NBER TECHNICAL WORKING PAPER SERIES

THE RAMSEY PROBLEM FOR CONGESTIBLE FACILITIES

Richard Arnott

Marvin Kraus

Technical Working Paper No. 84

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1990

This paper is part of NBER's research program in Taxation. Any opinions expressed are those of the authors not those of the National Bureau of Economic Research.

Technical Working Paper #84
February 1990

THE RAMSEY PROBLEM FOR CONGESTIBLE FACILITIES

ABSTRACT

In recent years, a new set of models drawing on Vickrey [1969] has been developed to analyze the economics of congestible facilities. These models are structural in that they derive the cost function from consumers' time-of-use decisions and the congestion technology. Standard models, in contrast, simply assume the general form of the cost function. We apply the new approach to analyze the Ramsey problem for a congestible facility, and show that the solution generally entails cost inefficiency. Standard models have failed to reveal this result because they treat the cost function as completely determined by technology.

Richard Arnott
Department of Economics
Boston College
Chestnut Hill, MA 02167

Marvin Kraus
Department of Economics
Boston College
Chestnut Hill, MA 02167

In recent years, a new set of models has been developed to analyze the economics of congestible facilities. These models are structural in that they derive the cost function from consumers' time-of-use decisions and the congestion technology. In contrast, standard models of congestible facilities have started with semi-reduced-form specifications of the cost function that are drawn from the public enterprise literature (e.g., Bos [1986], Crew and Kleindorfer [1986]). It turns out, however, that in some contexts the properties of semi-reduced-form cost functions for congestible facilities are qualitatively different from those for public enterprises, since the former incorporate consumers' time-of-use decisions while the latter are completely determined by technology. Thus, depending on context, the standard models of congestible facilities may be misspecified. This paper provides the first analysis of the Ramsey problem for congestible facilities using the new approach, and finds, in contrast to the standard approach, that *the Ramsey optimum generally entails cost inefficiency.*

The standard formulation of the Ramsey public enterprise problem[1] is presented in Braeutigam [1987, p. 43] (slightly modified):

> Consider the case of the N product firm, where y_i is the level of output of the i^{th} service produced by the firm, $i = 1, ..., N$. Let p_i be the price of the i^{th} output, $\mathbf{y}$ the vector of outputs $(y_1, y_2, ..., y_N)$, and $\mathbf{p}$ the vector $(p_1, p_2, ..., p_N)$. Let $y_i(\mathbf{p})$ be the demand schedule for the i^{th} service, $i = 1, ..., N$, and $\psi(\mathbf{p})$ be the consumer surplus at the price vector $\mathbf{p}$. Let w_j be the factor price of the j^{th} input employed by the firm, $j = 1, ..., J$, $\mathbf{w}$ be the vector of factor prices $(w_1, w_2, ..., w_J)$, and $C(\mathbf{y}, \mathbf{w})$ represent the firm's long run cost function. Finally, note that $\pi = \mathbf{py} - C(\mathbf{y}, \mathbf{w})$ corresponds to the economic profit of the firm.
>
> Formally one can represent the Ramsey pricing problem as follows. Ramsey optimal (second best) prices will maximize the sum of consumer and producer surplus, T, subject to a constraint on profits, $\pi \geq \pi_o$, where π_o is a constant.

$$\max_{\mathbf{p}} \; T = \psi(\mathbf{p}) + \mathbf{p}\mathbf{y} - C(\mathbf{y}, \mathbf{w})$$

$$\text{subject to} \quad (1) \;\; \pi = \mathbf{p}\mathbf{y} - C(\mathbf{y}, \mathbf{w}) \geq \pi_0$$

$$(2) \;\; \mathbf{y} = \mathbf{y}(\mathbf{p})$$

Note that the cost function $C(\cdot)$ is determined by technology.

The standard models of congestible facilities employ the above formulation (e.g., Mohring [1970], Berglas and Pines [1981]), but interpret $\mathbf{y}$ to be the vector of the number of users of the facility (indexed by group and/or time period). We shall demonstrate that, with congestion, the "cost" function is not completely technologically determined, but depends also on the vector of output prices, i.e., $C = C(\mathbf{y}, \mathbf{w}, \mathbf{p})$.[2] We shall also demonstrate that the Ramsey optimum will in general entail cost inefficiency, in the sense that, with $\mathbf{y}$ fixed at the vector of Ramsey optimal outputs, the set of price vectors for which $C(\mathbf{y}, \mathbf{w}, \mathbf{p})$ is at a minimum with respect to $\mathbf{p}$ does not in general contain the vector of Ramsey prices.

The new approach to the economics of congestible facilities has been developed almost completely in the context of morning rush-hour auto travel with bottleneck congestion on a point-input, point-output road. The seminal paper was by Vickrey [1969] who solved for the no-toll equilibrium, as well as for the optimal time-varying toll, with inelastic demand and linear costs. The basic idea is that commuters choose when to depart from home to work so as to minimize trip price. There are three components to trip price: travel time cost, the cost of arriving at work before or after the desired arrival time, termed schedule delay cost, and the toll. With basic bottleneck congestion, travel time is time spent in the queue behind the bottleneck. Suppose, to simplify, that everyone has the same desired arrival time, and is identical in other respects as well.[3] Then equilibrium[4] requires that all commuters have the same trip price. This is achieved through adjustment in the pattern of travel time over the rush hour, which depends on the evolution of the queue length and in turn on the distribution of departure times. For example, in the no-toll equilibrium, the time early cost of the first commuter to depart equals the time late cost of the last commuter to depart (since both encounter no queue), which equals the travel time (queuing) cost of the commuter who arrives on time.

Vickrey's model has been extended in many ways;[5] the present paper applies it to the Ramsey problem for a congestible facility. More specifically, we retain the context of Vickrey's model (bottleneck congestion for auto commuters on a single road in the morning rush hour) and consider a situation in which there are scale economies in the provision of bottleneck capacity and two groups of users who differ in their shadow values of time and elasticities of demand. We ask the question: What are the equilibrium characteristics of facility usage when the highway authority chooses optimal capacity and an optimal *anonymous* time-varying toll, subject to the constraint that the deficit incurred (capacity costs less toll revenues) be no larger than a specified amount? Since the only difference between the user groups is a subjective value of time, it is natural to assume that the toll is anonymous. The assumption is consequential; if the toll were nonanonymous, the cost inefficiency result would disappear.

In any equilibrium, all commuters in a given group face the same trip price, inclusive of the toll, P_1 and P_2 respectively. The problem can be insightfully posed in P_1-P_2 space -- see Figure 1. With or without the deficit constraint, it is optimal, at any point in P_1-P_2 space, to choose capacity and the toll function so as to minimize total system costs.[6] Thus, the problem can be decomposed. In the first stage, minimize system costs for every (P_1, P_2), and in the second stage conduct the analysis of equilibrium in P_1-P_2 space, with and without the deficit constraint.

An important feature of the first-stage solution is that because the toll is anonymous, for any P_1, there is one and only one P_2 for which the solution entails cost efficiency. In Figure 1, label the corresponding locus the cost-efficiency locus. At points along the locus, there is no queuing in the first-stage solution. Everywhere off the locus, the first-stage solution involves queuing.

We shall show that the first-best (no deficit constraint) optimum, θ, in which the government has full control over each group's departure pattern, can be decentralized with an anonymous time-dependent toll. The decentralized optimum lies on the cost-efficiency locus. However, the Ramsey optimum, R, will not in general lie on this locus. There are two sources of deadweight loss, inefficiently high costs of production and the costs associated with setting the two groups'

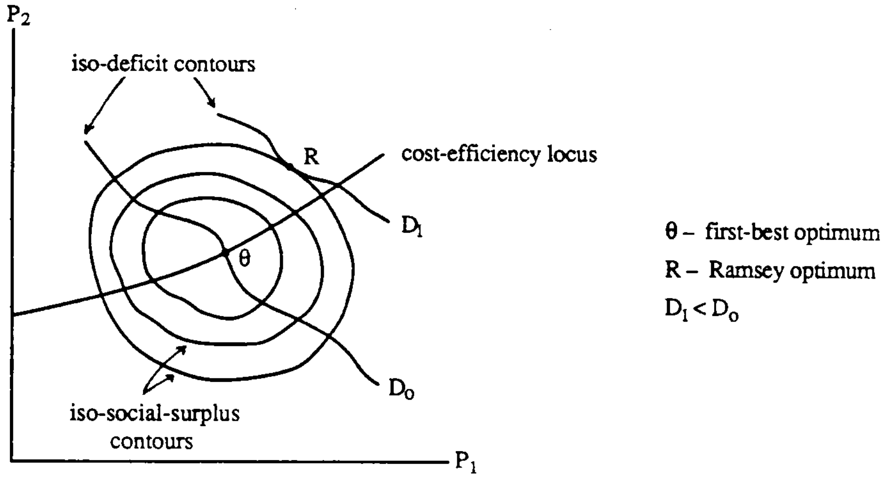


Figure 1. The Ramsey optimum may be characterized by cost inefficiency.

prices above their corresponding marginal social costs, and the Ramsey optimum will generally entail trading off these two sources of deadweight loss.

As noted earlier, the reason this phenomenon has been neglected in the standard approach is that there the user cost function is assumed to be technologically determined. In particular, user costs for the above problem would be written as $C = C(N_1, N_2, s)$, where N_1 and N_2 are the number of commuters in the two groups and s is capacity, without reference to prices.[7] However, user costs change if P_1 and P_2 are altered, holding N_1, N_2 and s fixed. Thus, the user cost function is in fact a semi-reduced-form function that captures both technology and demand.

Our result that the Ramsey optimum entails cost inefficiency is related to, but distinct from, several other results in the literature:

1. The possibility that queuing may be a feature of a second-best optimum has been explicitly considered by Bucovetsky [1984]. Furthermore, a number of papers (Weitzman [1977], Guesnerie and Roberts [1984] and Sah [1987]) have shown that it may be desirable to supplement the price system with nonprice rationing (of which queuing is a particular form) when distortions

are present. All these papers, however, demonstrate the possible desirability of nonprice allocation on second-best *equity* grounds, while we present a second-best *efficiency* argument for the desirability of queuing.

2. Our model is partial equilibrium, treating as exogenous all factor prices and all commodity prices other than those of the commuter trips. The possible desirability of production inefficiency is a general equilibrium phenomenon (see Guesnerie [1980] and Auerbach [1985]). Thus, the cost inefficiency we identify is distinct from the production inefficiency discussed in the optimal taxation literature.

3. Greenwald and Stiglitz [1986] have demonstrated that taxation is in general desirable in the presence of adverse selection to relax self-selection constraints. Because the time-varying toll in our model is anonymous, we encounter self-selection constraints. But endogenous queuing rather than taxation serves to relax the constraints.[8]

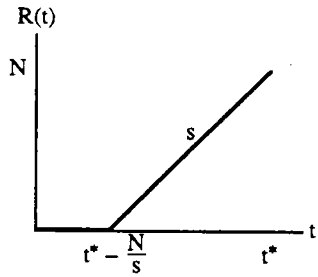
Sections 1-3 take up a series of increasingly complex first-best problems that lead up to the Ramsey problem. In Section 1, commuters are identical and demand is inelastic.[9] In Section 2, demand is still inelastic, but there are two commuter groups. Section 3 retains this heterogeneity, while introducing elastic demand.[10] The Ramsey problem is taken up in Section 4, and Section 5 presents some concluding remarks.

1. Identical Commuters, Inelastic Demand

Initially, commuters are assumed to be identical, and their number, N, is given. Every day, each 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 of capacity (the maximum rate at which cars can pass through the bottleneck per hour) s. If arrivals at the bottleneck ever occur at a rate exceeding s, then a queue forms.

For present purposes, nothing is lost by assuming that there are no travel costs other than queuing time costs at the bottleneck. Thus, a commuter's departure time from home is his arrival time at the bottleneck, and his arrival time at work is his departure time from the bottleneck.

Figure 2



Let d(t) denote the length of the queue at time t, and let q(t) denote queuing time for an individual who leaves home at time t. Then

$$q(t) = d(t)/s \tag{1}$$

Commuters have the same work start time t^* and must be at work by t^*.[11] An individual who leaves home at time t is early for work by an amount $t^* - (t + d(t)/s)$, for which he incurs a schedule delay cost of $\beta(t^* - (t + d(t)/s))$, where β is a parameter equal to the shadow value of time early. Let $\alpha > \beta$ be the shadow value of queuing time.[12] Total time costs for an individual who leaves home at time t are then

$$c(t) = \alpha d(t)/s + \beta(t^* - (t + d(t)/s)) \tag{2}$$

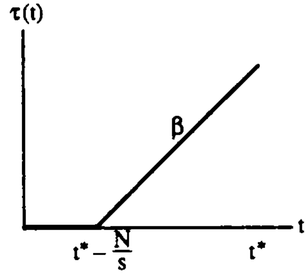
The optimal departure pattern minimizes queuing plus schedule delay costs. In the optimal configuration, individuals depart at a uniform rate of s over the interval $[t^* - N/s, t^*]$. This is shown in Figure 2, where R(t) is cumulative departures by time t. To see that this pattern is optimal, simply note that a queue never forms, and schedule delay costs cannot be made lower.

An increase in s permits a shorter departure interval and hence a reduction in schedule delay costs. The tradeoff between this reduction in schedule delay costs and the added capacity costs determines the optimal capacity.

We now consider the decentralized economy, in which there is a toll at the bottleneck at time t equal to $\tau(t)$. The price of a trip at departure time t is given by

$$p(t) = c(t) + \tau(t) \tag{3}$$

Figure 3



Taking q(t) as given, a consumer chooses a departure time t which minimizes p(t) subject to $t + q(t) \leq t^*$. Let T denote the set of all departure times that are chosen, and let $\overline{T}$ denote the complement of T in $\Omega = \{t: \ t + q(t) \leq t^*\}$. The conditions for equilibrium are that, for some trip price P,

$$p(t) = P \qquad t \in T \tag{4a}$$

$$p(t) \geq P \qquad t \in \overline{T} \tag{4b}$$

We now show that, if the toll gradient is that shown in Figure 3, then the equilibrium pattern of departures coincides with the optimum. To see this, suppose that the toll gradient is as shown, and that the pattern of departures is optimal. Let t_1 and t_2 be two departure times satisfying $t^* - N/s \leq t_1 < t_2 \leq t^*$. $\tau(t_2) - \tau(t_1) = \beta(t_2 - t_1)$, which is exactly the amount by which schedule delay costs are lower at t_2. As for (4b), the toll is the same to the left of $t^* - N/s$ as at $t^* - N/s$, but schedule delay costs are greater.

2. Two Commuter Groups, Inelastic Demand

Consider the same model as above, but with two distinct commuter groups. Type i commuters (i = 1, 2) have a shadow value of time early equal to β_i. The only difference between a type 1 and type 2 commuter is that $\beta_1 < \beta_2$.[13] We also assume that $\alpha > \beta_2$.

There are given numbers of type 1 and type 2 commuters, which we denote by N_1 and N_2, and $N \equiv N_1 + N_2$.

The optimal departure pattern is for type 2's to depart at a uniform rate of s over an interval from

$$t' \equiv t^* - N_2/s \tag{5}$$

to t^*, and 1's to depart at the same uniform rate over an interval from

$$t_0 \equiv t^* - N/s \tag{6}$$

to t'. The reason why the 1's depart first is simply that $\beta_1 < \beta_2$. As in the case of a single group, there is never a queue, and schedule delay costs are minimized.

In the decentralized economy, a type i commuter faces a trip price at departure time t given by

$$p_i(t) = \alpha d(t)/s + \beta_i(t^* - (t + d(t)/s)) + \tau(t) \tag{7}$$

The economy's equilibrium conditions are that, for some pair of trip prices P_1 and P_2,

$$p_1(t) = P_1 \qquad t \in T_1 \tag{8a}$$

$$p_1(t) \geq P_1 \qquad t \in \overline{T}_1 \tag{8b}$$

$$p_2(t) = P_2 \qquad t \in T_2 \tag{8c}$$

$$p_2(t) \geq P_2 \qquad t \in \overline{T}_2 \tag{8d}$$

where T_i is the set of departure times chosen by commuters of type i, and $\overline{T}_i$ is the complement of T_i in Ω.

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

$$\tau(t) = k \qquad t \leq t_0 \tag{9a}$$

$$= k + \beta_1(t - t_0) \qquad t \in [t_0, t'] \tag{9b}$$

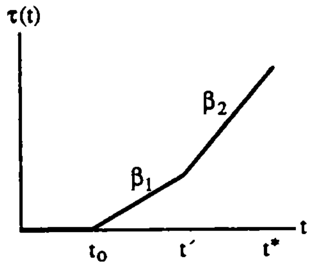
$$= k + \beta_1(t' - t_0) + \beta_2(t - t') \qquad t \in [t', t^*] \tag{9c}$$

results in decentralization of the optimal departure pattern. The proof is similar to that given in Section 1 for the single group case and is therefore omitted. The nature of the toll gradient defined by (9a)-(9c) is shown in Figure 4.

A couple of additional results are needed for later sections of the paper. One is that, when departures are cost-minimizing,[14] aggregate time costs as a function of N_1, N_2 and s are given by

$$C(N_1, N_2, s) = (\beta_1 N_1^2 + 2\beta_1 N_1 N_2 + \beta_2 N_2^2)/2s \tag{10}$$

Figure 4



To see this, note that, when departures are cost-minimizing, 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 (5) and (6) in these expressions then yields (10).

The second result concerns the equilibrium trip prices P_1 and P_2 under the toll gradient defined by (9a)-(9c). The easiest way to determine P_1 is as $p_1(t_o)$. The toll at t_o is k, and a type 1 individual's schedule delay costs are $\beta_1(t^* - t_o) = \beta_1 N/s$. Thus

$$P_1 = \beta_1 N/s + k \tag{11}$$

An easy way to determine P_2 is as $p_2(t^*)$. Since $p_2(t^*) = \tau(t^*)$,

$$P_2 = \beta_1 N_1/s + \beta_2 N_2/s + k \tag{12}$$

Finally, we note that

$$P_2 - P_1 = (\beta_2 - \beta_1)N_2/s \tag{13}$$

3. Two Commuter Groups, Elastic Demand

Consider the same model as in Section 2, but with demand given by

$$N_i = N_i(P_i) \tag{14}$$

We assume that $N_i'(P_i) < 0$, which ensures the existence of the inverse demand functions $P_i(N_i)$.

Society's problem is now one of maximizing the difference between the benefits and costs of trips,[15] where costs include capacity costs and benefits are given by

$$\int_0^{N_1} P_1(n_1)dn_1 + \int_0^{N_2} P_2(n_2)dn_2 \tag{15}$$

Benefits depend on N_1 and N_2, but not on departure times. The optimal departure pattern is thus the cost-minimizing pattern of Section 2.

Given this departure pattern, we can write the objective function

$$\int_0^{N_1} P_1(n_1)dn_1 + \int_0^{N_2} P_2(n_2)dn_2 - C(N_1, N_2, s) - K(s) \tag{16}$$

where $K(s)$ is the cost of providing a capacity of s, and $C(N_1, N_2, s)$ is given by (10). From (16), we obtain the marginal cost pricing conditions

$$P_1 = \partial C/\partial N_1 \tag{17}$$

$$P_2 = \partial C/\partial N_2 \tag{18}$$

and the optimal capacity rule

$$\partial C/\partial s + K'(s) = 0 \tag{19}$$

In the decentralized economy, how are the marginal cost pricing conditions to be fulfilled? From (10),

$$\partial C/\partial N_1 = \beta_1 N/s \tag{20}$$

while

$$\partial C/\partial N_2 = \beta_1 N_1/s + \beta_2 N_2/s \tag{21}$$

Comparing (20) with (11) and (21) with (12), we see that both of the marginal cost pricing conditions are fulfilled when the toll gradient parameter k is set at zero.

To demonstrate this result more heuristically, we first note that, if a marginal type 1 trip is made, there is no effect on t', but a decrease in t_o. The marginal cost of a type 1 trip is therefore the schedule delay cost of a type 1 individual who initially departs at the start of the rush hour. As previously noted, P_1 can be determined as $p_1(t_o)$. When $k = 0$, the toll at t_o is zero, and $p_1(t_o)$

consists of this same schedule delay cost. If a marginal type 2 trip is made, reductions occur in both t' and t_0. The marginal cost of a type 2 trip is therefore the schedule delay cost of the type 2 individual who is initially first to depart plus the difference between the schedule delay costs of the type 1 individuals who are initially first and last to depart. One need only now observe that, when $k = 0$, this schedule delay cost difference for type 1 individuals is the toll paid by the first type 2 individual to depart.

We assume that there are scale economies in providing capacity, by which we mean that the elasticity of K(s) with respect to s is less than unity.[16] Under this assumption, what will be the relationship between toll receipts and capacity costs, assuming optimal pricing and investment? From (10), the user cost function $C(N_1, N_2, s)$ is homogeneous of degree one in N_1, N_2, and s. Together with scale economies in capacity costs, this implies that the joint production process of producing type 1 and type 2 trips is characterized by declining long-run ray average costs. It is well known (e.g., Strotz [1965]) that, under this condition, the toll receipts from long-run marginal cost pricing are less than the capacity costs. A formal proof that the public authority incurs a deficit is provided in the Appendix.

4. The Ramsey Problem

Consider the same model as in Section 3, but with a binding constraint on the allowable deficit. We assume a fixed rather than proportionate deficit constraint.

It appears that the (second-best) optimal departure pattern may be inefficient (whenever we use the term "efficient," we mean "cost-minimizing"). If departures were efficient, (20) and (21) would hold, as well as (11) and (12). From these relationships, $P_1 - \partial C/\partial N_1$ and $P_2 - \partial C/\partial N_2$ would both be equal to k. The amount by which price is raised over marginal cost for one group would have to be exactly the same as for the other. There is therefore a tradeoff between efficiency in departures and price over marginal cost markups that better reflect demand elasticities.

In Section 2, we found that an efficient departure pattern has the characteristic that all type 1 departures occur prior to the time at which type 2 departures commence. In our treatment of the

Ramsey problem, departure patterns that do not have this characteristic will not be considered.[17]

We will see shortly that there is an important role in the problem for departure patterns in which a number of individuals depart in mass at a certain prescribed time. For a departure time that is a mass time, queuing time is assumed to be that of the middle person in the mass. The only change this requires is that we now define d(t) by

$$d(t) = D(t_-) + (D(t_+) - D(t_-))/2 \tag{22}$$

where $D(t_-)$ and $D(t_+)$ are respectively the limiting values from the left and the right of the actual queue length at time t and $D(t_+) - D(t_-)$ is therefore the size of the departure mass at time t.

The approach to the Ramsey problem that we have found to be most successful involves first solving what we will refer to as the feasibility problem. In the feasibility problem, N_1, N_2 and s are all exogenous. There are two parts to the problem. One is to find all price combinations (P_1, P_2) for which the equilibrium conditions (8a)-(8d) hold for some toll gradient τ.[18] Such price combinations will be referred to as feasible. A given price combination might be sustainable by alternative toll gradients, having different departure patterns in equilibrium. The second part of the problem is to determine the minimum user costs consistent with a given feasible price combination.

4.1. *The Feasibility Problem*

We shall first set out the formal derivations, and subsequently provide the intuition.

Our work on the feasibility problem begins with two pieces of groundwork. First, we introduce a new variable θ defined by

$$P_2 - P_1 = \theta(\beta_2 - \beta_1)N_2 / s \tag{23}$$

(23) maps each point (P_1, P_2) in price space into a specific value of θ. Since efficient price combinations are characterized by (13), they are mapped into $\theta = 1$.

Second, we establish a pair of inequalities implied by (8a)-(8d) that will prove to be useful. From (8a), (8d) and (7),

$$P_2 - P_1 \leq (\beta_2 - \beta_1)(t^* - (t + d(t)/s)) \qquad t \in T_1 \tag{24}$$

Similarily, from (8b), (8c) and (7),

$$P_2 - P_1 \geq (\beta_2 - \beta_1)(t^* - (t + d(t)/s)) \qquad t \,\varepsilon\, T_2 \tag{25}$$

Using (23), (24) and (25) can be rewritten

$$d(t) \leq s(t^* - t) - \theta N_2 \qquad t \,\varepsilon\, T_1 \tag{26}$$

$$d(t) \geq s(t^* - t) - \theta N_2 \qquad t \,\varepsilon\, T_2 \tag{27}$$

which are the desired inequalities.

The solution to the feasibility problem is contained in a set of four feasibility properties.

Property 1. A necessary condition for feasibility is $\theta \geq 1/2$.

Proof. Let t_2 denote the time at which type 2 departures commence. Imposing (27) at t_2 and using (22),

$$D(t_{2-}) + \phi N_2/2 \geq s(t^* - t_2) - \theta N_2 \tag{28}$$

where ϕ is the fraction of type 2 individuals who depart in mass at t_2. In order for all arrivals to occur by t^*,

$$s(t^* - t_2) \geq D(t_{2-}) + N_2 \tag{29}$$

From (28) and (29),

$$\theta \geq 1 - \phi/2 \tag{30}$$

(30) together with $0 \leq \phi \leq 1$ implies that $\theta \geq 1/2$. Q.E.D.

Properties 2-4 are proved using the following lemma.

Lemma 1. Given a price combination (P_1, P_2) and a departure pattern satisfying (26) and (27), there exists a toll gradient τ for which the given prices and departure pattern are sustainable.

Proof. The proof is by construction of the toll gradient. At times in T_1, tolls are set to satisfy (8a), while at times in T_2, they are set to satisfy (8c). Along with (8a), (26) implies that (8d) holds over $\bar{T}_2 \cap T_1$. Similarily, (27) along with (8c) implies that (8b) holds over $\bar{T}_1 \cap T_2$. It remains to satisfy (8b) and (8d) over $\bar{T}_1 \cap \bar{T}_2$. This is accomplished by setting sufficiently high tolls at such times. Q.E.D.

Property 2. Price combinations for which $1/2 \leq \theta \leq 1$ are feasible. For any such price combination

(A) The pattern of departures that results in minimum user costs is defined by

(1) Type 1 individuals depart at a uniform rate of s over $[t_0, t')$.

(2) A fraction $2(1-\theta)$ of type 2 individuals depart in a mass at t'.

(3) Type 2 individuals who do not depart at t' depart at a uniform rate of s over $[t' + 2(1-\theta)N_2/s, t^*]$.

(B) Minimum user costs are given by

$$\Gamma(N_1, N_2, \theta, s) = (\beta_1 N_1^2 + 2\beta_1 N_1 N_2 + \beta_2 N_2^2 + 4\alpha(1-\theta)^2 N_2^2)/2s \quad (31)$$

Proof. We first show that the departure pattern defined in (A) satisfies (26) and (27). Together with Lemma 1, this implies feasibility.

The right-hand side of (26) and (27) is a linear decreasing function of t that we denote by f(t), i.e.,

$$f(t) \equiv s(t^* - t) - \theta N_2 \quad (32)$$

When $\theta \leq 1$, f(t) has its root to the right of t' at $t' + (1-\theta)N_2/s$. The value of f at t' is $(1-\theta)N_2$. Now consider the departure pattern defined in (A). Everywhere in T_1, d is zero while f is positive. The given departure pattern thus satisfies (26). At $t' \varepsilon T_2$, d and f are both equal to $(1-\theta)N_2$. Everywhere else in T_2, f is negative. The given departure pattern thus satisfies (27).

The departure pattern defined in (A) will now be referred to as the reference solution to (26) and (27). We next show that it is the most efficient solution to (26) and (27), thereby establishing (A).

From the fact that departures in the reference solution begin at t_0, no departure pattern for which all arrivals occur by t^* can have lower schedule delay costs. We therefore proceed by showing that there is no solution to (26) and (27) that has lower queuing costs than the reference solution.

From the proof of Property 1, any solution to (26) and (27) must satisfy (30), which we now write as

$$\phi \geq 2(1-\theta) \tag{33}$$

The right-hand side of (33) is the fraction of type 2 individuals who are in the reference solution departure mass. The reference solution therefore involves minimum queuing costs among solutions to (26) and (27).

The simplest way to establish (B) is to use the fact that the departure pattern defined in (A) involves the same schedule delay costs as an efficient departure pattern. Γ can therefore be written as the sum of $C(N_1, N_2, s)$ and queuing costs for the departure pattern defined in (A). Queuing is limited to those in the mass, because the second round of type 2 departures begins just as the length of the queue generated by the mass falls to zero. There are $2(1-\theta)N_2$ individuals in the mass, each of whom incurs a queuing cost of $\alpha(1-\theta)N_2/s$. Q.E.D.

Property 3. Price combinations for which $1 \leq \theta \leq 1 + \gamma$ are feasible, where $\gamma \equiv \beta_1 N_1/4\alpha N_2$. For any such price combination

(A) The pattern of departures that results in minimum user costs is defined by

(1) A fraction $2(\theta - 1)N_2/N_1$ of type 1 individuals depart in a mass at time $t_m = t' - 2(\theta - 1)N_2/s$.

(2) Type 1 individuals who do not depart at t_m depart at a uniform rate of s over $[t_o, t_m)$.

(3) Type 2 individuals depart at a uniform rate of s over $[t', t^*]$.

(B) Minimum user costs are given by (31).

Proof. See the Appendix.

Remark 1. To take the cases covered by Properties 2 and 3, suppose that either $1/2 \leq \theta \leq 1$ or $1 \leq \theta \leq 1 + \gamma$. Properties 2 and 3 establish that, in both cases, the optimal time for departures to begin is t_o. This has the interesting implication that the optimal departure pattern involves the same schedule delay costs as for the case $\theta = 1$. The inefficiency thus comes about solely through queuing.

Property 4. Price combinations for which $\theta \geq 1 + \gamma$ are feasible. For any such price combination

(A) The pattern of departures that results in minimum user costs is defined by

(1) A fraction $\beta_1/2\alpha$ of type 1 individuals depart in a mass at time $t_m = t' - (\theta - 1 + \gamma)N_2/s$.

(2) Type 1 individuals who do not depart at t_m depart at a uniform rate of s over $[t_o - (\theta - 1 - \gamma)N_2/s, t_m)$.

(3) Type 2 individuals depart at a uniform rate of s over $[t', t^*]$.

(B) Minimum user costs are given by

$$\Gamma(N_1, N_2, \theta, s) = ((1 - \beta_1/4\alpha)\beta_1 N_1^2 + 2\theta\beta_1 N_1 N_2 + \beta_2 N_2^2)/2s \qquad (34)$$

Proof. See the Appendix.

Remark 2. Suppose that $\theta \geq 1 + \gamma$. Property 4 establishes that the optimal time for departures to begin in this case is before t_o. The inefficiency thus comes about partly through queuing and partly through inefficiently high schedule delay costs.

Some intuition can be gained concerning the qualitative properties of the departure pattern for various intervals of θ from Figures 5 and 6. Turn first to Figure 6b which portrays the equilibrium for $\theta = 1$, a useful reference case. Recall that with $\theta = 1$, the departure pattern is efficient and there is no queue. The slope of the toll gradient is β_1 for $t \in (t_o, t')$ and β_2 for $t \in (t', t^*]$. And the slope of the schedule delay cost function is $-\beta_i$ for individuals of type i. Since there is no queuing, the price function for group 1 is P_1 for $t \in [t_o, t']$ and $P_1 + (\beta_2 - \beta_1)(t - t')$ for $t \in [t', t^*]$;

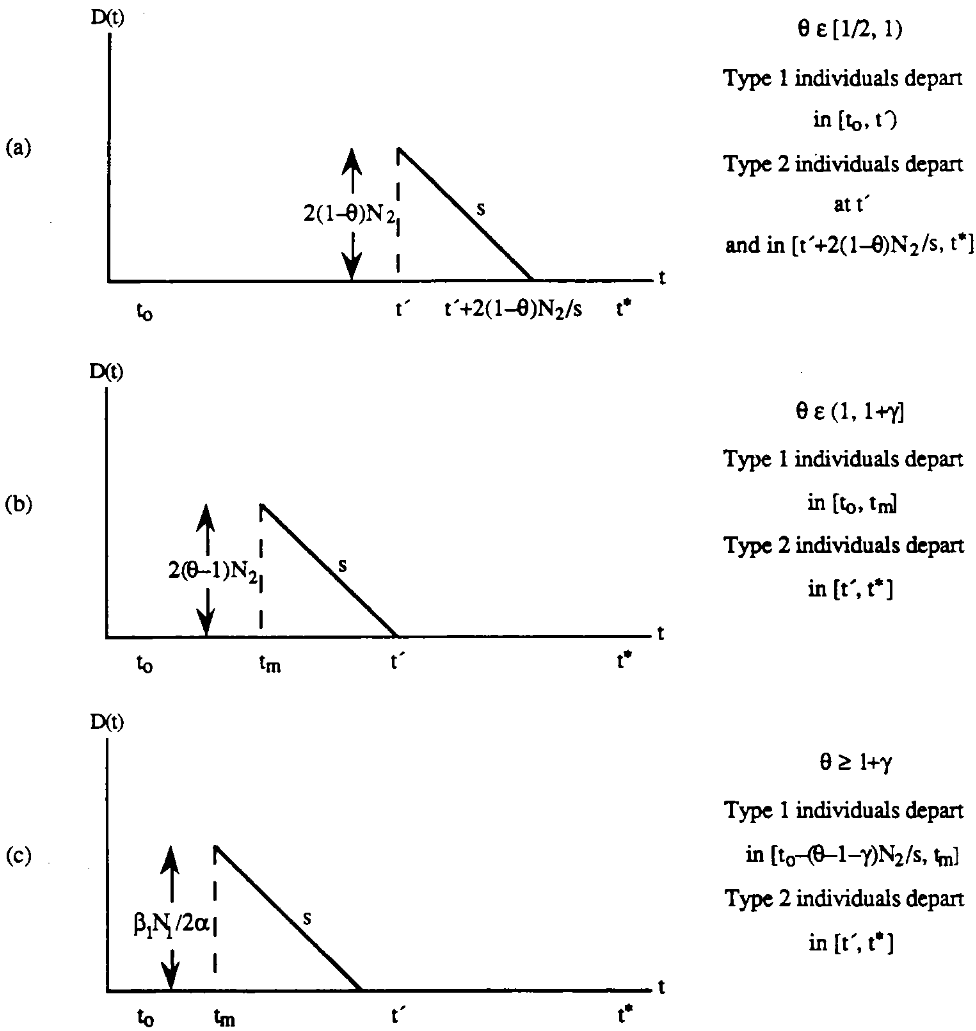


Figure 5. The queuing pattern for various intervals of θ.

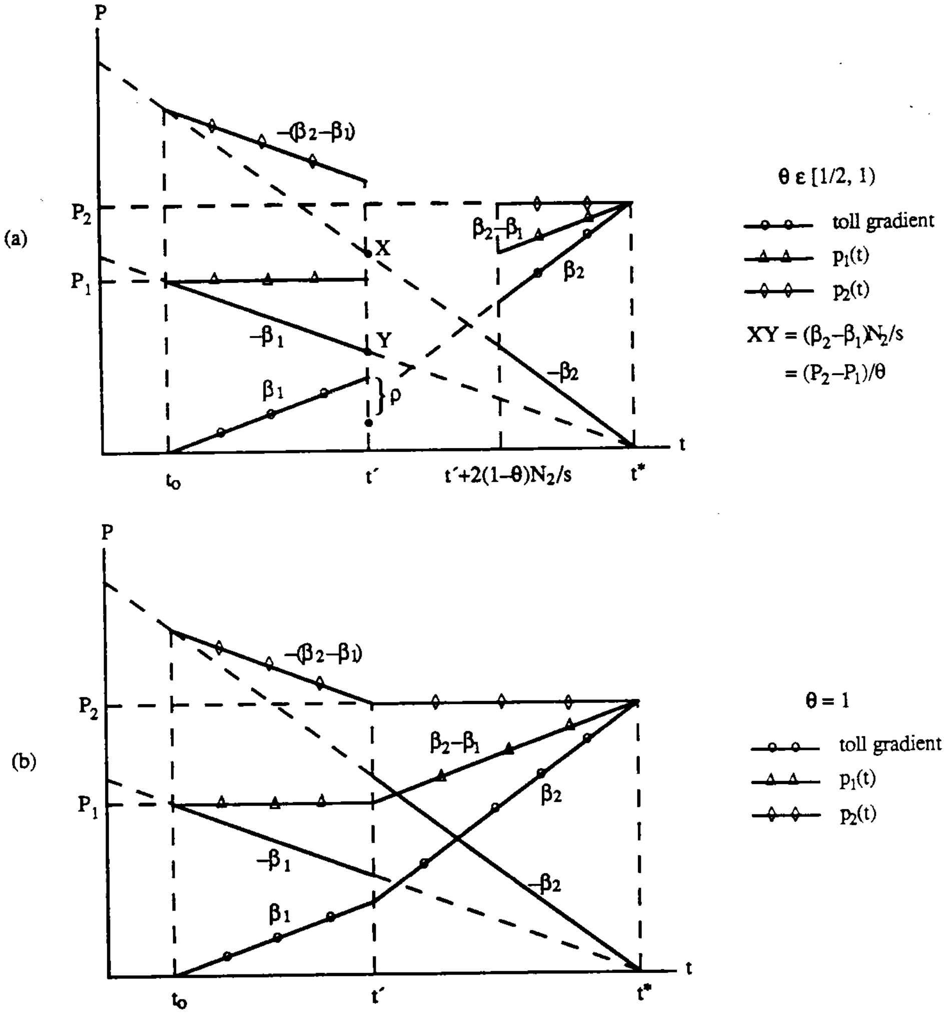


Figure 6. Derivation of the price functions.

and for group 2, $P_2 + (\beta_2 - \beta_1)(t' - t)$ for $t \varepsilon [t_0, t']$ and P_2 for $t \varepsilon [t', t^*]$. Thus, type 1 individuals travel in $[t_0, t']$ and type 2 in $[t', t^*]$.

Turn next to Figures 5a and 6a which portray the case $\theta \varepsilon [1/2, 1)$. Recall that $P_2 - P_1 = \theta(\beta_2 - \beta_1)N_2/s$. Thus, for $\theta \varepsilon [1/2, 1)$, the amount by which P_2 exceeds P_1 is less than that for which the departure pattern is efficient. To achieve this, the toll must be set so as to discourage type 1 individuals from departing in $[t', t^*]$. This is done by having a discontinuous decrease in the toll at t'. Let the magnitude of the toll decrease be ρ and the size of the queue it induces be $D(t'_+)$. Now consider the choice of type 1 and type 2 individuals as to whether to depart just before the queue forms or in the queue. An individual who departs in the queue will experience an expected increase in travel time, and an expected decrease in schedule delay, of $D(t'_+)/2s$. Thus,

$$p_1(t'_-) = \tau(t'_-) + \beta_1(t^* - t') \tag{35a}$$

$$p_1(t') = \tau(t'_-) + \beta_1(t^* - t') + (\alpha - \beta_1)D(t'_+)/2s - \rho \tag{35b}$$

$$p_2(t'_-) = \tau(t'_-) + \beta_2(t^* - t') \tag{35c}$$

$$p_2(t') = \tau(t'_-) + \beta_2(t^* - t') + (\alpha - \beta_2)D(t'_+)/2s - \rho \tag{35d}$$

Comparison of (35a)-(35d) indicates that type 2 individuals are willing to pay a higher premium to join the queue since they value the expected decrease in schedule delay more highly. At the Ramsey optimum, the constraint that type 1 individuals not wish to depart in $[t', t^*]$ must bind. Also, for type 2 individuals to travel in $[t', t^*]$, arrivals must be continuous. These conditions imply:

$$p_1(t') = p_1(t'_-) = P_1 \tag{36a}$$

$$p_2(t') = p_2(t' + D(t'_+)/s) = P_2 \tag{36b}$$

Solving these equations along with $P_2 - P_1 = \theta(\beta_2 - \beta_1)N_2/s$ yields

$$D(t'_+) = 2(1 - \theta)N_2 \tag{37a}$$

$$\rho = (\alpha - \beta_1)D(t'_+)/2s \tag{37b}$$

The resulting equilibrium is depicted in Figure 6a.

Next we consider $\theta > 1$. The amount by which P_2 exceeds P_1 is now greater than that for which the departure pattern is efficient. Thus, the toll must be set so as to discourage type 2 individuals from departing prior to t'. There are two qualitatively different ways to discourage type 2 individuals. First, a $\theta > 1$ can be sustained by an increase in schedule delay costs without queuing. The toll would be set prohibitively high over an interval (t'', t') such that no individuals of either type depart during this period; and so that $P_2 = p_2(t'')$ and $P_1 = p_1(t'')$. Then

$$P_2 - P_1 = p_2(t'') - p_1(t'') = (\beta_2 - \beta_1)(t^* - t'') > (\beta_2 - \beta_1)N_2/s \tag{38}$$

Since $P_2 - P_1 = \theta(\beta_2 - \beta_1)N_2/s$ and $t^* - t' = N_2/s$,

$$t' - t'' = (\theta - 1)N_2/s \tag{39}$$

With N_1 and N_2 fixed, the increase in the schedule delay costs of type 1 individuals rises *linearly* with $(\theta - 1)$.

Second, a $\theta > 1$ can be sustained by inducing a queue with no increase in schedule delay. This entails a mass, m, of type 1 individuals departing at $t_m = t' - m/s$. Since the constraint that type 2 individuals not depart before t' must bind at the Ramsey optimum,

$$p_2(t_m) = \tau(t_m) + \beta_2(t^* - t_m) + (\alpha - \beta_2)m/2s = P_2 \tag{40}$$

Also

$$p_1(t_m) = \tau(t_m) + \beta_1(t^* - t_m) + (\alpha - \beta_1)m/2s = P_1 \tag{41}$$

Then a $\theta > 1$ can be sustained by a mass of type 1 individuals of size $m = 2(\theta - 1)N_2$ departing at $t_m = t' - m/s$. Queuing costs increase as the square of m and therefore, with N_2 fixed, as the *square* of $(\theta - 1)$.

Considering both cases suggests the cost-minimizing means of sustaining a given $\theta > 1$. For θ below some critical value, which turns out to be $1 + \gamma$, the desired separation is achieved by inducing a queue, but no increase in type 1 schedule delay costs. For $\theta > 1 + \gamma$, cost-minimizing separation is obtained by inducing both a queue and an increase in type 1 schedule delay costs.

4.2. *The Ramsey Problem*

We will now see the advantage of first solving the feasibility problem. It permits the following formulation of the Ramsey problem, which has a purely static structure.

The problem is to maximize the social surplus function

$$\int_0^{N_1} P_1(n_1)dn_1 + \int_0^{N_2} P_2(n_2)dn_2 - \Gamma(N_1, N_2, \theta(N_1, N_2, s), s) - K(s) \tag{42}$$

subject to the deficit constraint

$$K(s) - (P_1(N_1)N_1 + P_2(N_2)N_2 - \Gamma(N_1, N_2, \theta(N_1, N_2, s), s)) \leq z \tag{43}$$

and the feasibility constraint

$$\theta(N_1, N_2, s) \geq 1/2 \tag{44}$$

where z is the maximum allowable deficit,

$$\theta(N_1, N_2, s) = (P_2(N_2) - P_1(N_1))s/(\beta_2 - \beta_1)N_2 \tag{45}$$

and

$$\Gamma(N_1, N_2, \theta(N_1, N_2, s), s) = (\beta_1 N_1^2 + 2\beta_1 N_1 N_2 + \beta_2 N_2^2 + 4\alpha(1 - \theta(N_1, N_2, s))^2 N_2^2)/2s$$

$$\text{for} \quad 1/2 \leq \theta(N_1, N_2, s) \leq 1 + \beta_1 N_1/4\alpha N_2 \tag{46a}$$

$$= ((1 - \beta_1/4\alpha)\beta_1 N_1^2 + 2\theta(N_1, N_2, s)\beta_1 N_1 N_2 + \beta_2 N_2^2)/2s$$

$$\text{for} \quad \theta(N_1, N_2, s) \geq 1 + \beta_1 N_1/4\alpha N_2 \tag{46b}$$

The choice variables are N_1, N_2 and s.

It is easily checked that Γ is smooth, even at points where $\theta = 1$ and $\theta = 1 + \gamma$. We can therefore apply Kuhn-Tucker theory. From the Lagrangean function

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

$$+ \mu(z + P_1(N_1)N_1 + P_2(N_2)N_2 - \Gamma(N_1, N_2, \theta(N_1, N_2, s), s) - K(s))$$

$$+ \nu(\theta(N_1, N_2, s) - 1/2) \qquad (47)$$

we obtain first order conditions for N_1, N_2 and s that we write

$$\frac{P_i - d\Gamma/dN_i}{P_i} = -\frac{\mu}{1+\mu}\frac{1}{\eta_i} - \frac{\nu}{1+\mu}\frac{\partial\theta/\partial N_i}{P_i} \qquad i = 1, 2 \qquad (48)$$

$$d\Gamma/ds + K'(s) = \frac{\nu}{1+\mu}\partial\theta/\partial s \qquad (49)$$

where $\eta_i = (P_i/N_i)dN_i/dP_i$. In (48),

$$d\Gamma/dN_i = \partial\Gamma/\partial N_i + \partial\Gamma/\partial\theta \cdot \partial\theta/\partial N_i \qquad (50)$$

where $\partial\Gamma/\partial N_i$ is the partial derivative of Γ with respect to its ith argument. Similarily,

$$d\Gamma/ds = \partial\Gamma/\partial s + \partial\Gamma/\partial\theta \cdot \partial\theta/\partial s \qquad (51)$$

where $\partial\Gamma/\partial s$ is the partial derivative of Γ with respect to its last argument.

In what follows, it will be useful to distinguish between two cases according to whether the feasibility constraint is nonbinding ($\nu = 0$) or binding ($\nu > 0$).

Case I. $\nu = 0$. In this case, (48) and (49) reduce to

$$\frac{P_i - d\Gamma/dN_i}{P_i} = -\frac{\mu}{1+\mu}\frac{1}{\eta_i} \qquad i = 1, 2 \qquad (52)$$

$$d\Gamma/ds + K'(s) = 0 \qquad (53)$$

(52) is the basic Ramsey rule, while (53) is the first-best capacity rule.[19]

Case II. $\nu > 0$. In order for (48) to hold, we must now have

$$\frac{P_1 - d\Gamma/dN_1}{P_1} < -\frac{\mu}{1+\mu}\frac{1}{\eta_1} \tag{54}$$

$$\frac{P_2 - d\Gamma/dN_2}{P_2} > -\frac{\mu}{1+\mu}\frac{1}{\eta_2} \tag{55}$$

since, from (45), $\partial\theta/\partial N_1 > 0$, while $\partial\theta/\partial N_2 < 0$. (54) states that the percentage markup of price over marginal cost for group 1 is less than the markup implied by the basic Ramsey rule (52). (55) requires the opposite relationship to hold in the case of group 2.

Equation (45) also implies that $\partial\theta/\partial s > 0$. Thus, where (49) holds,

$$d\Gamma/ds + K'(s) > 0 \tag{56}$$

The optimal level of capacity is thus greater than that which minimizes cost.

We stated above that the cost function Γ is everywhere smooth. Particularily noteworthy is the fact that it is not kinked at points where $\theta = 1$. This suggests that, for any specification of the capacity cost and demand functions, it would only be over a set having zero measure in parameter space that the solution to the Ramsey problem did not involve some degree of departure inefficiency.

In what follows, the smoothness of Γ where $\theta = 1$ is used to prove a proposition that will lead to a particularization of the Ramsey problem for which the optimal departure pattern is definitely inefficient.

Proposition 1. In order for a solution to the Ramsey problem to involve departure efficiency, it must meet the condition

$$P_2/P_1 = \eta_2/\eta_1 \tag{57}$$

Proof. If a solution to the Ramsey problem involves $\theta = 1$, then $\nu = 0$, so that (52) holds. Using (52) for both $i = 1$ and $i = 2$,

$$\frac{P_2}{P_1} = \frac{P_2 - d\Gamma/dN_2}{P_1 - d\Gamma/dN_1} \frac{\eta_2}{\eta_1} \tag{58}$$

From (50) and the fact that $\partial\Gamma/\partial\theta = 0$ when $\theta = 1$,

$$d\Gamma/dN_i = \partial\Gamma/\partial N_i = \partial C/\partial N_i \tag{59}$$

From (58) and (59),

$$\frac{P_2}{P_1} = \frac{P_2 - \partial C/\partial N_2}{P_1 - \partial C/\partial N_1} \frac{\eta_2}{\eta_1} \tag{60}$$

(57) then follows from the fact (proved at the beginning of this section) that, for $\theta = 1$,

$$P_1 - \partial C/\partial N_1 = P_2 - \partial C/\partial N_2 = k \qquad \text{Q.E.D.}$$

In any equilibrium, the price ratio P_2/P_1 in (57) is greater than one. Thus, if both groups had constant elasticity demand functions with $\eta_2 \geq \eta_1$ (so that $|\eta_2| \leq |\eta_1|$), then there could not be an equilibrium for which (57) holds. This proves

Corollary 1. Suppose that the demand functions $N_1(P_1)$ and $N_2(P_2)$ are constant elasticity demand functions satisfying $\eta_2 \geq \eta_1$. Then the solution to the Ramsey problem involves an inefficient departure pattern.

Corollary 1 is the main result of the paper. It establishes that, at least for a subclass of the Ramsey problem, the optimal departure pattern is inefficient.

5. Concluding Comments

In this paper, we have investigated the Ramsey problem for a congestible facility employing a structural model which *derives* the cost function from consumers' time-of-use decisions and the congestion technology. The standard approach, in contrast, *assumes* the form of the cost function without deriving it.

We employed a stylized model of a particular congestible facility – bottleneck congestion on a road in the morning rush hour – to illustrate our approach. We showed that, at least in this context, the standard approach is misspecified because it assumes that the congestion cost function is fully determined by technology, whereas in fact it depends on both technology and the pricing scheme. We also showed that the Ramsey optimum for a congestible facility generally entails cost inefficiency (manifested in our simple model as a queue behind the bottleneck or inefficiently high schedule delay costs). This phenomenon has not been identified by the standard approach because of its misspecification.

We conjecture that our main qualitative result, that the Ramsey optimum for a congestible facility generally entails cost inefficiency, is general. In our example, we took social surplus as our maximand, but the result should hold whatever social welfare function is employed. Also, we examined a particular congestible facility which has a particular form of cost inefficiency. The result should also apply to other congestible facilities, though the form of cost inefficiency will differ depending on context; for example, with telephone traffic, more calls will be blocked than is consistent with cost minimization.

Finally, we examined a particular second-best problem – Ramsey pricing for a congestible facility. But the cost inefficiency result should hold for any second-best problem for a congestible facility with heterogeneous users and anonymous pricing.

The general implication of our paper is that, in second-best analyses of congestible facilities, the form of the congestion cost function should be derived rather than assumed, which requires the use of structural models that explicitly treat consumers' time-of-use decisions and the congestion technology.

Appendix

Proof of the negative profit result of Section 3. Applying Euler's Theorem to the function $C(N_1, N_2, s)$ gives

$$N_1 \partial C/\partial N_1 + N_2 \partial C/\partial N_2 + s\partial C/\partial s = C \tag{61}$$

(61) together with (17)-(19) imply

$$P_1N_1 + P_2N_2 - C = sK'(s) \tag{62}$$

Letting λ denote the elasticity of $K(s)$ with respect to s, we can rewrite (62)

$$P_1N_1 + P_2N_2 - C = \lambda K(s) \tag{63}$$

The left hand side of (63) is the public authority's toll receipts. Thus, when $\lambda < 1$, the public authority incurs a deficit. Q.E.D.

Proof of Properties 3 *and* 4. Suppose that $\theta \geq 1$, in which case f(t) (recall (32)) has its root to the left of t' at $\tilde{t} = t' + (1 - \theta)N_2/s$. Then (26) and (27) are satisfied by a pattern of departures in which type 2 individuals depart at a uniform rate of s over $[t', t^*]$, while type 1 individuals depart at the same rate s over $[\tilde{t} - N_1/s, \tilde{t}]$. We will refer to this departure pattern as the reference solution to (26) and (27). In what follows, the reference solution is used to identify the most efficient solution to (26) and (27).

A departure pattern will be said to be *cost-improving* if and only if (a) it satisfies (26) and (27) and (b) it results in lower user costs than the reference solution. From (b), the only way that a departure pattern can be cost-improving is if it satisfies

(C1) Type 1 arrivals are not completed until after $\tilde{t}$.

From (C1) and (26),

(C2) At some time $t_m < \tilde{t}$, there is a mass of departures by type 1 individuals which makes $D(t_{m+}) > f(t_m)$.

From (C2) and the fact that the rate at which f(t) decreases is s,

(C3) $d(t) > f(t)$ for all $t > t_m$.

Together with (26), (C3) implies that there are no type 1 departures after t_m.

We next show that, for any cost-improving departure pattern for which $D(t_{m-}) > 0$, there is a dominant cost-improving departure pattern for which $D(t_{m-}) = 0$. Let x be a cost-improving departure pattern for which $D(t_{m-}) = r > 0$ and the size of the departure mass at t_m is m. Let y be the departure pattern in which (1) there is a type 1 mass of size m + r at the same time t_m, (2) type 1 individuals who do not depart at t_m depart at a uniform rate of s over $[t_m - (N_1 - m - r)/s, t_m)$, and (3) the pattern of type 2 departures is the same as in x. y satisfies (26) and (27) (from the fact that x does) as well as $D(t_{m-}) = 0$. Queuing costs are lower for y than for x, while schedule delay costs are not greater. y thus dominates x and is cost-improving.

It follows immediately that any cost-improving departure pattern that does not satisfy

(C4) Type 1 individuals who do not depart at t_m depart at a uniform rate of s over $[t_m - (N_1 - m)/s, t_m)$, where m denotes the size of the departure mass at t_m.

is dominated by a cost-improving departure pattern for which (C4) holds. The same is true for cost-improving departure patterns that do not satisfy

(C5) Type 2 individuals depart at a uniform rate of s over $[t', t^*]$.

We now consider the class of departure patterns defined by

(1) At some time $t_m \leq \tilde{t}$, there is a mass of departures by type 1 individuals whose size we denote by m.

(2) Type 1 individuals who do not depart at t_m depart at a uniform rate of s over $[t_m - (N_1 - m)/s, t_m)$.

(3) Type 2 individuals depart at a uniform rate of s over $[t', t^*]$.

(4) $t_m + m/s \leq t'$ (type 1 arrivals completed by t').

(5) $m/2 \leq s(t^* - t_m) - \theta N_2$ ((26) imposed at t_m).

The reference solution is the member of this class for which $t_m = \tilde{t}$ and $m = 0$. This and the preceding results imply that the most efficient solution to (26) and (27) is contained in this class.

Given a departure pattern in this class, user costs are given by

$$\beta_1 N_1(t^* - t_m + N_1/2s - m/s) + \alpha m^2/2s + \beta_2 N_2^2/2s \tag{64}$$

Initially, we will ignore the constraint

$$t_m + m/s \leq t' \tag{65}$$

Minimizing (64) subject to

$$m/2 \leq s(t^* - t_m) - \theta N_2 \tag{66}$$

requires that (66) hold with equality ((64) is monotonically decreasing in t_m, and (66) imposes an upper bound on t_m for any given value of m). The optimal values of t_m and m are given by

$$t_m = t^* - (\theta N_2 + \beta_1 N_1/4\alpha)/s \tag{67}$$

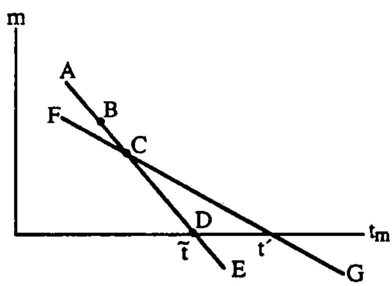
and

$$m = \beta_1 N_1/2\alpha \tag{68}$$

(67) and (68) are consistent with (65) if and only if $\theta \geq 1 + \gamma$, where $\gamma \equiv \beta_1 N_1/4\alpha N_2$. This and Lemma 1 imply (A) of Property 4. (B) of Property 4 follows from substituting (67) and (68) into (64).

What remains is to complete the analysis for the case $1 < \theta < 1 + \gamma$. The analysis is facilitated by the following diagram.

Figure 7



(66) constrains a (t_m, m) combination to lie on or below ABCDE, the slope of which is $-2s$. (65) constrains a (t_m, m) combination to lie on or below FCG, the slope of which is $-s$. When $\theta < 1 + \gamma$, (67) and (68) define a point like B. Since (64) is monotonically decreasing in t_m, an

optimal (t_m, m) combination must lie on FCD. Attention can further be restricted to CD, since any movement along FC towards C lowers queuing costs without affecting schedule delay costs. When $\theta < 1 + \gamma$, (64) increases monotonically as one moves along BD towards D. Costs are thus lowest over CD at C.

The optimal values of t_m and m are now

$$t_m = t' - 2(\theta - 1)N_2/s \tag{69}$$

and

$$m = 2(\theta - 1)N_2 \tag{70}$$

(69) and (70) imply (A) of Property 3. (B) of Property 3 follows from substituting (69) and (70) into (64). Q.E.D.

References

Arnott, R., de Palma, A. and Lindsey, R. (1988). "Information and Time-of-Use Decisions in Stochastically Congestible Facilities," Northwestern Center for Mathematical Studies in Economics and Management Science, Discussion Paper No. 788.

Arnott, R., de Palma, A. and Lindsey, R. (1989a). "Bottleneck Congestion with Elastic Demand," mimeo.

Arnott, R., de Palma, A. and Lindsey, R. (1989b). "Schedule Delay and Departure Time Decisions with Heterogeneous Commuters," mimeo.

Arnott, R., de Palma, A. and Lindsey, R. (1990a). "Departure Time and Route Choice for Routes in Parallel," *Transportation Research*, forthcoming.

Arnott, R., de Palma, A. and Lindsey, R. (1990b). "Economics of a Bottleneck," *Journal of Urban Economics*, forthcoming.

Auerbach, A.J. (1985). "The Theory of Excess Burden and Optimal Taxation," in A.J. Auerbach and M. Feldstein, eds., *Handbook of Public Economics*, Volume I, North-Holland, Amsterdam.

Baumol, W.J. and Bradford, D.F. (1970). "Optimal Departures from Marginal Cost Pricing," *American Economic Review*, 60, 265-283.

Boiteux, M. (1956). "Sur la Gestion des Monopoles Publics Astreints a l'Equilibre Budgetaire," *Econometrica*, 24, 22-40.

Bos, D. (1986). *Public Enterprise Economics*, North-Holland, Amsterdam.

Braeutigam, R.R. (1987). "Optimal Policies for Natural Monopolies," mimeo.

Braid, R.M. (1989). "Uniform versus Peak-Load Pricing of a Bottleneck with Elastic Demand," *Journal of Urban Economics*, 26, 320-327.

Bucovetsky, S. (1984). "On the Use of Distributional Waits," *Canadian Journal of Economics*, 17, 699-717.

Cohen, Y. (1987). "Commuter Welfare under Peak-Period Congestion Tolls: Who Gains and Who Loses?," *International Journal of Transport Economics*, 14, 239-266.

Crew, M.A. and Kleindorfer, P.R. (1986). *The Economics of Public Utility Regulation*, MIT Press, Cambridge.

de Palma, A. and Arnott, R. (1989). "The Temporal Use of a Telephone Line," mimeo.

Greenwald, B.C. and Stiglitz, J.E. (1986). "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics*, 101, 229-264.

Guesnerie, R. (1980). "Second-Best Pricing Rules in the Boiteux Tradition: Derivation, Review and Discussion," *Journal of Public Economics*, 13, 51-80.

Guesnerie, R. and Roberts, K. (1984). "Effective Policy Tools and Quantity Controls," *Econometrica*, 52, 59-86.

Mohring, H. (1970). "The Peak Load Problem with Increasing Returns and Pricing Constraints," *American Economic Review*, 60, 693-705.

Newell, G. (1987). "The Morning Commute for Non-identical Travellers," *Transportation Science*, 21, 74-88.

Ramsey, F.P. (1927). "A Contribution to the Theory of Taxation," *Economic Journal*, 37, 47-61.

Sah, R.K. (1987). "Queues, Rations, and Market," *American Economic Review*, 77, 69-77.

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

Smith, M.J. (1983). "The Existence and Calculation of Traffic Equilibria," *Transportation Research*, 17B, 291-303.

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.

Weitzman, M.L. (1977). "Is the Price System or Rationing More Effective in Getting a Commodity to Those Who Need it Most," *Bell Journal of Economics*, 8, 517-524.

Footnotes

1. Ramsey [1927], Boiteux [1956] and Baumol and Bradford [1970] are seminal papers related to this topic.
2. This is similar to the result in Arnott, de Palma and Lindsey [1989a] that the cost function depends on the form of pricing (e.g., uniform vs. optimal time-varying toll).
3. Vickrey allowed for a distribution of desired arrival times; in other respects, commuters were identical.
4. A pure strategy Nash equilibrium, with departure time as the strategy variable.
5. Heterogeneous users (Newell [1987], Cohen [1987], Arnott, de Palma and Lindsey [1989b]), elastic demand (Braid [1989], Arnott, de Palma and Lindsey [1989a]), stochastic capacity and demand (Arnott, de Palma and Lindsey [1988]), nonlinear costs (Smith [1983], Braid [1989]), simple networks (Arnott, de Palma and Lindsey [1990a]), intermediate toll regimes (Braid [1989], Arnott, de Palma and Lindsey [1990b, 1989a]), other congestible facilities (de Palma and Arnott [1989] on congestion in telephone traffic), and optimal capacity (Arnott, de Palma and Lindsey [1988, 1989a, 1989b, 1990a, 1990b]).
6. This is obvious without the deficit constraint. With the deficit constraint: Expenditure = user costs + toll revenues = $P_1N_1(P_1) + P_2N_2(P_2)$, where $N_1(P_1)$ and $N_2(P_2)$ are the demand functions of the two groups, while system costs = user costs + capacity costs. Since expenditure is fixed, minimizing system costs is equivalent to minimizing the deficit.
7. In the standard approach, the period of congestion is often divided into time intervals. Where $t \varepsilon T$ indexes these intervals, the cost function would then be written as $C(\{n_1^t\}_{t\varepsilon T}, \{n_2^t\}_{t\varepsilon T}, s)$, where n_i^t is the number of commuters in group i who use the congestible facility in interval t. Note that since the use of a congestible facility almost always takes time, there is ambiguity as to how the number of users in a time interval should be measured.
8. Our result on cost inefficiency at the Ramsey optimum is not altogether surprising. Consider a situation in which there are two groups who desire access to scarce medical facilities, the

provision of which is characterized by increasing returns. The two groups differ only in terms of an unobservable shadow value of waiting time, δ_1 and δ_2 respectively, with $\delta_1 > \delta_2$. If group 2's demand is more elastic than group 1's, then the Ramsey optimum will entail a pricing scheme whereby an individual has the choice of paying a higher price and incurring no wait (which group 1 would choose), or a lower price and incurring a wait (which group 2 would choose) -- cost inefficiency. The analogous phenomenon for traffic would be two parallel roads. Self selection would manifest itself in route choice -- spatial separation. In our model, with a single road and both travel time and schedule delay costs, the self selection manifests itself as temporal separation.

9. This is the problem solved in Arnott, de Palma and Lindsey [1990b].
10. This section extends to two consumer groups the elastic demand model in Arnott, de Palma and Lindsey [1989a].
11. This is equivalent to assuming that the shadow value of time late is infinite. This assumption is made to simplify the analysis, and does not alter the qualitative results.
12. There is strong evidence that $\alpha > \beta$ (see Small [1982]).
13. Allowing for different shadow values of queuing time or more than two commuter groups does not affect our basic result, but complicates the analysis.
14. In Sections 1-3, the optimal departure pattern is cost-minimizing. But in the Ramsey problem, the (second-best) optimal departure pattern is in general not cost-minimizing.
15. We employ this objective function since it is the standard one in the literature and since we wish to abstract from equity considerations.
16. We do not restrict this elasticity to be constant.
17. A slightly weaker assumption is that there are no type 1 departures after the time at which type 2 departures commence. This allows type 2 departures to begin with a mass of departures which includes type 1 as well as type 2 individuals. Allowing for this possibility does not affect the results, but greatly complicates the proofs. We therefore decided to employ the stronger assumption described in the text.

18. Any toll gradient having an equilibrium departure pattern that satisfies the "1's before the 2's condition" is admissible.

19. Note, however, that the marginal cost term $d\Gamma/dN_i$ in (52) and the cost derivative $d\Gamma/ds$ in (53) are generally associated with inefficiently high user costs for a given set of output levels N_1 and N_2 and a given capacity s.