We suppose that only a single installation of capacity is possible. Perhaps the facility is on a unique site. Once it has been put in place, it cannot be added to, nor can it be demolished with another facility taking its place. A public monument is an example. Let K denote facility size, ρ the cost per unit of facility size, and T construction date. Social surplus is

$$S = \int_T^a \int_{p(t)}^a N(p', t) e^{-rt} dp' dt + \int_T^a (p(t) - c(N(p(t), t), K)) N(p(t), t) e^{-rt} dt - \rho K e^{-rT} \quad (22)$$

The first-order conditions are:

$$p(t): \quad \int_T^a N_p(p - (c + N_{cN})) e^{-rt} dt = 0 \quad (23a)$$

$$K: \quad - \int_T^a N_{cK} e^{-rt} dt - \rho e^{-rT} = 0 \quad (23b)$$

$$T: \quad \left(- \int_{p(T)}^a N(p', T) dp' - ((p - c)N) \Big|_T + \rho K \right) e^{-rT} = 0 \quad (23c)$$

Equation (23a) is familiar. Equation (23b) states that capacity should be such that the reduction in the present value of user costs from expanding capacity by one unit equal the cost of the extra unit. Equation (23c) is an optimal timing condition. Capacity should be installed when the marginal benefit from postponing installation equals the marginal cost. Here the marginal benefit from postponing installation equals the opportunity cost of construction expenditure, while the marginal cost equals the foregone consumers' surplus plus the foregone toll revenue. We have

$$\begin{aligned} \int_T^a R^*(t) e^{-rt} dt &= \int_T^a N^* \tau^* e^{-rt} dt \\ &= \int_T^a (N^*)^2 c_N^* e^{-rt} dt \\ &= - \int_T^a N^* K^* c_K^* e^{-rt} dt \\ &= \rho K^* e^{-rT} \quad (\text{using (23b)}) \end{aligned} \quad (24)$$

Thus, discounted toll revenues equal discounted construction costs – the self-financing result holds. The assumption of constant costs to capital was crucial to the result. Note that (24) was derived without using (23c). Thus, *the self-financing result holds whether or not construction*

occurs at the optimal time. The explanation is that, with T fixed at any value, there are constant long-run ray average costs.

4.2 Capacity added at fixed time intervals

This case corresponds to a situation where the planner reviews the level of capacity every so many years – a planning cycle. Such a policy is often pursued with respect to maintenance and upgrading of roads in a network. There are constant costs to investment. Social surplus is

$$S = \int_{i=1}^a \int_{T_i}^{T_{i+1}} \frac{\partial N(p', t)}{\partial p(t)} N(p', t) e^{-rt} dp' dt - \int_{i=1}^a \rho I_i e^{-rT_i} + \int_{i=1}^a \int_{T_i}^{T_{i+1}} (p(t) - c(N(p(t), t), K_i)) N(p(t), t) e^{-rt} dt \quad (25)$$

where T_i is the fixed time the i th capacity increment is added, I_i is the size of the i th capacity increment, and $K_i = \sum_{j=1}^i I_j$. After substitution for K_i , the first-order conditions are:

$$p(t): (p - (c_i + N \frac{\partial c_i}{\partial N})) N_p e^{-rt} = 0 \quad (26a)$$

$$I_i: - \int_{j=i}^a \int_{T_j}^{T_{j+1}} \frac{\partial c_j}{\partial K_i} N e^{-rt} dt - \rho e^{-rT_i} = 0 \quad (26b)$$

Equation (26b) states that the size of the i th capacity increment (holding the size of future capacity increments fixed) should be such that the marginal cost equal the discounted marginal benefit, in terms of reduced user costs, from increasing capacity by one unit forever. Subtracting (26b) for I_{i+1} from that for I_i gives

$$\int_{T_i}^{T_{i+1}} \frac{\partial c_i}{\partial K_i} N e^{-rt} dt + \rho (e^{-rT_i} - e^{-rT_{i+1}}) = 0 \quad (27)$$

This corresponds to a perturbation in which capacity is increased by one unit from T_i to T_{i+1} , and

held at its original level thereafter. The benefit equals the discounted reduction in user costs over that interval, while the cost equals the increase in discounted construction costs from adding a unit of capacity at T_i rather than at T_{i+1} . Then

$$\begin{aligned}
R_i^* &= \int_{T_i}^{T_{i+1}} \tau^* N^* e^{-r(t-T_i)} dt \\
&= \int_{T_i}^{T_{i+1}} (N^2 \frac{\partial c_i}{\partial N})^* e^{-r(t-T_i)} dt \\
&= \int_{T_i}^{T_{i+1}} (-NK_i \frac{\partial c_i}{\partial K_i})^* e^{-r(t-T_i)} dt \\
&= \rho K_i^* (1 - e^{-r(T_{i+1}-T_i)}) \quad (\text{using (27)}) \tag{28}
\end{aligned}$$

where R_i^* is toll revenue obtained between the i th and $(i+1)$ st capacity additions, discounted to T_i . Thus, over each construction interval, discounted toll revenues equal discounted amortized construction costs. Summing over intervals yields

$$\begin{aligned}
\int_{i=1}^a R_i^* e^{-rT_i} &= \int_{i=2}^a \rho (K_i^* - K_{i-1}^*) e^{-rT_i} + \rho I_1^* e^{-rT_1} \\
&= \int_{i=1}^a \rho I_i^* e^{-rT_i} \quad (\text{since } I_i^* = K_i^* - K_{i-1}^*, i = 2, \dots, a) \tag{29}
\end{aligned}$$

– the self-financing result holds.

In the previous section, with continuous additions to capacity, we considered an efficient decentralization mechanism for capacity expansion. At every point in time, the facility operator collects the revenue from the optimal toll, and expands capacity at such a rate that the toll revenues exactly cover amortized capacity costs. We now consider how this decentralization procedure needs to be modified when capacity additions are intermittent.

The answer is suggested by (28). If capacity expansions are of optimal size, over each

construction interval, discounted optimal toll revenues equal discounted amortized construction costs. Hence, if the facility operator were instructed to finance capacity expansion through debt and were to retain toll revenues, she would find herself with a zero balance at the end of a construction interval. Furthermore, if demand grows continually, the cumulative balance is negative throughout each interval, except at the endpoints. Under this condition, a capacity expansion should occur when and only when the cumulative balance rises to zero. The optimal size of the expansion depends on the future time path of demand, about which the planner is assumed to be informed, but not the operator. Thus, an efficient decentralization procedure entails the planner saying to the operator: You collect the toll revenue and pay the amortized costs of capacity. When your cumulative balance rises to zero, call me, and I shall decide the size of the next capacity expansion. This decentralization mechanism works but since the planner is assumed to have perfect information he can do just as well controlling investment directly. The analogous mechanism under uncertainty is, however, nontrivial. Under uncertainty, it is reasonable to assume that the planner but not the operator knows the stochastic process generating demand while the operator but not the planner observes the realization of demand. Then the planner would choose the size of capacity expansion that would, on average, be appropriate for say five years. And he would initiate the next capacity expansion when the operator's cumulative surplus has risen to zero.

In the context of airport financing, the argument has been made that, with constant long-run average costs, investment of toll revenues (net of costs) in capacity expansion ensures optimal capacity. The argument is at best imprecise. One possible interpretation of this statement is that the facility is financed on a cash-as-you-go basis, and that surplus is channelled to finance capacity expansion. This is obviously incorrect. For one thing, the initial capacity increment cannot be financed on a cash-as-you-go basis; for another, the rule would result in capacity expansion when demand is stationary. A second interpretation is that capacity increments are financed by debt, and that expansions should occur when enough of a surplus has accumulated. This is too conservative a policy. A capacity expansion should occur whenever the cumulative surplus rises to zero.

4.3 Capacity replacement

Suppose that there is only one site at which a bridge can be placed across a river. Periodically, the bridge is torn down, and a new bridge with a larger capacity is put up in its place. Social surplus is

$$S = \int_{i=1}^a \int_{T_i}^{T_{i+1}} \frac{\partial N(p', t)}{\partial p} N(p', t) e^{-rt} dp' dt - \int_{i=1}^a \rho K_i e^{-rT_i} + \int_{i=1}^a \int_{T_i}^{T_{i+1}} (p(t) - c(N(p(t), t), K_i)) N(p(t), t) e^{-rt} dt \quad (30)$$

The timing of capacity replacements can be either endogenous or exogenous. When the timing is exogenous, the first-order conditions are:

$$p(t): (p - (c_i + N \frac{\partial c_i}{\partial N})) N_p e^{-rt} = 0 \quad (31a)$$

$$K_i: - \int_{T_i}^{T_{i+1}} \frac{\partial c_i}{\partial K_i} N e^{-rt} dt - \rho e^{-rT_i} = 0 \quad (31b)$$

Then

$$\begin{aligned} R_i^* &= \int_{T_i}^{T_{i+1}} \tau^* N^* e^{-r(t - T_i)} dt \\ &= \int_{T_i}^{T_{i+1}} (N^2 \frac{\partial c_i}{\partial N})^* e^{-r(t - T_i)} dt \\ &= \rho K_i^* \quad (\text{using (31b) and homogeneity of degree zero of } c(\cdot)) \end{aligned} \quad (32)$$

where R_i^* is the toll revenue obtained from the i th facility, discounted to the time of its construction. Thus, the self-financing result holds for each of the succession of facilities. The result hinges on constant costs to capacity, which ensures constant long-run ray average costs.

There is a nice decentralization rule here – replace when the facility has paid for itself. It would have to be the planner, though, who decides on the size of each capacity replacement.

4.4 *A more general capacity expansion technology*

A commentator on the paper argued that our analysis of the situations considered in the previous subsections was conceptually flawed since, he claimed, those situations could arise only if there were decreasing costs with respect to the size of capacity additions, which our analysis did not consider. We disagree. With respect to one-shot capacity installation, for example, one can conceive of a putty-diamond technology where there are constant costs to putting the putty in place, but once the putty has set it is immutable and indestructible. We concede that, in reality, when capacity is added intermittently, the technology is almost always characterized by decreasing costs with respect to the scale of capacity additions. But our intention in this paper is to examine whether the self-financing result holds for a wide range of constant cost technologies. For any constant cost technology for which the result holds, less than full cost recovery will obtain for the corresponding decreasing cost technology.

We now provide a more general treatment of the capacity expansion technology, which admits fixed costs and increasing returns to the size of capacity additions. We are interested in determining whether there are any specifications of the technology which simultaneously give rise to intermittent capacity and result in self-financing.

We define I_i and K_i as before, such that $K_i = \sum_{j=1}^i I_j$. And we describe the technology via the cost function for a capacity addition, $A_i = A(I_i, K_i, T_i - T_{i-1})$, $\partial A_i / \partial I_i > 0$, $\partial A_i / \partial K_i \geq 0$, $\partial A_i / \partial (T_i - T_{i-1}) \leq 0$. The idea behind the dependence of costs on $T_i - T_{i-1}$ is that costs are higher if planning for the capacity addition is rushed; thus, $\partial A_i / \partial (T_i - T_{i-1}) < 0$ incorporates the costs of haste.

Social surplus is

$$\begin{aligned}
S = & \int_{i=1}^a \int_{T_i}^{T_{i+1}} \frac{\partial A_i}{\partial p(t)} N(p', t) e^{-rt} dp' dt - \int_{i=1}^a A(I_i, K_i, T_i - T_{i-1}) e^{-rT_i} \\
& + \int_{i=1}^a \int_{T_i}^{T_{i+1}} (p(t) - c(N(p(t), t), K_i)) N(p(t), t) e^{-rt} dt
\end{aligned} \tag{33}$$

Where ϕ_i is the current-valued Lagrange multiplier on the constraint $\sum_{j=1}^i I_j - K_i \leq 0$, the first-order conditions are:

$$p(t): \quad (p - (c_i + N \frac{\partial c_i}{\partial N})) N_p e^{-rt} = 0 \tag{34a}$$

$$I_i: \quad - \frac{\partial A_i}{\partial I_i} e^{-rT_i} + \int_{j=i}^a \phi_j e^{-rT_j} = 0 \tag{34b}$$

$$K_i: \quad - \int_{T_i}^{T_{i+1}} \frac{\partial c_i}{\partial K_i} N e^{-rt} dt - \frac{\partial A_i}{\partial K_i} e^{-rT_i} - \phi_i e^{-rT_i} = 0 \tag{34c}$$

Then

$$\begin{aligned}
R_i^* &= \int_{T_i}^{T_{i+1}} \tau^* N^* e^{-r(t - T_i)} dt \\
&= \int_{T_i}^{T_{i+1}} (-NK_i \frac{\partial c_i}{\partial K_i})^* e^{-r(t - T_i)} dt \\
&= (\frac{\partial A_i}{\partial K_i} K_i + \phi_i K_i)^* \quad \text{(using (34c))} \\
&= (\frac{\partial A_i}{\partial K_i} K_i + (\frac{\partial A_i}{\partial I_i} - \frac{\partial A_{i+1}}{\partial I_{i+1}}) e^{-r(T_{i+1} - T_i)} K_i)^* \quad \text{(subtracting (34b) for } I_{i+1} \text{ from (34b) for } I_i) \tag{35}
\end{aligned}$$

Equation (35) has the following interpretation. Suppose that an extra unit of investment is added to the i th capacity expansion, and a unit of investment subtracted from the $(i + 1)$ st capacity expansion. This perturbation results in an increase in K_i by one unit, but no increase in capacity in

other intervals. The cost of this perturbation is therefore the shadow marginal cost of capital for the period between the i th and $(i + 1)$ st capacity additions, and equals

$$\frac{\partial A_i}{\partial K_i} + \frac{\partial A_i}{\partial I_i} - \frac{\partial A_{i+1}}{\partial I_{i+1}} e^{-r(T_{i+1} - T_i)}$$

Thus, (35) states that at the optimum the discounted toll revenue received over a construction interval equals the quantity of capital employed over that interval times the marginal cost of capital over that interval.

From (35),

$$\int_{i=1}^a R_i^* e^{-rT_i} = \int_{i=1}^a \left(\frac{\partial A_i}{\partial K_i} K_i + \frac{\partial A_i}{\partial I_i} I_i \right) e^{-rT_i} \quad (36)$$

Thus, the self-financing result holds if $A(\cdot)$ is homogeneous of degree one in I and K . There are two interesting situations in which this occurs with intermittent capacity additions being optimal. In the first, there are constant unit costs to the size of each capacity addition, with this constant being independent of the size of the facility but depending negatively on $T_i - T_{i-1}$. Thus, the intermittence of capacity additions can occur even with constant costs to the size of capacity additions if there are costs of haste. The second situation where self-financing results occurs when there are decreasing costs with respect to I_i , but increasing costs with respect to K_i , such that an equiproportional increase in I_i and K_i results in an equiproportional increase in capacity expansion costs.¹⁵ Again, the self-financing result holds with constant long-run ray average costs, but not generally otherwise.

A decentralization procedure here is as follows: The planner charges the facility operator a lump sum (equal to the right-hand-side of (35)) at the time each increment of capacity is added, and instructs the operator to charge the optimal toll. When the lump sum is paid off, the next capacity expansion should occur, its size being determined by the planner.

4.5 Fixed increments to capacity

In a recent paper, Oum and Zhang (1990) considered the situation where capacity must be

added in fixed increments. Their example is airport runways. The practical interest of this case is moot since with a fixed number of airport runways, capacity can be altered by varying the quality of runway maintenance, traffic control, etc. Nevertheless, the situation is of theoretical interest and of practical importance too with only one or two runways.

The social surplus function is

$$S = \int_{i=1}^a \int_{T_i}^{T_{i+1}} \int_{p(t)}^a N(p', t) e^{-rt} dp' dt - \int_{i=1}^a \bar{k} e^{-rT_i} + \int_{i=1}^a \int_{T_i}^{T_{i+1}} (p(t) - c(N(p(t), t), i\bar{I})) N(p(t), t) e^{-rt} dt \quad (37)$$

where $\bar{I}$ is the fixed size of each capacity increment, and $\bar{k}$ the fixed cost. The planner chooses $p(t)$ and T_i . The first-order conditions are:

$$p(t): (p - (c_i + N \frac{\partial c_i}{\partial N})) N_p e^{-rt} = 0 \quad (38a)$$

$$T_i: \left(- \int_{p(T_i^+)}^{p(T_i^-)} N(p', T_i) dp' - ((p - c_i)N) \Big|_{T_i^+} + ((p - c_{i-1})N) \Big|_{T_i^-} + r\bar{k} \right) e^{-rT_i} = 0 \quad (38b)$$

where $p(T_i^-) = a$ for $i = 1$, and T_i^- and T_i^+ denote just before and just after addition of the i th capacity increment. Equation (38b) states that the fixed increment to capacity should be added when the marginal benefit from postponing the capacity addition for a unit of time equals the marginal cost. The marginal benefit equals the interest on the delayed construction costs plus the toll revenue if construction is delayed. The marginal cost equals the toll revenue if construction is undertaken plus the gain in consumers' surplus. Note that this is a local condition. It ignores how demand will grow.

Oum and Zhang pointed out that the self-financing result does *not* in general hold in this situation. They provided the following intuition. Consider an alternative demand path which coincides with $N(p, t)$ at $T_1, \dots, T_a$, but is lower at all points in time between T_1 and T_2 , etc. Since

this alternative demand path satisfies (38b), capacity additions occur at the same times (assuming that the second-order conditions are satisfied) as with the original demand path. Thus, discounted capacity construction costs are the same for both demand paths. But since demand is almost everywhere higher with the original demand path, discounted toll revenues are higher. Thus, the self-financing result cannot hold for both cases.

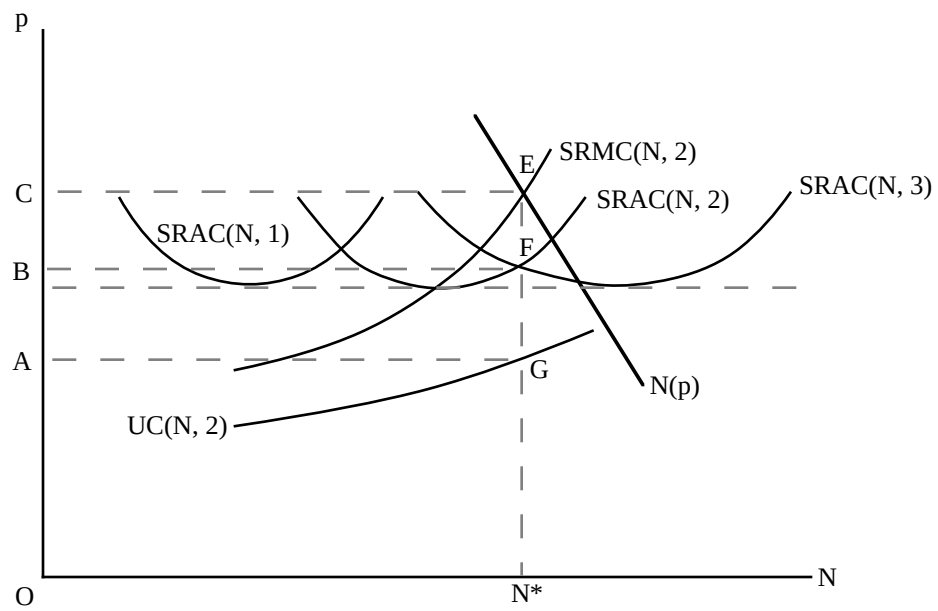
Some intuition concerning this negative result can be gleaned from considering the analogous static problem, which is displayed in Figure 2. Even though the convex hull of the short-run average cost curves is flat, the long-run average cost curve is scalloped-shaped. Since the degree of self-financing depends on the *local* degree of returns to scale, the self-financing result will not in general hold. As drawn, the optimum occurs with two units of capacity, at the point of intersection of the corresponding short-run marginal cost curve and the demand curve. Toll revenue is ACEG, while construction costs are ABFG.

Figure 2 can also be interpreted as the snapshot of a intertemporal equilibrium, where the short-run average cost curves are defined to include amortized construction costs. Over time, in a growing economy, the demand curve moves to the right. Losses will be made for a period of time after the addition of each capacity increment, followed by a period of profits prior to the addition of the next increment. The present value of profits between the addition of successive capacity increments will depend on the pattern of demand growth, the elasticity of demand, the congestion technology, and the interest rate.

We have argued that the self-financing result holds when long-run ray average costs are constant, and not generally otherwise. Here, long-run ray average costs are scalloped-shaped, not constant. Thus, the Oum-Zhang result is consistent with our argument.

We have tried, with only partial success, to determine the extent of cost recovery (in present value terms) with fixed capacity increments.¹⁶ We have constructed examples with more-than-full and others with less-than-full cost recovery, and also obtained bounds on the extent of cost recovery for special cases. But general results have eluded us.

Figure 2. Cost curves with fixed capacity increments



5. Literature Review

This is an appropriate point at which to review the literature. To our knowledge, there are no papers which examine the self-financing of *congestible* facilities in a growing economy. However, there is a close relationship between the theory of public utilities (e.g., electricity networks and water distribution systems) and the theory of congestible facilities, and the self-financing of public facilities in a growing economy *has* been studied. Congestible facilities differ from public facilities in that increased usage results in higher user costs in the former, and in higher production costs in the latter. In the first-best environment that we study, the two classes of problems are isomorphic.¹⁷ To illustrate, consider the simplest problem we analyzed with continuous additions to capacity. Social surplus was

$$S = \int_0^a \int_{p(t)}^a N(p', t) e^{-rt} dp' dt - \int_0^a \rho I(t) e^{-rt} dt - \rho K(0) + \int_0^a (p(t) - c(N(p(t), t), K(t))) N(p(t), t) e^{-rt} dt$$

The first term is discounted consumer surplus, the second and third discounted investment expenditures, and the fourth discounted toll revenues. Simple rearrangement gives

$$S = \int_0^a \int_{p(t)}^a \{ \int_{p(t)}^a N(p', t) dp' + p(t) N(p(t), t) \} e^{-rt} dt - \int_0^a \rho I(t) e^{-rt} dt - \rho K(0) - \int_0^a C(N(p(t), t), K(t)) e^{-rt} dt,$$

where $C(N(p(t), t), K(t)) = c(N(p(t), t), K(t))N(t)$. The first term is gross discounted social surplus, the second and third discounted investment expenditures, and the fourth can be interpreted as discounted production costs with output $N(p(t), t)$ at time t . With this rearrangement, we have the isomorphic public utilities problem.

We have not uncovered any references which deal with the self-financing of public utilities in a growing economy when there are continuous additions to capacity. Starrett (1978), however,

seemed to take this problem as solved. Presumably he was drawing on the result that, with constant returns to scale, the cost-recovery theorem holds with dated commodities. Instead, the literature has focused on the optimal management of public utilities in a growing economy when additions to capacity are "lumpy."¹⁸

The first modern paper on this topic appears to be Starrett (1978). The quality of argument suggests that Starrett had a deep understanding of the problem. Yet he made no mention of discounted cost recovery, and instead focused on the question: If demand growth is exponential and if there are decreasing costs to variable-size capacity additions, will the steady-state revenues generated over an investment cycle be sufficient to finance the *next* capacity increment? His answer is: normally yes. He acknowledges that losses will be made over the start-up period, but does not incorporate them into his analysis. Presumably his interest was in whether a public utility can be expected to pay for itself (in financial terms) after the government incurs the start-up costs. Woroch (1987) is based on Starrett's paper, and investigated those circumstances in which revenues generated over an investment cycle are not sufficient to finance the next capacity increment.

Rees (1986) examined the case of fixed capacity increments. In his review of the literature, he noted (p. 197) that "(discounted financial) profit (between capacity increments) may or may not be positive, even on the assumption of constant returns to scale – this depends on the precise growth path of demand." Thus, to some extent at least, he anticipated Oum and Zhang (1990). But the focus of his paper is on second-best issues – profit constraints, and pricing constraints to prevent "excessively" high prices immediately prior to capacity expansions.

Surprisingly, therefore, it appears that there has been no systematic analysis of discounted cost recovery in the public utilities literature. Thus, our paper contributes not only to the literature on congestible facilities, but also to that on public utilities.

6. Conclusion

In this paper we investigated the conditions for first-best self-financing of congestible facilities

in a variety of intertemporal economic environments. One can view the planner as producing a continuum of dated commodities – uses of the facility at different points in time. Associated with this production process is a cost function. If this cost function exhibits constant long-run ray average costs, efficient pricing results in the facility breaking even. This translates into self-financing/cost recovery – with optimal capacity management, the discounted revenue from the optimal toll covers the discounted costs of construction and operating expenses.

We examined the implications of this general result under a wide range of technological specifications. We first considered technological specifications which lead to continuous additions to capacity. We showed for example that with constant costs to capacity, and depreciation and maintenance, the self-financing result holds if the depreciation function exhibits constant returns to scale (to capacity and investment). If, alternatively, there are convex adjustment costs, cost recovery continues to hold as long as investment costs exhibit constant returns to scale. We then considered a variety of technological specifications which result in intermittent additions to capacity. One surprising implication of the sufficiency of constant long-run ray average costs for self-financing is that the timing of capacity additions need not be optimal, since such nonoptimality does not upset the constancy of long-run ray average costs; however, the size of capacity additions must be optimal conditional on the timing. We examined three exogenous sources of intermittence – a single facility size decision with no additions or replacements, a planning cycle, and a sequence of facility replacements – and then treated a more general specification of capacity addition costs. For the last model, we identified two situations where intermittent capacity additions are consistent with cost recovery. In the first, there are constant costs to the scale of capacity additions and costs to haste – unit investment costs fall the longer the period of time since the last capacity addition; in the second, decreasing costs to investment are offset by increasing costs to capacity.

We also considered the case of fixed increments to capacity, such as airport runways. We showed that self-financing generally does not occur, even when the costs of successive increments to capacity are the same. The reason is that this technology does not exhibit constant long-run ray

average costs – the convex hull is flat, but the cost curve itself is scalloped-shaped. Intuitively, because of the discreteness of capacity increments an incremental change in scale does not induce an equiproportional change in discounted costs.

While we focused on identifying situations in which full discounted cost recovery will occur, our analytics can be applied to determine the extent of first-best discounted cost recovery for congestible facilities under a wide range of technologies. Our analysis can also be extended to treat a variety of second-best situations – to solve for future discounted cost recovery given that current capacity is nonoptimal, and to determine second-best cost recovery under alternative pricing and investment constraints.¹⁹

We paid some attention to decentralization rules. With continuous additions to capacity, we found that capacity should be expanded at that rate at which toll revenues continually cover amortized capacity costs; with intermittent additions to capacity, the scope for decentralization appears more limited.

We hope, in the near future, to extend our analysis to stochastic dynamic environments. We hypothesize that with uncertainty, expected discounted cost recovery will occur in the same situations as does cost recovery in this paper. The intuition is that, if prices are fully flexible, while introducing uncertainty will augment the commodity space – facility uses will be both dated and state-contingent – with constant returns technologies, long-run ray average costs will remain constant. We also plan to investigate decentralization rules which are intrinsically more interesting when uncertainty is present.

There is now a sizeable literature which estimates static returns to scale for public utilities and congestible facilities. Our analysis demonstrates that discounted cost recovery in growing economies depends not only on static returns to scale but also the technological characteristics of facility maintenance and capacity expansion. We hope that our paper will stimulate econometric work along these lines.

Finally, we hope that our extension of the theory of congestion pricing and investment to an intertemporal setting will encourage more sophisticated policy analysis aimed at developing practicable congestion pricing and investment programs.

References

- Arnott, R. and M. Kraus. (1993). "The Ramsey Problem for Congestible Facilities," *Journal of Public Economics*, 50, 371-396.
- Arnott, R. and M. Kraus. (1995). "Financing Capacity in the Bottleneck Model," *Journal of Urban Economics*, forthcoming.
- Baumol, W.J., J.C. Panzar and R.D. Willig. (1982). *Contestable Markets and the Theory of Industry Structure*, Harcourt Brace, New York.
- Berglas, E. and D. Pines. (1981). "Clubs, Local Public Goods and Transportation Models: A Synthesis," *Journal of Public Economics*, 15, 141-162.
- Brock, W.A. and W.D. Dechert. (1985). "Dynamic Ramsey Pricing," *International Economic Review*, 26, 569-591.
- Keeler, T.E. and K.A. Small. (1977). "Optimal Peak-Load Pricing, Investment, and Service Levels on Urban Expressways," *Journal of Political Economy*, 85, 1-25.
- Kraus, M. (1981). "Scale Economies Analysis for Urban Highway Networks," *Journal of Urban Economics*, 9, 1-22.
- Mohring, H. and M. Harwitz. (1962). *Highway Benefits: An Analytical Framework*, Northwestern University Press, Evanston.
- Oum, T. and Y. Zhang. (1990). "Airport Pricing: Congestion Tolls, Lumpy Investment, and Cost Recovery," *Journal of Public Economics*, 43, 353-371.
- Rees, R. (1986). "Indivisibilities, Pricing and Investment: The Case of the Second Best," *Journal of Economics*, 55, 195-210.
- Royal Commission on National Passenger Transportation. (1992). *Directions: The Final Report of the Royal Commission on National Passenger Transportation*, Minister of Supply and Services, Ottawa.
- Starrett, D. (1978). "Marginal Cost Pricing of Recursive Lumpy Investments," *Review of Economic Studies*, 45, 215-227.

- Strotz, R.H. (1965). "Urban Transportation Parables," in J. Margolis, ed., *The Public Economy of Urban Communities*, Resources for the Future, Washington.
- Transportation Research Board. (1994). *Curbing Gridlock: Peak-Period Fees to Relieve Traffic Congestion*, Special Report 242, National Academy Press.
- Woroch, G. (1987). "On Pricing with Lumpy Investment: Two Problems of Boiteux," GTE Laboratories Economics Working Paper 87-2.

Footnotes

- * The authors would like to thank seminar participants at Australian National University, Boston College, Indiana University, the University of Montreal, the University of Munich, and Princeton University, as well as participants at the Berglas Conference, for helpful discussions. Avinash Dixit, Vesa Kanninen, Ray Rees, Hans-Werner Sinn, and Oded Hochman, the discussant of the paper at the Berglas Conference, provided particularly insightful comments. Arnott would like to thank the National Science Foundation for financial support under Grant SES-9210319, and the Department of Economics at the University of Melbourne, where the first draft of the paper was written, for its hospitality.
- ** This paper is dedicated to the memory of Eitan Berglas. Arnott recalls with fondness Eitan Berglas' intellectual enthusiasm and generosity of spirit, that contributed considerably to his enjoyment of visits to the Department of Economics, Tel-Aviv University in 1981, 1983 and 1985.

Eitan Berglas, in co-authorship with David Pines, wrote "Clubs, Local Public Goods and Transportation Models: A Synthesis," *Journal of Public Economics* 15 (1981), 141-162. That paper brought together and synthesized static models of clubs, local public goods, and congestible facilities, and remains one of the most widely-cited papers in each of those literatures. One topic it explored was the relationship between the Henry George Theorem of local public finance and the self-financing result for congestible facilities.

This paper is a fitting tribute to Eitan Berglas. Not only is it an intellectual descendant of the Berglas-Pines paper, but also the central insight of the Berglas-Pines paper enriches its interpretation – though couched in the language of congestible facilities, the models to be presented can be interpreted as club or local public goods models as well.

1. The result was originally derived by Herbert Mohring.
2. The extent of cost recovery when either pricing or capacity is nonoptimal can be calculated.

The cost recovery theorems have stimulated a large number of studies which have attempted to measure the degree of homogeneity of the capacity cost function for various

congestible facilities. For example, Kraus (1981) and Keeler and Small (1977) examine the degree of homogeneity of the capacity cost function for highways.

3. For example: Congestion pricing (and political constraints on its implementation) was a central concern of the (Canadian) Royal Commission on National Passenger Transportation (1992); the Transportation Research Board (1994) has recently shown interest in congestion pricing; many countries are considering the feasibility of urban auto congestion pricing as part of IVHS systems they are experimenting with; and there is increasing discussion of the application of congestion pricing to airports.
4. Cost recovery in a growing economy has, however, been discussed in the public utilities literature which is closely related to the literature on congestible facilities. We review the relevant public facilities literature in Section 5. Suffice it for the moment to say that the public utilities literature, while relevant, does not provide as systematic an analysis as we do, and does not consider the central concern of this paper – discounted cost recovery.
5. Throughout the analysis, we ignore second-order conditions. We are interested in the circumstances under which the self-financing results hold at the global optimum; it is of secondary concern whether there are other stationary points where the self-financing results hold as well. In most of our models, corner solutions are not of interest and are ignored. Where they are of interest, notably in the case of irreversible investment, we shall treat them explicitly. Hence, most of our analysis entails simple examination of the first-order conditions for a series of planning problems.
6. One seminar participant (mis-) interpreted our paper as advocating the market provision of congestible facilities on the grounds that, when the optimum is characterized by exact cost recovery, competition between congestible facilities decentralizes the optimum. We focus on exact cost recovery because it is the central case of theoretical interest, not because we judge constant long-run average costs to be the norm. In fact, many and perhaps most congestible facilities are characterized by decreasing long-run average costs, in which case the optimum cannot be attained through market provision.

7. We plan to address this issue in subsequent work. The difficulty is in developing a specification of the capacity expansion technology that is sufficiently general that all the specific models to be treated in this paper are special cases.
8. These assumptions together imply that optimal capacity grows throughout time. Thus, issues associated with irreversibility do not arise.
9. It turns out that here and throughout the rest of the paper we do not use the infinite-horizon transversality-condition. Consequently, in subsequent models when writing down the first-order conditions we shall omit the infinite-horizon transversality condition.
10. This step employs the condition $\lim_{t \rightarrow a} K^*(t)e^{-rt} = 0$. Its derivation goes as follows. From (7b) and (7c), N^*/K^* is the same at all times t . Then from (7a) the optimal toll is constant over time. Thus, at an optimum, discounted toll receipts are given by

$$\tau^* \int_0^a N(p^*, t)e^{-rt} dt \quad (39)$$

If it were not the case that $\lim_{t \rightarrow a} K^*(t)e^{-rt} = 0$, then $\lim_{t \rightarrow a} N(p^*, t)e^{-rt} \neq 0$, and the integral in (39) would diverge, violating our assumption that the objective function is bounded from above.

Analogous arguments apply to the analogous step in subsequent models. The proofs will be omitted.
11. With time variation in the interest rate and the price of investment, it may be optimal to disinvest even though the economy is growing. Thus, in this subsection we assume perfect reversibility of investment. This assumption is relaxed in the next subsection.
12. The constraint imposed by irreversibility binds only if, in its absence, disinvestment is optimal. To allow for this possibility, we drop the assumption made in Section 3.1 that population is continually growing.
13. Realistically, $Z_I(\cdot)$ would be discontinuous at $I = 0$. To avoid unnecessary complication, we

approximate this with a smooth function.

14. This is consistent with the interpretation of adjustment costs as the costs of adjusting to operating at a different level of capacity, but not with the interpretation of adjustment costs as the costs of installing capacity in a hurry.
15. In most cases, $A(\cdot)$ is not homogeneous of degree one in I and K . Typically, there are decreasing average costs for each capacity addition due to economies of scale in both planning and construction, which are little affected by the size of existing capacity. If, for example, $A(\cdot)$ is homogeneous of degree .8 in I and K , then with optimal tolling and with capacity additions optimal conditional on timing (which need not be optimal) the discounted cost-recovery ratio is .8.
16. A casual reading of Oum and Zhang might give the false impression that fixity of capacity increments typically results in more-than-full cost recovery.

Oum and Zhang calculated what might be termed *financial* cost recovery between capacity additions, viz. they calculated the ratio of the toll revenue between T_i^* and T_{i+1}^* discounted to T_i^* to the *cost of the capacity increment at T_i^** . This differs from *economic* cost recovery between capacity additions, the ratio of the toll revenue between T_i^* and T_{i+1}^* discounted to T_i^* to the discounted *amortized costs of capacity over the period*.

Discounted financial costs are, of course, the same as discounted amortized or economic costs. However their timing is different. Since the financial costs associated with a capacity increment are incurred at the time of construction, while the economic costs occur later, financial costs come earlier than economic costs.

Oum and Zhang gave the financial cost-recovery ratio for the second, third, and fourth (p. 369) increments, but not for the first. Since the financial cost-recovery ratio for the first increment is likely to be considerably lower than for subsequent increments, their results are misleading.

17. We have shown that there are differences between the two classes of problems in second-best

environments. See Arnott and Kraus (1993).

18. The term "lumpy investment" has been used to refer to situations where the size of capacity increments is fixed (Oum and Zhang (1990) and Rees (1986)) and where the size of capacity increments is a choice variable (Starrett (1978) and Woroch (1987)). To avoid this ambiguity, we have not used the term.
19. Some work along these lines has already been done in the public utilities literature. For example, Brock and Dechert (1985) have analyzed dynamic Ramsey pricing, and Rees (1986) has studied the effects of smoothing prices over a construction interval.