Financing Capacity on the Bottleneck Model

Richard Arnott
and
Marvin Kraus

Working Paper No. 222
September 1992

Financing Capacity in the Bottleneck Model

Abstract

It is well known that, for a congestible facility with constant long-run average cost, the revenue from the unconstrained optimal toll (set so that each individual faces the marginal (social) cost of a use) exactly covers the cost of optimal capacity. This paper investigates under what circumstances the first-best pricing and investment rules apply when the time variation of the toll is constrained, and when users differ in *unobservable* characteristics so that the same toll must be applied to heterogeneous users. Both the bottleneck model and the traditional flow congestion model are considered.

commuters, first-best pricing is feasible *even when the variation of the toll by time of day is constrained*; more specifically:

> Suppose that $\tau(t) = k + \lambda(t)$, where $\tau(t)$ is the toll at the bottleneck at time t, k is an unconstrained parameter, and the function $\lambda(\cdot)$ may be constrained. Suppose that the bottleneck's capacity and the toll function $\tau(\cdot)$ are set so that they are socially optimal, subject to the constraints on $\lambda(\cdot)$. Then all commuters face a trip price equal to marginal cost and the first-best self-financing result holds.

A corresponding result holds for the traditional flow congestion model *only when tolling is unconstrained* (Mohring and Harwitz, and Robert H. Strotz (1965)).

We shall focus on the following problem: Suppose that commuters are heterogeneous, with differences -- e.g., in work start time or the value of time -- that are unobservable to the tolling authority. Differentiating the toll across commuter types would therefore be impracticable. In the context of the bottleneck model, under what circumstances is it then feasible to establish marginal cost trip prices?

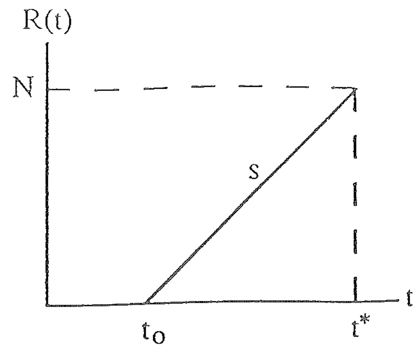
In Section 1, we model this problem in the simplest possible way: two commuter types who differ in terms of a single parameter. For the case in which $\lambda(\cdot)$ is unconstrained, we obtain the Arnott et al. result; for the case in which $\lambda(\cdot)$ is constrained, we establish a counterexample. In Section 2, the result for the unconstrained case is shown to hold for a general model of unobservable heterogeneity. In Section 3, the self-financing results of the Arnott et al. and present papers are reconciled with those implied by the traditional model.

1. A Simple Example

Consider an economy with two types of commuters. The number of commuters of type i ($i = 1, 2$) is denoted by N_i. For now, take N_1 and N_2 as given.

Every day, each commuter makes a work trip (driving his own car) from his residence at A to his workplace at B. A and B are connected by a single road which has a single bottleneck whose

Figure 1



When departures are optimal, aggregate time costs as a function of N_1, N_2 and s are given by

$$C(N_1, N_2, s) = (\beta_1 N_1^2 + 2\beta_1 N_1 N_2 + \beta_2 N_2^2)/2s \tag{5}$$

To see this, note that, when departures are optimal, the average schedule delay costs of type 1 and type 2 commuters are respectively $\beta_1(t^* - (t_o + t')/2)$ and $\beta_2(t^* - t')/2$. Using (3) and (4) in these expressions then yields (5). For later use, note that $C(N_1, N_2, s)$ is homogeneous of degree one in N_1, N_2 and s.

We now consider the decentralized economy, in which there is an *anonymous* toll at the bottleneck at time t equal to $\tau(t)$. A type i commuter faces a trip price at departure time t given by

$$p_i(t) = c_i(t) + \tau(t) \tag{6}$$

Taking q(t) as given, he chooses a departure time t which minimizes $p_i(t)$ subject to $t + q(t) \leq t^*$. Let T_i denote the set of departure times chosen by commuters of type i, and let $\overline{T}_i$ denote the complement of T_i in $\Omega = \{t: \ t + q(t) \leq t^*\}$. The economy's equilibrium conditions are that, for some pair of trip prices P_1 and P_2,

$$p_1(t) = P_1 \qquad t \in T_1 \tag{7a}$$

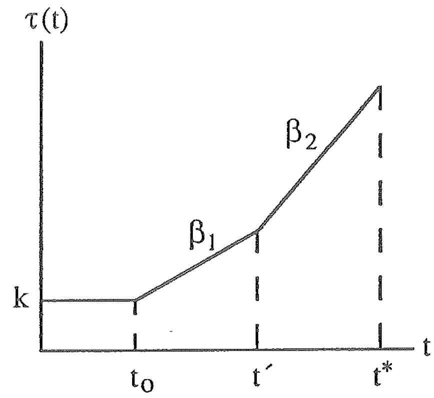
$$p_1(t) \geq P_1 \qquad t \in \overline{T}_1 \tag{7b}$$

$$p_2(t) = P_2 \qquad t \in T_2 \tag{7c}$$

$$p_2(t) \geq P_2 \qquad t \in \overline{T}_2 \tag{7d}$$

It follows that, for any value of the parameter k, the toll function defined by

Figure 2



Benefits depend on N_1 and N_2, but not on departure times. For any N_1 and N_2, the optimal departure pattern is thus the cost-minimizing pattern identified above.

Given this departure pattern, we can write the objective function

$$\int_0^{N_1} P_1(n_1)dn_1 + \int_0^{N_2} P_2(n_2)dn_2 - C(N_1, N_2, s) - K(s) \tag{13}$$

where K(s) is the cost of providing a capacity of s, and $C(N_1, N_2, s)$ is given by (5). From (13), we obtain the marginal-cost pricing conditions

$$P_1 = \partial C/\partial N_1 \tag{14}$$

$$P_2 = \partial C/\partial N_2 \tag{15}$$

and the optimal capacity rule

$$\partial C/\partial s + K'(s) = 0 \tag{16}$$

In the decentralized economy, how are the marginal-cost pricing conditions to be fulfilled? From (5),

$$\partial C/\partial N_1 = \beta_1 N/s \tag{17}$$

while

$$\partial C/\partial N_2 = \beta_1 N_1/s + \beta_2 N_2/s \tag{18}$$

Comparing (17) with (9) and (18) with (10), we see that both of the marginal-cost pricing departures are optimal, $\Gamma(\cdot)$ in Lemma 1.1 is equal to $C(\cdot)$ (see (5)). $\tau(\cdot)$ conforms to (8a)-(8c) with $k = 0$. Thus, $P_i = \partial C/\partial N_i$, $i = 1, 2$. As previously noted, $C(\cdot)$ is homogeneous of degree one. Thus

Proposition 1.1. Suppose that s and $\tau(\cdot)$ are unconstrained optimal. Then the ratio of toll receipts to capacity costs is given by the elasticity of capacity costs with respect to capacity.

1.2 *Other Forms of Toll*

To illustrate what happens when the variation of the toll by time of day is constrained, consider a tolling regime that sets the level, k, of a uniform toll. As above, we initially take N_1 and N_2 as given. Assuming that $\alpha > \max(\beta_1, \beta_2) = \beta_2$,[1] the equilibrium departure pattern is defined by:

(1) Type 1 individuals depart at a uniform rate of $\alpha s/(\alpha - \beta_1)$ over an interval from t_o (recall (3)) to $t'' \equiv t_o + N_1/(\alpha s/(\alpha - \beta_1))$.

(2) Type 2 individuals depart at a uniform rate of $\alpha s/(\alpha - \beta_2)$ over an interval from t'' to $\tilde{t} \equiv t'' + N_2/(\alpha s/(\alpha - \beta_2))$.

In Figure 3, *abcd* and *ad* respectively show cumulative departures from home and cumulative arrivals at work for the equilibrium departure pattern. The vertical distance between the curves is the length of the queue at the time in question. Arrival time at work for an individual who leaves home at time t is the time at which cumulative arrivals equal cumulative departures at time t. An individual who leaves home at time $\tilde{t}$ therefore arrives at work at exactly t^*.

To see that this pattern constitutes an equilibrium, first consider $\dot{p}_i(t)$ over T_i. Since $\dot{\tau}(t) = 0$, successive use of (6), (2) and (1) gives

$$\dot{p}_i(t) = \dot{c}_i(t) = (\alpha - \beta_i)\dot{q}(t) - \beta_i = (\alpha - \beta_i)\dot{D}(t)/s - \beta_i \tag{22}$$

For the given departure pattern and t in T_i, $\dot{D}(t) = \alpha s/(\alpha - \beta_i) - s$. Together with (22), this implies $\dot{p}_i(t) = 0$. $p_i(t)$ is therefore uniform over T_i. Next, consider a change in departure time by a type 1 individual to a time in T_2. Since queuing time increases over T_2 at a faster rate than over T_1, the individual would incur an increase in queuing costs greater than his schedule delay cost savings. For the same reason, a type 2 individual who changed his departure time to a time in T_1 would incur an increase in schedule delay costs greater than his queuing cost savings. Finally, since $q(t_o) = 0$, individuals have no incentive to depart before t_o.

We next derive aggregate time costs as a function of N_1, N_2 and s. Consider a type 1 individual who departs at t_o. He has no queuing costs and schedule delay costs of $\beta_1(t^* - t_o) = \beta_1(N_1 + N_2)/s$. Since the toll is uniform, all individuals of the same type have the same total time costs. Hence, aggregate time costs for type 1 individuals are $\beta_1 N_1(N_1 + N_2)/s$. Since there are no schedule delay costs at departure time $\tilde{t}$, the aggregate time costs of type 2 individuals are

$$
\begin{aligned}
N_2\alpha q(\tilde{t}) &= N_2\alpha D(\tilde{t})/s \\
&= N_2\alpha(t^* - \tilde{t}) && \text{(since } D(\tilde{t}) = s(t^* - \tilde{t})\text{, the number of indi-} \\
& && \text{viduals who arrive at work after } \tilde{t}) \\
&= N_2(\beta_1 N_1 + \beta_2 N_2)/s
\end{aligned}
$$

Aggregate time costs under a uniform toll are therefore

$$\tilde{C}(N_1, N_2, s) = (\beta_1 N_1^2 + 2\beta_1 N_1 N_2 + \beta_2 N_2^2)/s \tag{23}$$

Like $C(\cdot)$, $\tilde{C}(\cdot)$ is homogeneous of degree one.

From the preceding derivation, we can easily determine the equilibrium trip prices P_1 and P_2. P_1 can be expressed as the schedule delay costs of a type 1 individual who departs at t_o plus the toll. Thus

$$P_1 = \beta_1(N_1 + N_2)/s + k \tag{24}$$

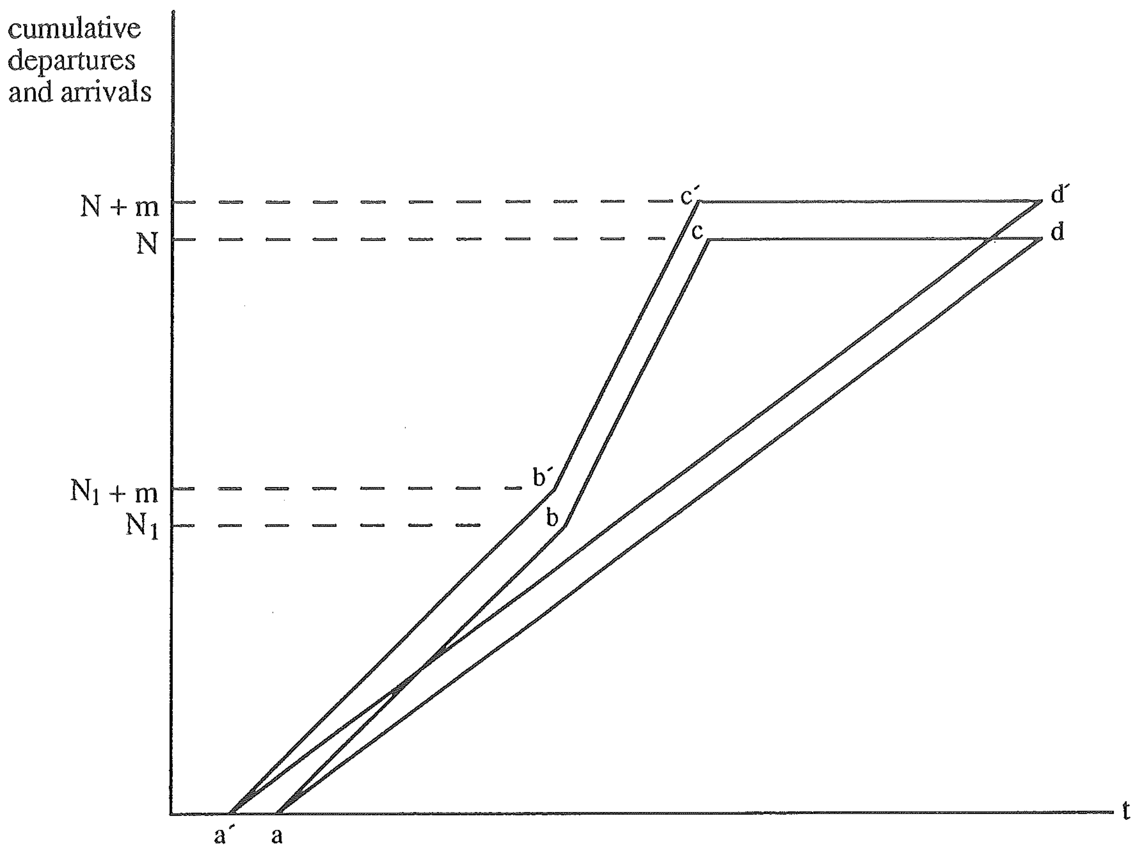
P_2 can be expressed as the sum of the toll and the queuing costs of a type 2 individual who departs at $\tilde{t}$. We can therefore write

$$P_2 = (\beta_1 N_1 + \beta_2 N_2)/s + k \tag{25}$$

Next, let us consider the marginal cost pricing requirements of Lemma 1.1. From (23),

$$\partial\tilde{C}/\partial N_1 = 2\beta_1(N_1 + N_2)/s \tag{26}$$

Figure 4



2's than for 1's, the incremental commuters cause more of a queue buildup if they are 2's.

Letting m go to zero in the limit, we have that a trip at departure time t'' imposes a congestion externality which is the sum of an anonymous schedule delay externality and a queuing externality which is greater if the trip is of type 2 than if it is of type 1. This differs from the unconstrained case, in which an anonymous schedule delay externality makes up the entire congestion effect of a trip.

When the toll is constrained to be uniform, optimal pricing and investment can result in a ratio of toll receipts to capacity costs that is either greater than or less than the elasticity of capacity costs with respect to capacity, depending on demand elasticities. To see this, consider a case in which the demand for trips by individuals of one of the types is completely inelastic, while the demand for trips by individuals of the other type is a decreasing function of price. If it is type 1 trips whose demand depends on price, then the optimal pricing policy is $P_1 = \partial\tilde{C}/\partial N_1$, which is accomplished by setting k equal to the congestion externality imposed by a type 1 trip. Since a type 2 trip imposes a larger congestion externality than a type 1 trip, this results in $P_2 < \partial\tilde{C}/\partial N_2$. This does not distort the optimal investment rule, since the demand for type 2 trips is completely inelastic. It follows from a straightforward adaptation of the proof of Lemma 1.1 that the ratio of toll receipts to capacity costs is *less than* the elasticity of capacity costs with respect to capacity.

If it is type 2 trips whose demand depends on price, then the optimal pricing policy is $P_2 = \partial\tilde{C}/\partial N_2$, which is accomplished by setting k equal to the congestion externality imposed by a type 2 trip. The relationship between the congestion externalities imposed by type 1 and type 2 trips now implies $P_1 > \partial\tilde{C}/\partial N_1$. The ratio of toll receipts to capacity costs will therefore be *greater than* the elasticity of capacity costs with respect to capacity.

2. General Results

The preceding model is now generalized and extended. First, the number of commuter types is generalized from 2 to G. Second, a commuter is no longer prohibited from arriving at work late. We now write

Our key result is:[5]

Lemma 2.2. Let $\mathcal{D}$ represent a cost-minimizing departure pattern. If the toll function is unconstrained, then $\mathcal{D}$ can be decentralized in such a way that $P_i = \partial C/\partial N_i$, $i = 1, \ldots, G$.

Proof. For the departure pattern $\mathcal{D}$, let T_i denote the set of times at which type i commuters depart, and let $\varphi_i(t)$ denote the marginal cost of a type i trip at departure time t when the departure times of other commuters are adjusted, if necessary, to minimize costs. Clearly,

$$\varphi_i(t) = c_i(t) + v_i(t) \tag{30}$$

where $v_i(t)$ is the congestion effect of a type i trip at departure time t. Also, $\varphi_i(t) = \partial C/\partial N_i \ \forall\ t \in T_i$, while, $\forall\ t \notin T_i$, $\varphi_i(t) \geq \partial C/\partial N_i$.

We now assume that the departure pattern is $\mathcal{D}$ and that, for all $i = 1, \ldots, G$, $\tau(t) = v_i(t)\ \forall\ t \in T_i$. (There is no problem in having nondisjoint T_i's, since, at any time t, $v_i(t) = v_j(t)\ \forall\ j \neq i$.[6]) Since $p_i(t) = c_i(t) + \tau(t)$, then $\forall\ t \in T_i$, $p_i(t) = c_i(t) + v_i(t) = \varphi_i(t) = \partial C/\partial N_i$. Thus, for all $i = 1, \ldots, G$, (29a) holds with $P_i = \partial C/\partial N_i$.

It remains to show that, for all $t \notin T_i$, $p_i(t) \geq \partial C/\partial N_i$. If $t \notin \bigcup_{i=1}^{G} T_i$, this can be met by setting $\tau(t) \geq \max\{\partial C/\partial N_i, i = 1, \ldots, G\}$. Suppose, then, that $t \in T_j$ for some $j \neq i$. We have $p_i(t) = c_i(t) + \tau(t) = c_i(t) + v_j(t) = c_i(t) + v_i(t) = \varphi_i(t) \geq \partial C/\partial N_i$. Q.E.D.

Remark 2.1. The preceding proof not only makes Lemma 2.2 intuitive, but brings out how the lemma depends critically on the feature of the model that at any time t, $v_i(t) = v_j(t)\ \forall\ j \neq i$. A simple way to generate a model for which this condition does not hold is to make the bottleneck's capacity specific to trip type. It is easily shown that with capacity specific to trip type in the model of Section 1, decentralization of a cost-minimizing departure pattern with marginal cost trip prices for both groups requires group-specific toll functions. But this is a case in which the two commuter types are observationally distinguishable, so there is nothing impracticable about group-specific toll functions.

The function $C(\cdot)$ is homogeneous of degree one (see the Appendix). This property together

For the case of unconstrained tolling,[9] we have the marginal-cost pricing result, Lemma 2.2. The result derives from the fact that, when departures are cost-minimizing, the congestion externality of a trip at a given departure time consists of an *anonymous* schedule delay externality.

When the variation of the toll by time of day is constrained, equilibrium generally involves queuing, and the congestion externality of a trip includes a *nonanonymous* queuing externality. It follows that if t is a departure time used by more than one commuter type, then not all trips departing at t can be priced at marginal cost. Proposition 1.1 therefore holds only when tolling is unconstrained.

Another possibility is that there are different user groups across which the bottleneck's capacity differs. In this case, even a trip's schedule delay externality is nonanonymous, and pricing trips at marginal cost for all groups requires nonanonymous tolling. But this is a case in which different user groups are observationally distinguishable, so there is nothing impracticable about nonanonymous tolling.

Turning to the traditional model, we begin with the case in which heterogeneity arises from differences in the valuation of travel time.[10] In this case, total user costs are written as $\sum_i\sum_j N_{ij}f_j(N_i, s)$, where N_{ij} is the number of trips by users of type j in subperiod i, $N_i = \sum_j N_{ij}$, and $f_j(N_i, s)$ is average user cost to a type j user in subperiod i. Given this specification, all trips in a given subperiod impose the same congestion externality.[11] It follows that marginal cost pricing can be achieved with anonymous tolling, but only if the toll in each subperiod is unconstrained.

The other case we consider is that in which there are different user groups across which the highway's capacity differs. The way this is typically modeled is to choose one of the groups as a reference group and to write total user costs as $\sum_i\sum_j N_{ij}f_j(\tilde{N}_i, s)$, where, with group 1 as the reference group, s is the highway's capacity for type 1 trips, $\tilde{N}_i = N_{i1} + \sum_{j\neq 1} b_jN_{ij}$, and, for $j \neq 1$, b_j is a parameter equal to the type 1 congestion equivalent of a type j trip. Under this specification, the congestion externality of a trip is nonanonymous. For marginal cost pricing, tolls must be group-specific and unconstrained.

Our results are practically important in two respects, one positive, one negative. On the positive side, they indicate that when there are no constraints on the time variability of the toll, unobservable user heterogeneity does not upset the applicability of first-best pricing and capacity rules for congestible facilities. On the negative side, they indicate that when there are constraints on the time variability of the toll, unobservable user heterogeneity renders the first-best pricing and capacity rules inapplicable. These insights should be useful not only in the context of our analysis – congestion pricing on urban roads – but for other congestible facilities as well.

References

Arnott, R., de Palma, A. and Lindsey, R. (1992). "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand," *American Economic Review*, forthcoming.

Atkinson, A.B. and Stiglitz, J.E. (1980). *Lectures in Public Economics*, McGraw-Hill, Maidenhead.

Braeutigam, R.R. (1989). "Optimal Policies for Natural Monopolies," in R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization*, Volume II, North-Holland, Amsterdam.

Catling, I. and Harbord, B.J. (1985). "Electronic Road Pricing in Hong Kong: The Technology," *Traffic Engineering and Control*, 26, 608-615.

Feldstein, M.S. (1972). "Equity and Efficiency in Public Sector Pricing: The Optimal Two-Part Tariff," *Quarterly Journal of Economics*, 86, 175-187.

Laffont, J.J. and Tirole J. (1990). "The Regulation of Multiproduct Firms. Part I: Theory," *Journal of Public Economics*, 43, 1-36.

Mohring, H. and Harwitz, M. (1962). *Highway Benefits: An Analytical Framework*, Northwestern University Press, Evanston.

Small, K.A. (1982). "The Scheduling of Consumer Activities: Work Trips," *American Economic Review*, 72, 467-479.

Strotz, R.H. (1965). "Urban Transportation Parables," in J. Margolis, ed., *The Public Economy of Urban Communities*, Resources for the Future, Washington.

Vickrey, W.S. (1969). "Congestion Theory and Transport Investment," *American Economic Review Proceedings*, 59, 251-260.