Discussion Paper No. 974

INFORMATION AND USAGE OF CONGESTIBLE FACILITIES UNDER FREE ACCESS

by

Richard Arnott*
André de Palma**
Robin Lindsey***

January 1992

Financial support from the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, and NSF under grant SES-8912335 is gratefully acknowledged. Helpful comments were received from colleagues at the University of Alberta and participants at the 1988 Canadian Economic Theory Conference, the 1988 European Econometric Society meetings in Bologna and the 1988 North American Regional Science meetings in Toronto. Thanks are also due to Qin Liu for research assistance.

*Boston College, Chestnut Hill, Mass., U.S.A. 02167-3806

**Northwestern University, Evanston, IL., U.S.A. 60208-3109.

***University of Alberta, Edmonton, Alberta, Canada T6G 2H4

Abstract

We investigate the effect of information about congestion on participation and time-of-use decisions in a free-access delay system subject to predictable and unpredictable fluctuations in capacity and demand intensity. Expected welfare is greater with perfect than with zero information, while optimal design capacity is greater if and only if demand elasticity is less than one. Imperfect information can reduce welfare by inducing concentration in the arrival times of users at the facility. This suggests that route guidance systems for automobile travel and other public information dissemination schemes must be designed and implemented with care.

Notational Glossary

Greek characters

α	Unit cost of travel time
β	Unit cost of arriving early
γ	Unit cost of arriving late
$\Gamma(d)$	c.d.f of days
δ	Scale factor for average cost function
ε	Price elasticity of demand
η	Elasticity of average cost wrt level of usage
θ	n/σ
ν	Certainty-equivalent intensity of demand
$\rho(t)$	Density function of arrival times
σ	Ratio of realized capacity to design capacity
ϕ	N/s
$\phi(t)$	Largest ϕ such that queueing never occurs at time t
$\phi^*(t)$	ϕ such that user arriving at time t is served at t^*

English characters

$C(t)$	Cost for user arriving at facility at time t
CS	Consumers' surplus
d	Index of days
e	(superscript) denoting equilibrium value
F	(superscript) denoting full information regime
$F_d(n,\sigma)$	Joint c.d.f. of n and σ on day d
$G(n)$	c.d.f of n
$H(\sigma)$	c.d.f. of σ
I	(superscript) denoting imperfect information regime
$J(\phi)$	c.d.f. of ϕ from perspective of users
k	Marginal cost of capacity expansion
m	Index of messages
M	c.d.f. of m
MCS	Marginal consumers' surplus from expanding design capacity
n	Intensity of demand
N	Number of users
p	Price (= user cost) of using facility
$Q(t)$	Number of users in queue at time t
$Q(t,\phi)$	Number of users in queue at time t when $N/s = \phi$
$r(t)$	Arrival rate of users at time t
$R(t)$	c.d.f. of arrivals (integral of $\rho(t)$)
s	Realized capacity of facility
$\hat{s}$	Design capacity of facility
SB	Social benefit
t	Arrival time
t^*	Preferred time of usage

t_n	Full information regime: Arrival time for which user served at t^*
	Imperfect information regime: Latest arrival time for which user is never served late
t_0	Earliest arrival time
t_1	Latest service time
t_r	Latest arrival time
$T(t)$	Queueing time for user arriving at time t
$\bar{T}$	Time spent using facility in absence of a queue
TC	Total costs
U	(superscript) denoting user
U	Denotes event that an individual is a potential user
Z	(superscript) denoting zero information regime
$Z(t)$	Rate of increase wrt arrival time in expected queueing time
$Z^e(t)$	Rate of decrease wrt arrival time in expected early time
$Z^l(t)$	Rate of increase wrt arrival time in expected late time

A little learning is a dangerous thing

[Alexander Pope[1]]

1. Introduction

Congestion occurs when the social marginal cost of using a facility exceeds the private marginal cost (Walters (1987)). Congestion is fundamental to the theory of clubs and to peak-load pricing. Probably the most costly form of congestion is on urban roads, as reflected in the voluminous literature in economics and civil engineering on traffic congestion. But the costs of congestion are also apparent at airports, sea ports and with other modes of transport, at recreational facilities, in telecommunications, *etc.*

The purpose of this paper is to investigate how information about congestion affects individuals' participation and time-of-use decisions in free-access congestible facilities. The paper contributes to the peak-load pricing literature, which we now briefly review.

In the original peak-load model (Steiner (1957), Williamson (1966), Mohring (1970), *inter alios*) the timing and intensity of peak and off-peak demands are assumed known. Demand uncertainty was introduced by Brown and Johnson (1969), whose work encouraged a series of further papers. Uncertainty about capacity has been relatively neglected.

Despite the inherently dynamic nature of the peak-load problem, most theoretical models have been static. Empirical studies (notably of electricity demand) typically divide the period of use into time intervals, and capture intertemporal demand effects with parametric cross-price elasticities. This approach has several ambiguities (Arnott *et al.* (1991b)) which derive from a failure to model explicitly user behavior and the congestion technology of the particular facility.

The first systematic treatment of time-of-use decisions was undertaken by Vickrey (1969) in the context of morning rush hour traffic. Vickrey used a deterministic queueing model to describe the evolution of congestion. Though a number of authors have applied and extended his model[2], they have continued to work in a nonstochastic setting. Some work has been done in transportation in developing probabilistic choice models of travel time (*e.g.* Alfa and Minh (1979), de Palma *et al.* (1983, 1987)). Multinomial logit demand models have been estimated by Cosslett (1977), Abkowitz (1981a,b), Small (1982), Hendrickson and Plank (1984) and Moore, Jovanis and Koppelman (1984). Naor (1969) and De Vany (1976) have developed queueing models in which the service time of individuals is random. However, all these models deal only with stochasticity at the level of the individual. Stochasticity at the aggregate level has been modeled by De Vany and Saving (1977, 1980), Kraus (1982) and D'Ouville and McDonald (1990), but in a steady-state framework that abstracts from the peak-load problem. It is thus fair to say that the dynamic peak-load problem under uncertainty has yet to receive a definitive treatment.

Another gap in the peak-load literature is a lack of focus on the effect of information on individuals' usage decisions, in either a static or a dynamic setting. Where congestion is sensitive to the capacity utilization rate of a facility, and where users have discretion over their intensity and/or time of use, information can materially alter the time pattern of congestion.

Our interest in information and congestion derives from more than intellectual curiosity. Research has been underway for over a decade on the design and operation of Route Guidance Systems (RGS) for facilitating vehicular travel. The main function of RGS is to provide information to drivers on the location, timing and magnitude of congestion on the road network they will be traveling on. The information can be historical, in the sense that it concerns predictable patterns of congestion, or dynamic in the sense of real-time information about current driving conditions. Information can be conveyed in the form of a list of options amongst which the driver is free to choose, or as directions on when to depart, which route to take, *etc.*

A number of field experiments have taken place at various sites around the world on the impact of RGS information on congestion and accidents.[3] Simulation studies have also been conducted recently (Koutsopoulos and Lotan (1989), Sullivan and Wong (1989), Gardes and May (1990), Hamerslag and Van Berkum (1991), Mahmassani and Jayakrishnan (1991), *inter alios*). While these early modeling efforts have been insightful, most suffer from one or more limitations or conceptual problems. Some studies have focused on the impact of information on test vehicles that receive it while ignoring the general equilibrium effects of information on traffic overall. Other studies have overlooked the incentive compatibility constraint that drivers will follow advice only if they believe it will benefit them. A third problem is that the traffic engineering models employed are complex and require numerical solution, so that basic economic insights are obscured. The models are also specific to traffic flow, so that lessons cannot easily be transfered to other congestible facilities.

Given the very substantial costs of designing and implementing RGS or information systems in other types of facilities it is highly desirable to assess the potential benefits from them beforehand. One question that can be asked is whether the benefits are necessarily positive. To answer this it is useful to distinguish what traffic engineers call the user equilibrium and the system optimum. The user equilibrium is a Nash equilibrium under free access in which no driver can unilaterally do better by changing his behavior (decision to travel, departure time, route, *etc.*) The system optimum as typically formulated entails minimization of the sum of travel costs of a fixed set of users. Clearly, better information can only improve the system optimum. But in user equilibrium congestion is an uninternalized externality. It is conceivable that information induces changes in user behavior that increase the deadweight loss associated with unpriced congestion. Information is then welfare-reducing.

It is well known that the value of information to a single, rational, agent cannot be negative in a decision-theoretic context (Marschak (1954)). However, instances where better information can be disadvantageous, either to the agent that receives it or to society at large, are known in the literature. These can be grouped into psychological, strategic and transactional situations. A *psychological* preference for ignorance exists, according to Drèze (1987, p.108), when "the psychological contents of a piece of information ... prove so harmful as to offset the advantages of flexibility in choosing a course of action". The "Curse of Knowledge", whereby agents with superior information are unable to ignore it when trying to forecast the decisions of less well-informed agents (Camerer *et al.* (1989)), also fits this category. A *strategic* preference for ignorance can arise in a game setting when preplay communication would be harmful, or where ignorance enables a player to commit to an advantageous course of action (see, for example, Gal-Or (1988)).

Finally, information is undesirable when it destroys the opportunity for certain *transactions*, such as risk-sharing. As Hirshleifer (1971) has shown, all agents may suffer if the information becomes public, while if it is private, the gains to the owner can be outweighed by the losses of others.

The mechanism that concerns us by which information may be welfare-reducing is similar to Hirshleifer's public information, although in free-access facilities market forces are absent. In fact, there are two ways in which information could be counterproductive. The first, termed *overreaction* by Ben-Akiva and de Palma (1991), occurs when too many individuals react to information, by switching routes for example. Such out-of-equilibrium behavior has been produced in a simulation model by Mahmassani and Jayakrishnan (1991) in the case of one-shot adjustments, and by Ben-Akiva *et al*. (1986) in the case of oscillatory behavior. Information can also have perverse results through *concentration* when it induces users to make more similar spatial and/or temporal choices *in the equilibrium of a repeated game*. An example of concentration will be given in Section 3.

Before turning to the analysis it is useful to describe its scope and our modeling approach. First, we are concerned with delay systems, in which all users who desire service receive it, but at a quality that degrades with the level of usage. Automobile transportation, indoor and outdoor recreational activities, and interactive computer systems are examples of delay systems. Depending on the facility and the circumstances, declining quality can manifest itself in queues, flow congestion or crowding. By contrast, loss systems provide all-or-nothing service at a (generally) constant quality. Electricity, natural gas and telephone service are examples.

Travelers and recreationists may choose not to participate if congestion is particularly heavy, and in this respect transportation and recreational facilities may behave like loss systems. However, individuals often make a substantial commitment in time and/or money to use a facility, and will follow through with their plans even if they regret having made them. (Of course, in a traffic jam drivers have no choice but to put up with congestion.) We shall rule out balking, which means that individuals base their participation decision on the (unconditional) expected cost of usage.

Second, we are concerned with free-access facilities. Roads and recreational facilities are the main examples. Most roads are publicly provided and free of charge, at least in North America. Despite numerous studies in favour of road pricing, political opposition on distributional and other grounds has prevented it. Electonic road pricing is now receiving increasing support, and is being experimented with in several countries. But it will be some years, if at all, before pricing is widespread. Similarly, admission to recreational areas in the U.S. has traditionally been provided either free, or at a nominal fee that remains constant for long periods.[4] Resistance to price-rationing derives in part from concern about equity and the cost of collecting fees at sites with low usage rates or multiple access points (Walsh (1986)).

Since roads and recreational facilities are both delay systems and (usually) free-access, our analysis is directly primarily at them.[5] In fact there is a strong parallel between highways and recreational areas, both in the physical interaction between users and in problems of operations management (Sandler and Tschirhart (1980, Table 2)).[6] And in line with the development of RGS attempts to promote efficient usage of recreational areas have been made by disseminating information to users (Schechter and Lucas (1978)).[7]

Finally, a word on modeling uncertainty. One approach is to assume that nature is deterministic, but individuals are imperfectly informed about demand and/or capacity. Information is assumed to reduce the variance of their perception errors. Koutsopoulos and Lotan (1989) and Mahmassani and Jayakrishnan (1991) took this approach in modeling route guidance systems. Haltiwanger and Waldman (1985) adopted a variant in which there are two groups of individuals: the "sophisticated", who have rational expectations and correct beliefs, and the "naive", who have "pure limited RATEX" and hold the same, incorrect, beliefs.

The alternative view is that nature is stochastic. In modeling RGS this has been done either by assuming travel times on links of the road network are variable (*e.g.* Tsuji *et al.* (1985)) or by modeling incidents and other shocks explicitly (Gardes and May (1990)).[8] For two reasons we have taken the stochastic view. First, traffic studies have found that nonrecurring congestion (due to incidents, traffic signal failures, *etc.*) contributes as much or more to traffic delay than does recurring congestion (Ju *et al.* (1987), Lindley (1987), OECD (1988)). Second, regular users of a facility (such as commuters) are likely to be familiar with patterns of recurring congestion, and would not learn much from information broadcast about it.

Our analysis is a general one in allowing for fluctuations in both demand and capacity, and for both predictable (recurring) and unpredictable (nonrecurring) fluctuations. There are two limitations that deserve mention. First, demand and capacity are assumed to remain constant during the period of use, which means that there is is no role for, say, information received after participation decisions are made but before time of use. Second, attention is restricted to a single facility. There is no role for information received after participation decisions are made but before choice of facility, or after choice of facility but before time of use, and *vice versa*. Also bypassed is the question of how policies adopted at one facility can affect the deadweight loss from unpriced congestion at other facilities.

Our analysis consists of two parts. In Section 2 we adopt a static model in which only the decision *whether* to use a facility is considered, not *when*. This approach is reasonable if individuals lack strong time-of-use preferences, or if service entails a 'batch' operation, as is the case with entertainment or sporting events, or once-a-year parades. As shown in Section 3 the static model also serves as a reduced-form of our dynamic model when users have full information, so that all the results go through when time-of-use is made endogenous.

On the assumption that user costs are homogeneous of degree zero in demand and capacity and that demand is isoelastic we derive the optimal design capacity for a facility, and show how it varies with the joint probability distribution of demand intensity and capacity availability. We also show that information is welfare-improving in the sense that expected consumers' surplus with information is greater than with zero information. However, this result is not robust to the functional form of the cost and demand curves.

In Section 3 we extend consideration to time-of-use decisions by adopting Vickrey's (1969) model of queueing behind a bottleneck. Full information again turns out to be welfare-improving but, in contrast to the static model, imperfect information need not be.

2. Static Equilibrium

2.1 The Model

The static equilibrium model is based on conventional supply and demand curves. Following Arnott *et al.* (1991b) demand is assumed to be isoelastic:

$$N = np^{-\varepsilon}, \tag{2.1}$$

where n is a parameter characterizing demand intensity and p is the 'price' or full cost of using the facility. The constant price elasticity assumption is crucial to the analysis. Though it is somewhat restrictive, it need only hold within the range of price variation; not necessarily 'near the axes'. Average user cost is assumed to be homogeneous of degree zero in the ratio of demand to capacity:

$$AC = \delta\left(\frac{N}{s}\right)^{\eta}, \tag{2.2}$$

where s is capacity and $\eta > 0$ characterizes the rate at which costs rise as usage increases, capacity given.[9] Capacity is presumed to be available in homogeneous units. At this point we are agnostic as to whether rising costs are due to queueing, flow congestion, crowding or some other type of interference between users. Queueing congestion is treated explicitly in Section 3.

N and s are assumed to fluctuate over time. The fluctuations can range from being fully predictable to completely unpredictable. In the case of recreational capacity, seasonal openings and closures of hiking trails and campsites may follow a fixed timetable well known to prospective visitors. Traffic lane closures due to maintenance and repairs are predictable if publicized well in advance, and completed according to schedule. Seasonal variations in lighting that affect travel and various outdoor activities are predictable to an extent, as is bad weather (rain, snow, fog, flooding *etc.*), depending on the accuracy of weather forecasts and how speedily and widely they are disseminated. In contrast, accidents, vehicular breakdowns and signal failures are unpredictable, as are mechanical breakdowns in public utilities and other facilities.

Turning to fluctuations in demand it is typically the case that the demands for travel, communications, recreation, utilities and other services have regular daily, weekly, seasonal and/or holiday cycles. Indeed, it might be argued that demand should be *fully* predictable. On the one hand, if individual usage decisions are statistically independent, then by the law of large numbers fluctuations in aggregate demand will be insignificant in proportional terms.[10] On the other hand, if individual usage decisions are governed by common causal factors, then knowledge of these factors should be enough to predict total usage. In either case, there should be no demand surprises.

Yet surprises do occur in reality. Previous attendance at annual events can be forgotten, or become obsolete through changing demographics, incomes, tastes *etc.* And for relatively rare or distinctive events such as transit strikes, periods of gasoline rationing or epidemics, or for one-off events such as world fairs, there may be little or no precedent on which to base predictions.[11]

To formalize these ideas we assume for concreteness that the day is the appropriate unit of time measurement, and that demand intensity and capacity are determined before daily participation decisions are made. The realized values of demand intensity and capacity are assumed constant until the following day. Fluctuations in capacity are modeled by writing $s = \sigma\hat{s}$, where

$\hat{s}$ is design capacity (capacity under ideal conditions) and $\sigma \leq 1$ is the fraction of capacity available. Fluctuations in demand are captured through changes in demand intensity, n. n and σ are thus random variables.

To allow for systematic fluctuations in n and σ we let d be an index of 'days', with each d corresponding to a distinct frequency distribution of n and σ. In the case of commuting traffic, which has a predominantly weekly cycle, a suitable choice for d is a discrete index with seven possible values, one for each day of the week. In other applications it may be appropriate to treat d as continuous. We let $\Gamma(d)$ denote the cumulative distribution function of d, and $F_d(n,\sigma)$ be the joint c.d.f. of n and σ on day d. Both $\Gamma(d)$ and $F_d(n,\sigma)$ are assumed to be upper hemicontinuous (they can have mass points). The support of the distribution of n/σ defined by $F_d(n,\sigma)$ is assumed to have a finite upper bound. Finally, $\Gamma(d)$ and $F_d(n,\sigma)$ are assumed to be common knowledge among users.

We will be concerned with three informational regimes: perfect (or full) information, zero information and imperfect information. Under full information, users learn the precise realizations of n and σ before making their participation decisions. Under zero information they learn nothing, and have only $F_d(n,\sigma)$ to go on. Both full and zero information are limiting cases of imperfect information, which we discuss next.

In the imperfect information regime there is a message system that reports daily on demand and capacity after they are realized, but before usage decisions are made. This message system is an idealized representation of weather forecasts, early morning traffic reports, reports on ski conditions, *etc.*[12] Let m be an index of messages, and $M(m|d,n,\sigma)$ be the c.d.f. of m conditional on d and the realization (n,σ). The probability of message m, conditional only on d, is

$$dM(m|d) = \int_n \int_\sigma dM(m|d,n,\sigma)\ dF_d(n,\sigma). \tag{2.3}$$

(Where limits on the range of integration are omitted, the full range of variation is implicit.) Messages convey, explicitly or implicitly, information about n and/or σ in the form of a c.d.f. $\hat{F}_{dm}(n,\sigma)$. It is assumed that this distribution coincides with the true probability distribution of n and σ conditional on d and m, $F_{dm}(n,\sigma)$:

$$\hat{F}_{dm}(n,\sigma) = F_{dm}(n,\sigma). \tag{2.4}$$

Condition (2.4) can be viewed as a condition of unbiasedness, or irrefutability, in the sense that the frequency that a given state occurs on a particular day, when a particular message is announced, coincides with its forecast frequency. This seems reasonable in the long run, once forecasters have had the opportunity to correct biases in their forecasts, and users to compare the frequencies of forecasts and realizations.

By Bayes' theorem, the joint probability on day d of state (n,σ) and message m can be written

$$\Pr(n,\sigma,m|d) = dM(m|d,n,\sigma)\ dF_d(n,\sigma) = dF_{dm}(n,\sigma)\ dM(m|d), \tag{2.5}$$

whence

$$dF_{dm}(n,\sigma) = \frac{dM(m|d,n,\sigma)\ dF_d(n,\sigma)}{dM(m|d)}. \tag{2.6}$$

Substituting (2.6) into (2.4) and using (2.3) one obtains

$$d\hat{F}_{dm}(n,\sigma) = dF_{dm}(n,\sigma) = \frac{dM(m|d,n,\sigma)\ dF_d(n,\sigma)}{\int_{n'}\int_{\sigma'} dM(m|d,n',\sigma')\ dF_d(n',\sigma')}.$$

This implies

$$d\hat{F}_d(n,\sigma) = dF_d(n,\sigma),$$

that is, states are forecast with the same probability they occur in reality.[13] Since assumption (2.4) is maintained throughout the paper, the superscript ^ is henceforth omitted.

The probability function $dF_{dm}(n,\sigma)$ has been defined from the perspective of the analyst or outside observer; that is, someone who monitors the facility at all times and in all states. A subtle but crucial point is that this distribution generally differs from the corresponding distribution $dF^{u}_{dm}(n,\sigma)$ of the user. This fact is well known in queuing theory.[14] The reason in the present context is that the number of users varies from day to day.

Before showing this formally it is helpful to consider a stylized example. Suppose that buses on a certain route travel either full or half full, with equal probability. From the perspective of an outside observer (*e.g.* the bus company) the two states do indeed each occur with probability 1/2. But for bus riders, the respective probabilities of traveling on a full bus and on a half-full bus are 2/3 and 1/3, because twice as many people travel on full buses. The user's probability distribution is weighted toward high demand states. Of course, users may know which state is in effect because of regularity in their travel behavior vis à vis other riders. This can be captured in the model by making the probability distributions of demand day-specific.

We now derive the relationship between the probability distribution of the outside observer and that of the user. Let U denote the event that on a given day a particular individual is a user (or, more precisely, a potential user if demand is price-sensitive). If n_p is the number of potential users in the population then $\Pr(U|n,\sigma) = \frac{n}{n_p}$ (which is independent of σ). The joint probability of (n,σ) and U is by Bayes' rule

$$\Pr(n,\sigma,U|d,m) = \frac{n}{n_p}\, dF_{dm}(n,\sigma) = dF^u_{dm}(n,\sigma)\, \Pr(U|d,m), \tag{2.7}$$

where

$$\Pr(U|d,m) = \int_n \int_\sigma \Pr(U|n,\sigma)\, dF_{dm}(n,\sigma).$$

Let $\bar{n}_{dm}$ denote the expected value of n conditional on d and m. Then

$$\bar{n}_{dm} \equiv n_p \Pr(U|d,m),$$

and (2.7) can be written

$$dF^u_{dm}(n,\sigma) = \frac{n}{\bar{n}_{dm}}\, dF_{dm}(n,\sigma). \tag{2.8}$$

As true of the bus example, the user's probability distribution is weighted toward high-demand intensity states. Of course, with perfect information $\bar{n}_{dm} = n$, and the distributions of the outside observer and the user coincide.

2.2 Full Information

In this section we first solve for equilibrium price and level of usage for given realizations of demand and capacity, on the assumption that the realizations are known to users. Then we derive optimal design capacity. Finally, we examine how optimal design capacity varies with the joint distribution of n and σ.

Let n and s be the demand intensity and capacity realized on a given day. Absent admission charges, the average cost borne by users equals the full price of usage, p. Equilibrium price, p^e, and usage, N^e, can hence be solved by equating p and AC, given respectively by the inverse demand curve defined by (2.1), and (2.2):

$$p^e = \left(\frac{n}{N^e}\right)^{1/\varepsilon} = AC = \delta\left(\frac{N^e}{s}\right)^{\eta}. \tag{2.9}$$

This yields

$$p^e(n,s) = \delta^{\frac{1}{1+\varepsilon\eta}} \left(\frac{n}{s}\right)^{\frac{\eta}{1+\varepsilon\eta}}, \tag{2.10}$$

$$N^e(n,s) = \delta^{-\frac{\varepsilon}{1+\varepsilon\eta}} n^{\frac{1}{1+\varepsilon\eta}} s^{\frac{\varepsilon\eta}{1+\varepsilon\eta}}. \tag{2.11}$$

Consumers' surplus in state (n,s) is simply

$$CS(n,s) = \int_{p^e} N(p)dp.$$

Expected consumers' surplus, taking into account both systematic and unpredictable variations in demand and capacity, is

$$\overline{CS}^F(\hat{s}) = \int_d \int_n \int_\sigma \int_{p^e} np^{-\varepsilon}dp\ dF_d(n,\sigma)\ d\Gamma(d), \tag{2.12}$$

where superscript F denotes full information. The increase in consumers' surplus from a marginal capacity expansion, which we call marginal consumers' surplus, is

$$\overline{MCS}^F(\hat{s}) \equiv \frac{d\overline{CS}^F(\hat{s})}{d\hat{s}}.$$

We now assume that $F_d(n,\sigma)$ is independent of design capacity.[15] Then, from (2.10), and (2.12):

$$\overline{MCS}^F(\hat{s}) = \frac{\eta}{1+\varepsilon\eta}\delta^{\frac{1-\varepsilon}{1+\varepsilon\eta}} (\hat{s})^{-\frac{(1+\eta)}{1+\varepsilon\eta}} \int_d \int_n \int_\sigma n^{\frac{1+\eta}{1+\varepsilon\eta}} \sigma^{\frac{\eta(\varepsilon-1)}{1+\varepsilon\eta}} dF_d(n,\sigma)\ d\Gamma(d). \tag{2.13}$$

If capacity costs are linear, with constant[16] marginal cost k, then optimal capacity is given by

$$\hat{s}^F_* = \text{argmax} \left[\overline{CS}^F(\hat{s}) - k\hat{s}\right].$$

Using the first order condition $\overline{MCS}^F(\hat{s}) = k$ and (2.9) one obtains

$$\hat{s}^F_* = \left[\frac{1+\varepsilon\eta}{\eta} k\right]^{-\frac{1+\varepsilon\eta}{1+\eta}} \delta^{\frac{1-\varepsilon}{1+\eta}} \nu, \tag{2.14a}$$

where

$$\nu \equiv \left[\int_d \int_n \int_\sigma n^{\frac{1+\eta}{1+\varepsilon\eta}} \sigma^{\frac{\eta(\varepsilon-1)}{1+\varepsilon\eta}} dF_d(n,\sigma)\ d\Gamma(d) \right]^{\frac{1+\varepsilon\eta}{1+\eta}} . \tag{2.14b}$$

ν can be interpreted as a certainty-equivalent intensity of demand, *i.e.* the constant demand intensity for which under ideal conditions (σ = 1) design capacity is $\hat{s}_*^F$. In the limiting case where neither systematic nor random fluctuations in capacity or demand occur, ν reduces to n.

While the model underlying (2.14) is obviously a simple one, it has the merit of incorporating both demand and cost parameters, as well as systematic and nonsystematic fluctuations in demand and capacity. This contrasts with some formulas, or rather rules of thumb, used in capacity design. A good illustration of this is in the design of road capacity as described by the Institute of Transportation Engineers (1982) - hereafter ITE, and the Highway Capacity Manual of the Transportation Research Board (1985) - hereafter HCM.[17]

According to HCM (Section 2) it is standard practice in the U.S. to design roads to provide a given quality, or 'level', of service (there are 6 levels, ranked from A to F) at a particular rate of usage. For urban roads the usage rate typically chosen is that which occurs at the 10th or 20th busiest hour of the year. For rural roads the critical hour is higher (*e.g.* the 50th busiest hour) because of the relatively low traffic experienced on such roads over much of the year. Short-term variations in demand, such as those experienced during a rush hour, are accounted for by use of a peak-hour factor (PHF), defined as the ratio of the total volume occurring during an hour to the peak flow (expressed as an hourly flow rate) during a shorter time period, typically between 5 and 15 minutes (ITE, p.475). The PHF, which by definition is ≤ 1, is used to scale down the peak measured service volume to a more realistic measure of sustainable design capacity.

This procedure for capacity design is flawed in several respects. First, it allows for demand variability only to the extent of considering a particular fractile of the demand distribution. In contrast, the formula given in (2.14) indicates that the whole demand distribution is relevant.[18] Second, there is no clear indication in ITE or HCM how the target level of service is chosen. Third, the design procedures ignore capacity fluctuations. Incidents are a major cause of road capacity reductions, and as noted may contribute more to delay than recurrent peak period congestion. Yet the design of roads to meet a given level of service is based on the assumption of no incidents (HCM, 6-10). Road capacity is also affected by weather. Several studies have found that precipitation reduces effective capacity appreciably. To a lesser extent, so does poor visibility. Yet the HCM procedures do not take weather into account, though the manual does recommend this be done in areas where bad weather is common (HCM, 2-10, 6-15).[19]

The effect of capacity and demand fluctuations on optimal design capacity as given in (2.14) is summarized in the following proposition (for a precise statement and proof see Appendix 1):

PROPOSITION 1: If $\varepsilon < 1$ then, for any mean demand intensity, optimal design capacity is greater with variable capacity availability than when full capacity is always available. Furthermore, optimal design capacity is the larger: a) the greater the variability in demand, b) the greater the variability in capacity availability, c) the lower the ratio of mean capacity to design capacity, and d) the lower the correlation between demand and capacity. The opposite is true of all the above when $\varepsilon > 1$.

Prop. 1 establishes that if demand for trips is relatively inelastic (which is the case in most commuting contexts)[20] greater investment in capacity is warranted if loss of capacity can occur. There are two opposing forces at work. On the one hand, since only a fraction of design capacity is sometimes available, the marginal cost of constructing *working* capacity is greater in the stochastic case. On the other hand, the expected marginal benefit from working capacity is greater if the marginal value of usage declines with level of use. If $\varepsilon < 1$, the second factor dominates, and optimal design capacity is greater in the stochastic case.[21]

Prop. 1 also states that, if $\varepsilon < 1$, optimal design capacity increases with a mean-preserving spread (MPS) in the distribution of demand intensity[22] or a MPS in the distribution of capacity availability. Capacity is also greater the more states of *high* demand tend to coincide with *low* capacity, and *vice versa*.

2.3 Imperfect Information

In the preceding section it is assumed that users learn the exact realizations of demand and capacity before making decisions. In this section we consider the more realistic situation in which they know the systematic fluctuations, but have only imperfect information about nonsystematic fluctuations. First we solve for the expected equilibrium price on a given day with a given message. Next we solve for optimal design capacity, and recover the limiting cases of full information and zero information. Finally, we rank optimal design capacities and consumer welfare under imperfect information and zero information.

Let $\bar{p}^u_{dm}$ denote the price expected by a prospective user when message m is received on day d. If the realized demand intensity is n, the number of users is

$$N = n\left(\bar{p}^u_{dm}\right)^{-\varepsilon}. \tag{2.15}$$

Let $p(n,\sigma|d,m)$ be the resulting price given realizations n and σ. By (2.9) this is

$$p(n,\sigma|d,m) = \delta\left(\frac{N}{s}\right)^{\eta}, \tag{2.16}$$

which depends on d and m because N does (*viz.* equation (2.15)). Substituting (2.15) into (2.16):

$$p(n,\sigma|d,m) = \delta\left(\frac{n\left(\bar{p}^u_{dm}\right)^{-\varepsilon}}{\sigma\hat{s}}\right)^{\eta}. \tag{2.17}$$

Now by (2.8)

$$\bar{p}^u_{dm} = \int_n\int_\sigma p(n,\sigma|d,m)\ dF^u_{dm}(n,\sigma) = \int_n\int_\sigma p(n,\sigma|d,m)\ \frac{n}{\bar{n}_{dm}}dF_{dm}(n,\sigma). \tag{2.18}$$

Substituting (2.17) into (2.18)

$$\bar{p}^u_{dm} = \int_n\int_\sigma \delta\left(\frac{n\left(\bar{p}^u_{dm}\right)^{-\varepsilon}}{\sigma\hat{s}}\right)^{\eta}\frac{n}{\bar{n}_{dm}}\ dF_{dm}(n,\sigma),$$

which resolves to

$$\bar{p}^u_{dm} = \left[\frac{\delta(\hat{s})^{-\eta}}{\bar{n}_{dm}}\int_n\int_\sigma n\ \left(\frac{n}{\sigma}\right)^{\eta}dF_{dm}(n,\sigma)\right]^{\frac{1}{1+\varepsilon\eta}}. \tag{2.19}$$

Consumers' surplus is more tricky to evaluate than under full information. The number of users, given by (2.15), is proportional to n, and independent of σ. Gross social benefit SB equals the area under the inverse demand curve for demand intensity n, which is the sum of the areas above and below $\bar{p}^u_{dm}$. Integrating, one obtains

$$\overline{SB}^{I} = \int_d \int_m \left[\int_n \int_\sigma \left\{ \int_{\bar{p}^u_{dm}} np^{-\varepsilon} dp + n\left(\bar{p}^u_{dm}\right)^{1-\varepsilon} \right\} dF_{dm}(n,\sigma) \right] dM(m|d)\, d\Gamma(d), \tag{2.20}$$

where the superscript I denotes imperfect information. From gross social benefit must be deducted expected costs, $\overline{TC}^{I}$, to arrive at net expected consumers' surplus. Now

$$\overline{TC}^{I} = \int_d \int_m \overline{TC}^{I}_{dm}\, dM(m|d)\, d\Gamma(d), \tag{2.21}$$

where

$$\overline{TC}^{I}_{dm} = \int_n \int_\sigma TC_{dm}(n,\sigma|d,m)\, dF_{dm}(n,\sigma) = \int_n \int_\sigma p(n,\sigma|d,m)\, N(n,\sigma|d,m)\, dF_{dm}(n,\sigma)$$

$$= \int_n \int_\sigma p(n,\sigma|d,m)\, n\left(\bar{p}^u_{dm}\right)^{-\varepsilon} dF_{dm}(n,\sigma)$$

$$= \int_n \int_\sigma p(n,\sigma|d,m) \left(\bar{p}^u_{dm}\right)^{-\varepsilon} \bar{n}_{dm}\, dF^u_{dm}(n,\sigma) \qquad \text{(by (2.8))}$$

or by (2.18)

$$\overline{TC}^{I}_{dm} = \bar{n}_{dm}\left(\bar{p}^u_{dm}\right)^{1-\varepsilon}. \tag{2.22}$$

Substituting (2.22) into (2.21) and subtracting the result from (2.20) yields finally

$$\overline{CS}^{I} = \int_d \int_m \left[\int_n \int_\sigma \left\{ \int_{\bar{p}^u_{dm}} np^{-\varepsilon} dp \right\} dF_{dm}(n,\sigma) \right] dM(m|d)\, d\Gamma(d). \tag{2.23}$$

Define expected marginal consumers' surplus with imperfect information as

$\overline{MCS}^{I}(\hat{s}) \equiv \dfrac{d\overline{CS}^{I}(\hat{s})}{d\hat{s}}$. Differentiating[23], and using (2.19):

$$\overline{MCS}^{I}(\hat{s}) = \frac{\eta}{1+\varepsilon\eta} \frac{1}{\hat{s}} \int_d \int_m \bar{n}_{dm}\left(\bar{p}^u_{dm}\right)^{1-\varepsilon} dM(m|d)\, d\Gamma(d). \tag{2.24}$$

Following the same procedure as earlier, optimal capacity is found to be

$$\hat{s}^{I}_{*} = \left[\frac{1+\varepsilon\eta}{\eta}\, k\right]^{-\frac{1+\varepsilon\eta}{1+\eta}} \delta^{\frac{1-\varepsilon}{1+\eta}}\, \nu, \tag{2.25a}$$

$$\nu = \left[\int_d \int_m \bar{n}_{dm} \left(\bar{n}^{u}_{dm} \right)^{\frac{\varepsilon-1}{1+\varepsilon\eta}} \left\{ \int_n \int_\sigma \left(\frac{n}{\sigma} \right)^{\eta} n dF_{dm}(n,\sigma) \right\}^{\frac{1-\varepsilon}{1+\varepsilon\eta}} dM(m|d) d\Gamma(d) \right]^{\frac{1+\varepsilon\eta}{1+\eta}}. \qquad (2.25b)$$

As before, ν can be interpreted as a certainty-equivalent demand intensity. In the case of perfect information, (2.25b) reduces to the corresponding formula for full information, (2.14b). To see this, note that a perfect message system maps messages into unique states $M(m|d)$: $m \rightarrow (n,\sigma)$. The integral with respect to m in (2.25b) is thus replaced by a double integral with respect to n and σ, while the double integral inside the braces disappears.

The opposite extreme of zero information obtains when messages are completely uninformative. Optimal design capacity is given by (2.25) with the integral with respect to m removed:

$$\hat{s}^{Z}_{*} = \left[\frac{1+\varepsilon\eta}{\eta} k \right]^{-\frac{1+\varepsilon\eta}{1+\eta}} \delta^{\frac{1-\varepsilon}{1+\eta}} \left[\int_d \bar{n}_d \left(\bar{n}^{u}_{d} \right)^{\frac{\varepsilon-1}{1+\varepsilon\eta}} \left[\int_n \int_\sigma \left(\frac{n}{\sigma} \right)^{\eta} n dF_d(n,\sigma) \right]^{\frac{1-\varepsilon}{1+\varepsilon\eta}} d\Gamma(d) \right]^{\frac{1+\varepsilon\eta}{1+\eta}}, \qquad (2.26)$$

where the superscript Z denotes the zero information regime.

2.4 Comparison of the Information Regimes

In this subsection we rank information regimes with respect to optimal design capacity and efficiency. Regime A is said to be more efficient than regime B if

$$\overline{CS}^{A}(\hat{s}) > \overline{CS}^{B}(\hat{s}) \quad \forall \ \hat{s} > 0.$$

To rank efficiency, use is made of the following proposition (proved in Appendix 2):

PROPOSITION 2: If $\varepsilon < 1$ (resp. $\varepsilon > 1$) the more efficient information regime has a lower (resp. higher) marginal consumers' surplus for *any* design capacity.

Since a lower marginal consumers' surplus implies a lower optimal design capacity, with $\varepsilon < 1$ (resp. $\varepsilon > 1$) the more efficient regime has the lower (resp. higher) design capacity. Both the efficiency and optimal design capacities of two information regimes can thus be ranked by comparing the respective marginal consumers' surpluses from capacity expansion.

It turns out that an unambiguous ranking obtains when one of the regimes is zero information. This is formalized in:

PROPOSITION 3: Imperfect information increases efficiency relative to zero information. Furthermore, if $\varepsilon < 1$ (resp. $\varepsilon > 1$) then optimal design capacity is smaller (resp. larger) with information than without it.

Prop. 3 is proved in Appendix 3. It establishes that if demand is inelastic, information reduces the benefit from capacity expansion. As is the case with Prop. 1 there are two opposing forces at work. The fact that information increases efficiency relative to zero information tends to lower the benefit of investment. But information also reduces the efficiency loss from latent demand due to unpriced congestion. If $\varepsilon < 1$ the first factor dominates, and if $\varepsilon > 1$ the second does.

Prop. 3 is analogous to a result derived by Arnott *et al.* (1991b). There, we applied the Vickrey queueing model (see Section 3) in which users choose when as well as whether to use a facility. For a deterministic setting in which capacity and demand are nonstochastic we considered four pricing regimes; in order of increasing efficiency: no toll, a time-invariant toll, a step-function toll and a continuously time-varying toll. We showed that optimal design capacity decreases with the efficiency of the pricing regime if $\varepsilon < 1$, and *vice versa* if $\varepsilon > 1$.

Pricing in that paper serves a role similar to information here as a means of improving the efficiency with which a given facility is used. Pricing and information both reduce the return from capacity expansion when demand is inelastic, and increase it when demand is elastic.

2.5 Robustness

We have assumed that user costs are homogeneous of degree zero in demand and capacity, and demand is isoelastic. If either assumption is relaxed, the results do not go through. In particular, Prop. 3 does not hold up: information is not necessarily welfare-improving. To see this in the case of full information, let N(p) be an arbitrary demand curve and AC(N,s) the average cost curve. Using (2.23), the welfare change in going from zero to full information can be written

$$\overline{CS}^F - \overline{CS}^Z = \int_d \int_n \int_\sigma \left\{ \int_{p(n,\sigma)}^{\bar{p}_d^u} N(p)dp \right\} dF_d(n,\sigma)\ d\Gamma(d). \tag{2.27}$$

Consider a simple example in which there are no day-specific fluctuations, demand intensity is fixed, and capacity is either high, s_H, or low, s_L, with equal probability. Then $\bar{p}^u = (p(s_H) + p(s_L))/2$, and (2.27) reduces to

$$\overline{CS}^F - \overline{CS}^Z = \frac{1}{2}\left[\int_{p(s_H)}^{\bar{p}^u} N(p)dp - \int_{\bar{p}^u}^{p(s_L)} N(p)dp\right],$$

which is one half the difference between the diagonally shaded area and the vertically shaded area in Figure 1. In panel (a), with homogeneous costs and isoelastic demand, the difference is positive. But in panel (b) with nonhomogeneous costs, and panel (c), where the demand curve is sharply curved in the neighborhood of $\bar{p}^u$, the difference is negative. In (b), $p(s_H)$ is relatively high because the cost curve is steep to the right of $N(\bar{p}^u)$. In (c), $p(s_H)$ is relatively high because the demand curve is relatively elastic below $\bar{p}^U$. In both cases, the message that capacity is high induces extra usage that drives up costs appreciably.

The welfare effect of full information in the example is considered another way in Figure 2 (the average cost curves are made linear for simplicity). The marginal cost curves for the high and low capacity states are $MC(N,s_H)$ and $MC(N,s_L)$. The deadweight loss with zero information from overuse in the two states is indicated by the heavily bordered areas (underusage is possible with high capacity). With full information, usage expands to N_H with high capacity and falls to N_L with low capacity. The corresponding efficiency loss and efficiency gain relative to zero information are indicated by the vertically and diagonally shaded areas. Again, the shapes of the cost curves and the demand curve determine whether the net gain from information is positive or negative.

3. Dynamic Equilibrium

In this section we endogenize individuals' time of use by adopting a variant of Vickrey's (1969) bottleneck queueing model.[24] The facility is assumed to have a bottleneck with a maximum service rate of s users per unit time. In the commuting context the facility is a road or freeway, and the bottleneck a bridge, tunnel, lane drop, *etc.*[25] If the arrival rate of users at the bottleneck exceeds s, a queue develops. The time taken to use the facility for a user arriving at time t is

$$T(t) = \bar{T} + \frac{Q(t)}{s},$$

where $\bar{T}$ is usage time in the absence of a queue. Without loss of generality we assume $\bar{T}$ is zero. Q(t) is the length of the queue, equal to

$$Q(t) = \int_{\hat{t}}^{t} r(\tau)d\tau - s(t-\hat{t}), \tag{3.1}$$

where $\hat{t}$ is the time at which the queue was last zero, and r is the arrival rate at the bottleneck.[26]

If individuals are indifferent about when they use the facility, capacity is enough to serve everyone during the period of operation, and queueing time is unpleasant, then queueing will not occur in equilibrium. However, in most real world situations time of use does matter. For example, commuters typically incur penalties from arriving at work after the official starting time. They also prefer not to arrive very early, with the associated costs of an early morning rise and wasted time before work begins. Non-work activities are better scheduled at certain times to conform with work and household time constraints, and individual biorhythms. And outdoor recreational opportunities are influenced by diurnal, weekly and/or seasonal cycles that determine when usage is best, or even possible.

To model time-of-use preferences, we assume that individuals have a common preferred time, t^*, for being served - that is, for *ending* usage of the facility, and incur a 'schedule delay cost' (SDC) if served at a different time.[27] For an individual arriving at the facility at time t, queueing plus schedule delay costs are assumed to be

$$C(t) = \alpha\frac{Q(t)}{s} + \beta \text{ Max}[0,(t^{*} - t - \frac{Q(t)}{s})^{\eta}] + \gamma \text{ Max}[0,(t + \frac{Q(t)}{s} - t^{*})^{\eta}]. \qquad (3.2)$$

The parameter α measures disutility (inclusive of any monetary, such as vehicle operating, cost) of queueing time. β, γ and η determine the shape of the SDC function. If $\eta = 1$, as has been assumed in most of the theoretical literature, C(t) is piecewise linear; β is the unit cost of being early and γ the unit cost of being late. If $\eta > 1$, the marginal disutility from schedule delay increases with the size of the delay.[28] For reasons described below we rule out $\eta < 1$. To admit differences in the cost of earliness and lateness we allow $\beta \neq \gamma$; for morning commuters with fixed work hours one expects $\beta << \gamma$. The power factor η is assumed the same for early and late service to admit a closed-form solution; see below.

3.2 Equilibrium with full information

In deciding when to use the facility, individuals face a trade-off between queueing time and schedule delay. With full information they know the distribution of arrival times. In equilibrium, no one can reduce usage costs by arriving at a different time. With identical individuals this means that costs are constant during the period when usage occurs.

Equilibrium for the Vickrey model in a deterministic setting, which is equivalent to that under full information, has been exposited elsewhere (*e.g.* Hendrickson and Kocur (1981), Arnott *et al.* (1990a, 1991b)) for $\eta = 1$, and requires only minor modifications for the more general specification given in (3.2).

Equilibrium for $\eta > 1$ is shown in Figure 3. The number of individuals in the queue is measured by the vertical distance between the curve depicting cumulative arrivals at the facility, and the straight line with slope s depicting the cumulative number served. Arrivals occur over the time interval $[t_0, t_1]$. A derivation of the equilibrium arrival rate and number of users, follows.

Let t_n be the arrival time for which an individual gets served at t^*. During $[t_0, t_n)$, individuals are served early. Setting the derivative of (3.2) with respect to t equal to zero, one obtains for the equilibrium arrival rate

$$r(t) = \frac{\alpha s}{\alpha - \beta\eta(t^* - t - Q(t)/s))^{\eta-1}}, \quad t \in (t_0, t_n). \tag{3.3}$$

If $\eta < 1$, r(t) becomes infinite before $t + Q(t)/s$ reaches t^*. To avoid this we assume (as noted above) that $\eta \geq 1$. If $\eta = 1$, $r(t) = \alpha s/(\alpha-\beta)$, which is positive and finite if $\alpha > \beta$. We assume henceforth that this condition holds.[29] If $\eta > 1$, r(t) is strictly decreasing for $t \in (t_0, t_n)$. At t_0,

$$r(t) = \frac{\alpha s}{\alpha - \beta\eta\left(\frac{\gamma^{1/\eta}}{\beta^{1/\eta}+\gamma^{1/\eta}}\frac{N}{s}\right)^{\eta-1}}. \tag{3.4}$$

This is positive and finite provided

$$\frac{N}{s} < \frac{\alpha^{1/(\eta-1)}}{\beta\eta}\frac{\beta^{1/\eta}+\gamma^{1/\eta}}{\gamma^{1/\eta}},$$

which we assume is satisfied for the equilibrium value of N.

Over the arrival interval (t_n, t_1), individuals are served late. Setting $\dot{C}(t) = 0$ as before, one gets

$$r(t) = \frac{\alpha s}{\alpha + \alpha\eta(t + Q(t)/s - t^*))^{\eta-1}}, \quad t \in (t_n, t_1). \tag{3.5}$$

which is positive and finite.

The first user to arrive naturally encounters no queue, and suffers only the cost of being served early. The last user also escapes queueing. (Proof: If he did encounter a queue, he could arrive slightly later and reduce queueing time without any change in schedule delay cost.) Equating the schedule delay costs of the first and last users:

$$\beta(t^*-t_0)^\eta = \gamma(t_1-t^*)^\eta. \tag{3.6}$$

Since the bottleneck operates at capacity throughout $(t_0,\ t_1)$,

$$t_1 = t_0 + \frac{N}{s}. \tag{3.7}$$

Solving (3.6) and (3.7) for t_0, t_1 and the equilibrium user cost, C:

$$t_0 = t^* - \frac{\gamma^{1/\eta}}{\beta^{1/\eta}+\gamma^{1/\eta}}\frac{N}{s}, \qquad t_1 = t^* + \frac{\beta^{1/\eta}}{\beta^{1/\eta}+\gamma^{1/\eta}}\frac{N}{s}, \tag{3.8}$$

$$C = \frac{\beta\gamma}{(\beta^{1/\eta}+\gamma^{1/\eta})^\eta}\left(\frac{N}{s}\right)^\eta. \tag{3.9}$$

Comparing (3.9) with (2.2) it is evident that the equilibrium cost, for a given number of users, is as given in the static model, with

$$\delta = \frac{\beta\gamma}{(\beta^{1/\eta}+\gamma^{1/\eta})^\eta}.$$

The equilibrium number of users can hence be solved as in Section 2.2, and the analysis of that section, including Prop. 1, goes through for the dynamic model when users have full information.[30]

For the linear SDC function, $\eta = 1$, (3.8) and (3.9) reduce to

$$t_0 = t^* - \frac{\delta}{\beta}\frac{N}{s}, \qquad t_1 = t^* + \frac{\delta}{\gamma}\frac{N}{s}, \tag{3.10}$$

$$C = \delta\left(\frac{N}{s}\right), \tag{3.11}$$

with

$$\delta = \frac{\beta\gamma}{\beta+\gamma}.$$

(3.10) and (3.11) will be compared with the imperfect information equilibrium, considered next.

3.3 Equilibrium with imperfect information

In the imperfect information regime, individuals face uncertainty about the capacity of the facility and/or the number of users. The equilibrium arrival rate is more difficult to compute than under full information, for several reasons: (1) the bottleneck may not operate at capacity throughout the arrival period, (2) depending on when they arrive, users may not know whether they will be served early or late, (3) if the number of users is uncertain, the arrival rate itself is unknown.

To simplify the algebra we assume henceforth that schedule delay costs are linear; that is, $\eta = 1$. We first show by construction how to solve for the dynamic equilibrium under uncertainty. We then establish that imperfect information can raise expected user costs, even though full information is necessarily welfare-improving. In this sense there is a nonconcavity in the value of information. This result stands in contrast to that of the static model, where both perfect and imperfect information improve efficiency.

With capacity and/or demand unknown, costs will vary according to when users arrive. We shall characterize equilibrium by a normalized arrival rate, $\rho(t)$, where $\rho(t)dt$ is the fraction of users who arrive between t and t+dt. $\rho(t)$ will in general depend on the type of day and on the message received. To economize on notation, d and m subscripts will be suppressed where unnecessary for clarity. If there are N users, the arrival rate at time t is $N\rho(t)$. The assumption that $\rho(t)$ is invariant from day to day can be justified by the law of large numbers if there are many users choosing arrival times independently.[31]

To derive the equilibrium some additional variables are required. Let $R(t)$ be the c.d.f. of arrivals, and t_r the time at which the last user arrives, which may be before he is served, t_1. By definition, $R(t_0) = 0$ and $R(t_r) = 1$. Let $\phi \equiv N/s$. Given the assumption that n/σ has a finite upper bound, so does ϕ; call it ϕ_M. The c.d.f. of ϕ, from the perspective of users, is denoted $J(\phi)$.[32] Let $Q(t,\phi)$ be the length of the queue at time t when $N/s = \phi$. Define $\phi(t) \equiv \mathrm{Max}\ \{\phi |\ Q(t,\phi) = 0\}$ as the largest ϕ such that there is no queue at time t. A user arriving at t thus queues iff $\phi > \phi(t)$. Define $\phi^*(t)$ by $t + \dfrac{Q(t,\phi^*(t)}{s} = t^*$ as the ϕ such that a user arriving at t gets served at t^*. ($\phi(t)$ and $\phi^*(t)$ may or may not be elements of the support of $J(\phi)$.) Finally, define t_n by $t_n + \dfrac{Q(t,\phi_M)}{s} = t^*$ as the latest arrival time for which a user is never served late.

The qualitative characteristics of equilibrium are shown in Figure 4.[33] There are 4 regions, defined by whether or not users queue, and whether they are served early or late. In regions NE and QE users are served early, in regions NL and QL they are served late. In regions NE and QE they do not queue, whereas in QE and QL they do. If there is no queue, a user is served early if $t < t^*$, and late if $t > t^*$. Regions NE and NL thus abut at $t = t^*$. The boundary between regions QE and QL is defined by the locus $\phi^*(t)$. Regions NE and NL are separated from QE and QL by the locus $\phi(t)$. $\phi(t) < \phi_M$ for all $t \in (t_0,t_r)$, since otherwise the last user would never queue, and could reduce his cost by arriving earlier. Moreover, $\phi^*(t) > \phi(t)$ for $t < t^*$ and $\phi^*(t) < \phi_M$ for $t > t_n$.

Expected costs in each region are:

$$C^{NE}(t,\phi) = \beta(t^*-t), \tag{3.12a}$$

$$C^{QE}(t,\phi) = \beta(t^* - t - \frac{Q(t,\phi)}{s}) + \alpha\frac{Q(t,\phi)}{s}, \tag{3.12b}$$

$$C^{NL}(t,\phi) = \gamma(t-t^*), \tag{3.12c}$$

$$C^{QL}(t,\phi) = \gamma(t + \frac{Q(t,\phi)}{s} - t^*) + \alpha\frac{Q(t,\phi)}{s}. \tag{3.12d}$$

As is clear from Figure 4 there are 3 departure intervals to consider:

$t \in (t_0, t_n)$ in which users are always served early,

$t \in (t_n, t^*)$ in which, depending on ϕ, users may be served early or late,

$t \in (t^*, t_r)$ in which users are always served late.

In equilibrium, C(t) must be independent of t in each of these intervals,

$t \in (t_0, t_n)$

To see that this interval is non-empty, note that the first user must arrive before t^*, since otherwise he could arrive at t^* and escape both queueing and schedule delay, which is not possible for all users. By continuity of R(t) there is a nonzero time interval over which users are always served early.

Given (3.12a), (3.12b) and Figure 4, expected costs are

$$C(t) = \beta\left\{ \int_0^{\phi(t)} (t^*-t)dJ(\phi) + \int_{\phi(t)}^{\phi_M} (t^*-t-\frac{Q(t,\phi)}{s})dJ(\phi) \right\} + \alpha \int_{\phi(t)}^{\phi_M} \frac{Q(t,\phi)}{s}dJ(\phi)$$

$$= \beta(t^*-t) + (\alpha-\beta) \int_{\phi(t)}^{\phi_M} \frac{Q(t,\phi)}{s}dJ(\phi). \tag{3.13}$$

Given $Q(t,\phi(t)) = 0$,

$$\dot{C}(t) = -\beta + (\alpha-\beta) \int_{\phi(t)}^{\phi_M} \frac{\dot{Q}(t,\phi)}{s}dJ(\phi).$$

Given $\dot{Q}(t,\phi) = N\rho(t) - s$ for $\phi > \phi(t)$, we have as a condition for equilibrium

$$\dot{C}(t) = \alpha Z(t) - \beta(1+Z(t)) = 0, \tag{3.14}$$

where

$$Z(t) \equiv \int_{\phi(t)}^{\phi_M} (\phi\rho(t) - 1)dJ(\phi). \tag{3.15}$$

$Z(t)$ is the rate of increase wrt t in expected queueing time. A user who postpones arrival by dt incurs an increase in expected queueing time cost of $\alpha Z(t)dt$, which is the first term on the RHS of (3.14). Since users are also served early, arriving dt later decreases expected schedule delay cost by $\beta(1+Z(t))dt$, which is the second RHS term of (3.14). In equilibrium the two terms must balance. Appendix 4 establishes that (3.14) defines a constant value of $\rho(t)$. The departure rate is thus constant over (t_0, t_n).

$t \in (t_n, t^*)$

In this arrival interval, users are sometimes served early, and sometimes late. Given (3.12a), (3.12b), (3.12d) and Figure 4, expected cost is

$$C(t) = \beta\left\{ \int_0^{\phi(t)} (t^*-t)dJ(\phi) + \int_{\phi(t)}^{\phi^*(t)} \left(t^*-t-\frac{Q(t,\phi)}{s}\right)dJ(\phi) \right\} \tag{3.16}$$

$$+ \gamma \int_{\phi^*(t)}^{\phi_M} \left(t+\frac{Q(t,\phi)}{s}-t^*\right)dJ(\phi) + \alpha \int_{\phi(t)}^{\phi_M} \frac{Q(t,\phi)}{s}dJ(\phi).$$

$$\dot{C}(t) = \alpha Z(t) - \beta Z^e(t) + \gamma Z^l(t) = 0, \tag{3.17}$$

where

$$Z^e(t) \equiv \int_{\phi(t)}^{\phi^*(t)} (\phi\rho(t) - 1)dJ(\phi) + J(\phi^*(t)) \tag{3.18}$$

is the rate of decrease in expected early time and

$$Z^1(t) \equiv \int_{\phi^*(t)}^{\phi_M} (\phi\rho(t) - 1)dJ(\phi) + 1 - J(\phi^*(t)) \tag{3.19}$$

is the rate of increase in expected late time. Equation (3.17) has an interpretation analogous to (3.14). Appendix 4 establishes that $\rho(t)$ is weakly decreasing for $t \in (t_n, t^*)$.

$t \in (t^*, t_r)$ (if $t_r = t^*$ this interval is degenerate)

In this arrival interval, users are always served late. Given (3.12c), (3.12d) and Figure 4, expected costs are

$$C(t) = \gamma\left\{ \int_0^{\phi(t)} (t-t^*)dJ(\phi) + \int_{\phi(t)}^{\phi_M} (t+\frac{Q(t,\phi)}{s} - t^*)dJ(\phi) \right\} + \alpha \int_{\phi(t)}^{\phi_M} \frac{Q(t,\phi)}{s}dJ(\phi),$$

$$= \gamma(t-t^*) + (\alpha+\gamma) \int_{\phi(t)}^{\phi_M} \frac{Q(t,\phi)}{s}dJ(\phi). \tag{3.20}$$

$$\dot{C}(t) = \alpha Z(t) + \gamma(1+Z(t)) = 0. \tag{3.21}$$

(3.21) has an interpretation analogous to (3.14) and (3.17). Again, $\rho(t)$ is weakly decreasing on the interval (see Appendix 4).

Conditions (3.14), (3.17) and (3.21) ensure that expected costs are constant in each arrival interval. Since $\phi^*(t_n) = \phi_M$, and $\phi(t^*) = \phi^*(t^*)$, expected costs are also continuous at t_n and t^*, and hence constant over the whole interval $[t_0, t_r]$. It is now possible to state:

PROPOSITION 4: The normalized arrival rate $\rho(t)$ is weakly decreasing over the arrival period, and the cumulative arrivals distribution R(t) is concave.

Prop. 4 follows from the fact that $\rho(t)$ is constant on (t_0, t_n) and weakly decreasing thereafter. To see why, note that after t_n the later a user arrives the more likely he is to be served late, and to suffer increased lateness from marginally postponing arrival rather than reduced earliness. To compensate, expected queueing costs must grow at a decreasing rate and eventually decline, which requires that the arrival rate decrease over time.[34]

Given concavity of R(t) it follows that if queueing occurs on a given day it must be over a continuous time interval beginning at t_0. Queueing time is

$$\frac{Q(t,\phi)}{s} = \text{Max}\ [0,\ t_0 + \phi R(t) - t]. \tag{3.22}$$

t_n is defined implicitly by the condition

$$t_0 + \phi_M R(t_n) = t^*,$$

while

$$\phi(t) = (t-t_0)/R(t), \tag{3.23}$$

$$\phi^*(t) = (t^*-t_0)/R(t). \tag{3.24}$$

With regard to the timing of arrivals it turns out that there are two possibilities: $t_r < t^*$ and $t_r = t^*$. To establish which occurs define:

$$\bar{\phi} \equiv \int_0^{\phi_M} \phi dJ(\phi),$$

$$\tilde{\phi} \equiv J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\},$$

and the mean of ϕ for values greater than the $\left\{\frac{\alpha}{\alpha+\gamma}\right\}$ fractile:

$$\hat{\phi} \equiv \int_{\tilde{\phi}}^{\phi_M} \phi dJ(\phi)\ /(1-J(\phi)) = \frac{\alpha+\gamma}{\gamma}\int_{\tilde{\phi}}^{\phi_M} \phi dJ(\phi). \tag{3.25}$$

The case $t_r > t^*$ is described in Lemma 1 (for the proof, see Appendix 5)

LEMMA 1:

$$\text{If } \tilde{\phi} > \frac{\gamma}{\beta+\gamma}\hat{\phi} \tag{3.26}$$

then the first and last arrivals occur at

$$t_0 = t^* - \frac{\delta}{\beta}\hat{\phi}, \tag{3.27}$$

$$t_r = t_0 + \tilde{\phi} > t^*. \tag{3.28}$$

Expected costs are

$$C = C(t_0) = \beta(t^*-t_0) = \delta\hat{\phi}. \tag{3.29}$$

As true of the full information equilibrium, arrivals occur both before and after t^*. It follows from (3.10) and (3.27) that the first arrival is earlier on average with imperfect information than with full information.[35] By (3.11) and (3.29) expected user costs are thus greater with imperfect information than full information.

Both the arrival interval and expected user costs are independent of the distribution of ϕ below the $\left\{\frac{\alpha}{\alpha+\gamma}\right\}$ fractile. To understand this, as well as (3.28), suppose the last user arrives at $t_0 + \phi_r$ for some ϕ_r. The user escapes queueing with probability $J(\phi_r)$. If he delays arrival by dt his expected cost changes by $[J(\phi_r)\gamma - (1-J(\phi_r))\alpha]dt$. In equilibrium this must be zero, which is why $\phi_r = J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}$ and $t_r = t_0 + J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}$. If $\phi < J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}$ the last user escapes queueing and incurs a late arrival cost determined by the $\frac{\alpha}{\alpha+\gamma}$ fractile of the $J(\phi)$ distribution. If $\phi > J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}$ his costs are determined by the shape of the $J(\phi)$ distribution *above* this fractile. In either case, costs are independent of the distribution below the $\frac{\alpha}{\alpha+\gamma}$ fractile. Since in equilibrium all users incur the same cost, this is also true of the system.

The case $t_r = t^*$ is described by Lemma 2 (for the proof see Appendix 6):

LEMMA 2:

$$\text{If } \tilde{\phi} \le \frac{\gamma}{\beta+\gamma}\hat{\phi} \tag{3.30}$$

then t_0 is defined implicitly by the equation

$$J(t^*-t_0) + \frac{1}{t^*-t_0}\int_{t^*-t_0}^{\phi_M} \phi dJ(\phi) - \frac{\alpha+\beta+\gamma}{\alpha+\gamma} = 0, \tag{3.31}$$

$$t_r = t^*,$$

and

$$C = C(t_0) = \beta(t^*-t_0). \tag{3.32}$$

By (3.31) the equilibrium is again independent of the left-hand tail of the ϕ distribution.

Since conditions (3.26) and (3.30) in Lemmas 1 and 2 are complementary, there are no equilibria with $t_r < t^*$. To see why, suppose $t_r < t^*$ and consider $t \in (t_r, t^*)$. If $\phi < \phi(t)$ there is no queue at t and a user is better off arriving at t than t_r, since he is served less early. If $\phi > \phi(t)$ he is also better off, since he is served at the same time as when arriving at t_r, but spends less time queueing. Thus, if $t_r < t^*$, arrival after t_r is preferable to arrival at t_r, a contradiction. This yields

PROPOSITION 5: With imperfect information the last user arrives no earlier than t^*.

3.4 Comparison of the Information Regimes

Prop. 2 of Section 2 concerned the relationship between consumers' surplus and marginal consumers' surplus for the static model. The Proposition carries over without alteration to the dynamic model (Proof: see Appendix 7). For completeness we repeat the proposition here as

PROPOSITION 2′: If $\varepsilon < 1$ (resp. $\varepsilon > 1$) the more efficient information regime has the lower (resp. higher) marginal consumers' surplus for *any* design capacity.

In place of Prop. 3 we have a weaker result (Proof: see Appendix 8).

PROPOSITION 3′: Full information increases efficiency relative to zero information. Furthermore, if $\varepsilon < 1$ (resp. > 1) then optimal design capacity is smaller (resp. larger) with full information than with zero information.

Prop. 3′ differs from Prop. 3 in that only *full* information is necessarily welfare-improving relative to zero information.

We now show by example that imperfect information can be welfare-reducing. Suppose there are no day-specific fluctuations (the subscript d is redundant), and demand intensity is fixed, but capacity fluctuates with an (unconditional) p.d.f

$$s = \begin{cases} \hat{s} \text{ with probability } 1-\pi \\ \sigma\hat{s} \text{ with probability } \pi, \quad \sigma < 1. \end{cases} \tag{3.33}$$

In the case of commuting, π can be interpreted as the probability of events (*e.g.* road repairs, bad weather, accidents) that reduce freeway capacity below its design level for the duration of the morning commute. In Appendix 9 we show that, if $\sigma < \frac{\gamma}{\beta+\gamma}$, then $(\bar{p})^{1-\varepsilon}$ varies with π as shown in Figure 5.[36]

Now, with n fixed, no systematic fluctuations, and $\eta = 1$, we have by (2.24)

$$\overline{MCS}^{I}(\hat{s}) - \overline{MCS}^{Z}(\hat{s}) = \frac{1}{1+\varepsilon}\frac{n}{\hat{s}}\left[\int_{m}\left(\bar{p}^{u}_{m}\right)^{1-\varepsilon}dM(m) - \left(\bar{p}^{u}\right)^{1-\varepsilon}\right]. \tag{3.34}$$

By Prop. 2

$$\overline{CS}^{I}(\hat{s}) \gtrless \overline{CS}^{Z}(\hat{s}) \text{ if } \left[\overline{MCS}^{I}(\hat{s}) - \overline{MCS}^{Z}(\hat{s})\right](\varepsilon - 1) \gtrless 0. \tag{3.35}$$

With zero information, the probability from the users' perspective of reduced capacity is π every day. With full information, it is 0 with probability $1-\pi$ and 1 with probability π. It is clear from Figure 5 that (3.34) is negative if $\varepsilon < 1$ and positive if $\varepsilon > 1$. By (3.35), full information is thus welfare-improving.

With imperfect information, the probability of low capacity conditional on the message received fluctuates around π. At $\pi_c \equiv \frac{\beta\sigma}{(\alpha+\gamma)(1-\sigma)}$, there is a kink in the function $\left(\bar{p}^{u}\right)^{1-\varepsilon}(\pi)$. Suppose $\pi \cong \pi_c$, and messages are not very informative, in the sense that the probability of capacity reduction conditional on a message rarely differs much from π. By the unbiasedness of the message system and Jensen's inequality (3.34) is then *positive* if $\varepsilon < 1$ and *negative* if $\varepsilon > 1$. Imperfect information is welfare-reducing; there is a nonconcavity in the value of information.[37] This is the inspiration for the quote from Alexander Pope at the head of the article: "A little learning is a dangerous thing".[38]

A somewhat brute-force explanation for the kink in the $\left(\bar{p}^{u}\right)^{1-\varepsilon}(\pi)$ is possible with the aid of Figure 6. Suppose, to consider a general bivariate distribution, that $\frac{N}{s} = \phi$ with probability $1-\pi$ and $\frac{N}{s} = \phi' > \phi$ with probability π. If $\pi = 0$, then in equilibrium with the linear SDC function (panel (a) of Figure 6) the first user arrives at $t_0 = t^* - \frac{\gamma}{\beta+\gamma}\phi$ and the last user at $t_r = t^* + \frac{\beta}{\beta+\gamma}\phi$. The times at which the two users are served, and their usage costs, are depicted by the points A_0 and A_1. Both users avoid queueing and incur the same schedule delay cost.

Suppose now that $\pi > 0$. With the original arrival schedule the last user would incur an extra cost $(\alpha+\gamma)(\phi'-\phi)$ with probability π. To maintain equilibrium, arrivals must begin earlier. This reduces the expected cost of the last user, since if $\frac{N}{s} = \phi$ he still escapes queueing and is served earlier, while if $\frac{N}{s} = \phi'$, he suffers both less queueing and less late cost. The first user ends up at point B_0, and the last user at B_1 or B_1' according to whether ϕ or ϕ' is realized.

The shift toward earlier arrivals continues as π increases until $t_r = t^*$. The last user does not arrive any sooner if π increases further, as explained earlier. The last user thus bears the full brunt of the increase in expected costs, so that the expected price of usage rises more sharply as π increases above π_c.

The fact that $\tilde{p}^u(\pi)$ in Figure 5 is constant for $\pi \in [\frac{\gamma}{\alpha+\gamma}, 1]$ follows from the fact, also explained earlier, that equilibrium is invariant to the distribution of ϕ below the $\left\{\frac{\alpha}{\alpha+\gamma}\right\}$ fractile. Changes in π within this range induce changes in the arrival *rate*, but not the arrival *interval*. Furthermore, $\bar{p} = \delta\phi$; expected costs are the same as if $\frac{N}{s} = \phi'$ with certainty.

Attempts by a facility manager to increase welfare by improving the reliability of the facility through reductions in π or reductions in ϕ (holding ϕ' fixed) would be defeated by offsetting changes in user behavior.

The fact that imperfect information necessarily increases welfare in the static model, but can decrease it in the dynamic model, illustrates the importance of modeling all margins of user behavior. Neglect of users' time-of-use decisions can lead to qualitative errors in analysis, as well as overlooking of interesting and counterintuitive behavior.

3.5 Robustness

The kink in $\bar{p}^u(\pi)$ at $\pi = \pi_c$ derives from the fact that the linear SDC function has a kink at t^*. However, nonconcavities in $\bar{p}^u(\pi)$ can arise even if schedule delay costs are everywhere differentiable. Consider the SDC function $D(t-t^*) = \beta(t-t^*)^\eta$, $\eta > 0$. It is possible to construct examples with a bivariate capacity distribution in which $\bar{p}^u(\pi)$ is strictly convex over a range of π for any $\eta \geq 1$. Some intuition for this can be gleaned from panel (b) of Figure 6. Again, the equilibrium service times and costs of the first and last users are shown for $\pi = 0$ by A_0 and A_1, and for $\pi > 0$ by B_0, B_1 and B_1'. As π increases, the shift toward earlier arrivals slows progressively as the first user is driven up the increasingly steep early arrival branch of the SDC curve. The expected cost of the last user (and in equilibrium all users) thus increases at an increasing rate.

4. Summary and Concluding Remarks

Most congestible facilities are subject to nonsystematic fluctuations in demand and/or capacity. In the case of transportation, recreational areas, telecommunications and other facilities, real-time or near real-time information may be available on the state of the system. Where the provision of information is a policy decision, as will presumably be the case with route guidance systems for automobile travel, it is important that the welfare effects of information be understood.

In this paper we investigate the effect of information on usage of a congestible facility under free access. We consider three information regimes: zero information, full (perfect) information and imperfect information, which contains as a limiting case the other two regimes. We begin with a static model in which only the aggregate level of usage is endogenous. We assume average user cost (price) is homogeneous of degree zero in usage and capacity, and demand is isoelastic. We then consider a dynamic model in which individuals' time of usage as well as their participation decisions are endogenous. In the case of full (perfect) information the reduced-form equilibrium cost function of the dynamic model is a special case of the static model.

For the static model with homogeneous costs and isoelastic demand both imperfect and perfect information are welfare-improving relative to zero information. Full information is also welfare-improving for the dynamic model with full information, but not necessarily with imperfect information. And neither perfect nor imperfect information need be welfare-improving in either model for arbitrary user cost and demand functions. These results suggest that a cautious approach be adopted in designing and implementing information dissemination schemes, at least absent tolls or other means of regulating usage.

There are several directions in which the analysis could be extended. We have assumed the facility is a delay system, in which users are neither turned away nor balk when congestion is heavy. This is reasonable for travel because of the substantial sunk costs and commitments generally involved. With other facilities, such as the telephone, the cost of attempting usage may be quite low, so that users can balk and try again later.

We have considered a single facility in isolation. This is unsatisfactory if the facility is part of a system, such as one road in a corridor with several roads, or a ski area close to competing facilities. Information may affect users' choice of facility as well as their decision whether to participate and when. Furthermore, with unpriced congestion, capacity expansion or other policies adopted at one facility will affect the efficiency loss due to congestion elsewhere in the system.[39]

It has been assumed that all users have the same information. For various reasons some individuals tend to be better informed than others: greater experience, preferential access to information, superior cognitive capabilities, *etc.* Haltiwanger and Waldman (1985) have shown that informed decision makers have an impact disproportionate to their fraction of the population in determining equilibrium in congestible systems. This would suggest that assuming all users know the available information may be reasonable. In the case of automobile travel, however, it has been theorized as well as observed in the field that overreaction (as defined in the introduction) to information can occur if too many drivers are tuned in and respond (Ben Akiva and de Palma (1991), Mahmassani and Jayakrishnan (1991)). Given this, and the costs of installing and operating route guidance systems in individual vehicles, the fraction of drivers to equip becomes a policy issue.

In the dynamic model used here, queue discipline is FIFO. This is reasonable for transportation and some recreational facilities. Other facilities have different congestion technologies and service disciplines. With time sharing, computer users are served quasi-simultaneously. On a telephone network, capacity is determined by the number of telephone lines, while the completion rate of calls is governed by the average length of a call. Too, in some exchanges users gain access to a line at random rather than in order of placing a call. In other facilities the quality of service may degrade with cumulative usage; *e.g.* due to buildup of heat in a museum, or wear and tear on nature trails over the course of a season. It remains to be seen whether the welfare effects of information are sensitive to these considerations.

The probability distribution of demand intensity (n) and capacity availability (σ) has been assumed independent of design capacity ($\hat{s}$). In practice, there may be a tradeoff between the utilization rate of a facility and the probability or magnitude of loss of capacity. For example, if highway shoulders are used as travel lanes during peak hours, which effectively increases design capacity, they will be unavailable for stalled vehicles and rescue equipment, so that the average capacity utilization rate may fall. Furthermore, σ may depend on demand, as is the case for road traffic to the extent that the probability of incidents increases with traffic volume (Newbery (1990)). And demand may be a function of service reliability (Kay (1979), Saving and DeVany (1981), Coate and Panzar (1988)). A definition of capacity ought thus to include a reliability coefficient, which would be a factor in determining optimal design capacity, as well as maintenance and repair policy.

Perhaps the most important next step is to consider the use of pricing in conjunction with information as a policy tool. For the dynamic model it is necessary to derive the optimal time-varying toll under imperfect (or zero) information about demand and/or capacity. In the case of the bottleneck queueing model with full information it has been shown (*viz.* Arnott *et al.* (1990a, 1991b)) that time-varying tolls can provide substantial efficiency gains through induced changes in time-of-usage decisions. It remains to be seen how large these gains are in a stochastic environment, and how they vary with the quality of information.

APPENDIX 1

PROOF OF PROPOSITION 1

Proposition 1 concerns the ratio of optimal design capacity with anticipated fluctuations in capacity and demand, $\hat{s}_*^F$ given in (2.14), to optimal design capacity, $\hat{s}_*$, when capacity remains at its design value and demand intensity is fixed at its mean value.

Define $F(n,\sigma) \equiv \int_d F_d(n,\sigma)\, d\Gamma(d)$. Then

$$\hat{s}_*^F/\hat{s}_* = \nu/\bar{n} = \left[\int_n \int_\sigma n^{\frac{1+\eta}{1+\varepsilon\eta}} \sigma^{\frac{\eta(\varepsilon-1)}{1+\varepsilon\eta}} dF(n,\sigma) \right]^{\frac{1+\varepsilon\eta}{1+\eta}} / \bar{n} ,$$

where $\bar{n}$ is mean demand intensity. The various statements in Proposition 1 are proved by a series of lemmas.

LEMMA 1:

$$\nu \gtreqless \bar{n} \quad \text{as } \varepsilon \lesseqgtr 1.$$

PROOF: Let x and y be variates. By Hölder's inequality

$$\int\int x^\alpha y^{1-\alpha} dF(x,y) \lesseqgtr (\bar{x})^\alpha (\bar{y})^{1-\alpha} \quad \text{as} \quad \alpha \begin{cases} \in (0,1) \\ = 0,\ 1 \\ < 0 \text{ or } > 1 \end{cases}$$

where $F(x,y)$ is an arbitrary nondegenerate joint c.d.f. Replacing x with σ, y with n, α with $\eta(\varepsilon-1)/(1+\varepsilon\eta)$ and $F(x,y)$ with $F(n,\sigma)$ yields

$$\nu \gtreqless (\bar{\sigma})^{\frac{\eta(\varepsilon-1)}{1+\eta}} \bar{n} \gtreqless \bar{n} \quad \text{as } \varepsilon \lesseqgtr 1. \quad \text{QED.}$$

Thus, optimal design capacity is greater in the stochastic case if $\varepsilon < 1$. The reason is that, with $\varepsilon < 1$, marginal consumers' surplus is a convex function of n and σ, so that by Jensen's inequality expected marginal consumers' surplus from capacity expansion is greater. If $\varepsilon > 1$, marginal consumers' surplus is concave in n and σ and the inequality is reversed.

LEMMA 2:

Let $H(\sigma)$ be the c.d.f. of σ and $G(n|\sigma)$ be the c.d.f. of n conditional on σ. Then if $\varepsilon < 1$, $\nu/\bar{n}$ increases with a mean-preserving spread to $G(n|\sigma)$ for any set of σ of nonzero probability measure. The opposite is true if $\varepsilon > 1$.

PROOF: From (2.14b)

$$\nu = \left[\int_{\sigma} \sigma^{\frac{\eta(\varepsilon-1)}{1+\varepsilon\eta}} \nu(\sigma) dH(\sigma) \right]^{\frac{1+\varepsilon\eta}{1+\eta}}, \quad \text{where } \nu(\sigma) \equiv \int_{n} n^{\frac{1+\eta}{1+\varepsilon\eta}} dG(n|\sigma).$$

Let i_1 and i_2 be indexes. Assume that, for all σ, $G(n|\sigma, i_1)$ either coincides with $G(n|\sigma, i_2)$ or can be obtained by a mean-preserving spread, with the latter true for a subset of σ of positive probability measure.

If $\varepsilon < 1$, $n^{\frac{1+\eta}{1+\varepsilon\eta}}$ is strictly convex in n, so that by Rothschild and Stiglitz (1970), $\nu(\sigma, i_1) \geq \nu(\sigma, i_2)$ for all σ, with a strict inequality for σ of nonzero measure. Then $\nu(i_1) > \nu(i_2)$. If $\varepsilon > 1$, $n^{\frac{1+\eta}{1+\varepsilon\eta}}$ is strictly concave in n, and by analogous reasoning $\nu(i_1) < \nu(i_2)$. The proof is completed by noting that the construction holds $\bar{n}$ fixed. QED.

LEMMA 3:

Let $G(n)$ be the c.d.f. of n, and $H(\sigma|n)$ be the c.d.f. of σ conditional on n. Then if $\varepsilon < 1$, $\nu/\bar{n}$ increases with a mean-preserving spread to $H(\sigma|n)$ for any set of n of nonzero probability measure. The opposite is true if $\varepsilon > 1$.

PROOF: By (2.14b)

$$\nu = \left[\int_n n^{\frac{1+\eta}{1+\varepsilon\eta}} w(n)dG(n) \right]^{\frac{1+\varepsilon\eta}{1+\eta}}, \quad \text{where } w(n) \equiv \int_\sigma \sigma^{\frac{\eta(\varepsilon-1)}{1+\varepsilon\eta}} dH(\sigma|n).$$

If $\varepsilon < 1$ (resp. > 1), $\sigma^{\frac{\eta(\varepsilon-1)}{1+\varepsilon\eta}}$ is strictly convex (resp. concave) in σ. The proof follows by analogy with the proof of Lemma 2. QED.

To establish the two remaining lemmas some preliminaries are required. Let x and y be variates with joint c.d.f. F(x,y) and marginal distributions H(x) and G(y). Define

$$M_\alpha(x) \equiv \left\{ \int x^\alpha dH(x) \right\}^{1/\alpha}, \quad M_\alpha(y) \equiv \left\{ \int y^\alpha dG(y) \right\}^{1/\alpha},$$

$$M_\alpha(x,y) \equiv \iint x^\alpha y^{1-\alpha} dF(x,y),$$

and

$$\rho_\alpha(x,y) \equiv \left\{ \frac{M_\alpha(x,y)}{(M_\alpha(x))^\alpha (M_{1-\alpha}(y))^{1-\alpha}} \right\} - 1. \tag{A1.1}$$

$M_\alpha(x)$, $M_\alpha(y)$ and $M_\alpha(x,y)$ are generalized means, and $\rho_\alpha(x,y)$ is a generalized correlation coefficient. If x and y are independent then:

$$M_\alpha(x,y) = \left[\int y^{1-\alpha} dG(y)\right]\left[\int x^\alpha dH(x)\right] = (M_\alpha(x))^\alpha (M_{1-\alpha}(y))^{1-\alpha},$$

so that $\rho_\alpha(x,y) = 0$. From (2.14b) and (A1.1)

$$\nu = \left[(M_\alpha(\sigma))^\alpha (M_{1-\alpha}(n))^{1-\alpha}(1 + \rho_\alpha(\sigma,n))\right]^{1/(1-\alpha)}, \tag{A1.2}$$

where

$$\alpha = \frac{\eta(\varepsilon-1)}{1+\varepsilon\eta} < 1.$$

We can now proceed to lemmas 4 and 5.

LEMMA 4:

If $\varepsilon < 1$ (resp. > 1), $\nu/\bar{n}$ is larger the smaller (resp. larger) the α-mean of σ.

PROOF: Fix $M_{1-\alpha}(n)$, $\bar{n}$ and $\rho_\alpha(\sigma,n)$. By (A1.2) $\nu/\bar{n}$ is larger the larger is $(M_\alpha(\sigma))^\alpha$, and hence with $\varepsilon < 1$ ($\alpha < 0$) the *smaller* is $M_\alpha(\sigma)$. If $\varepsilon > 1$, $\nu/\bar{n}$ is larger the *larger* is $M_\alpha(\sigma)$. QED.

LEMMA 5:

If $\varepsilon < 1$ (resp. > 1) $\nu/\bar{n}$ is larger the smaller (resp. larger) the correlation between n and σ.

PROOF: Fix $M_\alpha(\sigma)$, $M_{1-\alpha}(n)$ and $\bar{n}$. By (A1.2), $\nu/\bar{n}$ is larger the larger is $\rho_\alpha(\sigma,n)$. If $\varepsilon < 1$, $\alpha < 0$, and $\nu/\bar{n}$ is larger the *smaller* the correlation between σ and n. If $\varepsilon > 1$, $\alpha > 0$, and $\nu/\bar{n}$ is larger the *larger* the correlation. QED.

APPENDIX 2

PROOF OF PROPOSITION 2

Consumers' surplus with imperfect information is given by (2.23), which we write here with the order of integration altered:

$$\overline{CS}^{I} = \int_{d}\int_{n}\int_{\sigma}\left[\int_{m}\left\{\int_{\bar{p}^{u}_{dm}} np^{-\varepsilon}dp\right\} dM(m|d,\sigma,n)\right] dF_{d}(n,\sigma)d\Gamma(d).$$

Let I and I′ be two imperfect information regimes with message distributions $M(m|d,\sigma,n)$ and $M'(m'|d,\sigma,n)$ respectively. The difference in expected consumers' surplus between the two regimes is

$$\overline{CS}^{I} - \overline{CS}^{I'} = \int_{d}\int_{n}\int_{\sigma}\left[\int_{m}\left\{\int_{\bar{p}^{u}_{dm}} np^{-\varepsilon}dp\right\} dM(m|d,n,\sigma) - \int_{m'}\left\{\int_{\bar{p}^{u}_{dm'}} np^{-\varepsilon}dp\right\} dM'(m'|d,n,\sigma)\right]dF_{d}(n,\sigma)d\Gamma(d).$$

$$= \int_{d}\int_{n}\int_{\sigma}\left[\int_{m}\int_{m'}\left\{\int_{\bar{p}^{u}_{dm}} np^{-\varepsilon}dp-\int_{\bar{p}^{u}_{dm'}} np^{-\varepsilon}dp\right\}dM'(m'|d,n,\sigma)dM(m|d,n,\sigma)\right]dF_{d}(n,\sigma)d\Gamma(d).$$

Define $S(m) \equiv \left\{m|\ \bar{p}^{u}_{dm'} \le \bar{p}^{u}_{dm}\right\}$. Then

$$\overline{CS}^{I} - \overline{CS}^{I'} = \int_{d}\int_{n}\int_{\sigma}\left[\int_{m}\left[\int_{m'\in S(m)}\int_{\bar{p}^{u}_{dm'}}^{\bar{p}^{u}_{dm}} np^{-\varepsilon}dp\ dM'(m'|d,n,\sigma) - \int_{m'\notin S(m)}\int_{\bar{p}^{u}_{dm}}^{\bar{p}^{u}_{dm'}} np^{-\varepsilon}dp\ dM'(m'|d,n,\sigma)\right] dM(m|d,n,\sigma)\right] dF_{d}(n,\sigma)d\Gamma(d).$$

$$=\int_{d}\int_{n}\int_{\sigma}\left[\int_{m}\left[\int_{m'\in S(m)}\frac{n}{1-\varepsilon}\left(\left(\bar{p}^{u}_{dm}\right)^{1-\varepsilon} - \left(\bar{p}^{u}_{dm'}\right)^{1-\varepsilon}\right)dM'(m'|d,n,\sigma) - \int_{m'\notin S(m)}\frac{n}{1-\varepsilon}\left(\left(\bar{p}^{u}_{dm'}\right)^{1-\varepsilon}-\left(\bar{p}^{u}_{dm}\right)^{1-\varepsilon}\right)dM'(m'|d,n,\sigma)\right]dM(m|d,n,\sigma)\right]dF_{d}(n,\sigma)d\Gamma(d). \qquad (A2.1)$$

Now $\overline{MCS}^{I}(\hat{s}) - \overline{MCS}^{I'}(\hat{s}) = \frac{d}{d\hat{s}}\left[\overline{CS}^{I}(\hat{s}) - \overline{CS}^{I'}(\hat{s})\right]$

$$= \int_d \int_n \int_\sigma \Bigg[\int_m \Bigg[\int_{m' \in S(m)} \frac{n}{1-\varepsilon} \frac{d}{d\hat{s}} \left[\left(\bar{p}^{u}_{dm}\right)^{1-\varepsilon} - \left(\bar{p}^{u}_{dm'}\right)^{1-\varepsilon} \right] dM'(m'|d,n,\sigma) \tag{A2.2}$$

$$- \int_{m' \notin S(m)} \frac{n}{1-\varepsilon} \frac{d}{d\hat{s}} \left[\left(\bar{p}^{u}_{dm'}\right)^{1-\varepsilon} - \left(\bar{p}^{u}_{dm}\right)^{1-\varepsilon} \right] dM'(m'|d,n,\sigma) \Bigg] dM(m|d,n,\sigma) \Bigg] dF_d(n,\sigma) d\Gamma(d).$$

(By Leibnitz's rule terms involving derivatives of the limits of integration cancel.) Now by (2.19)

$$\frac{d\bar{p}^{u}_{dm}}{d\hat{s}} \frac{\hat{s}}{\bar{p}^{u}_{dm}} = - \frac{\eta}{1+\varepsilon\eta}, \tag{A2.3}$$

whence

$$\frac{d}{d\hat{s}} \left(\bar{p}^{u}_{dm}\right)^{1-\varepsilon} = - \frac{\eta(1-\varepsilon)}{1+\varepsilon\eta} \frac{1}{\hat{s}} \left(\bar{p}^{u}_{dm}\right)^{1-\varepsilon}. \tag{A2.4}$$

Substituting (A2.4) and its counterpart for m′ into (A2.2) and using (A2.1) one finds

$$\overline{MCS}^{I}(\hat{s}) - \overline{MCS}^{I'}(\hat{s}) = \frac{\eta(\varepsilon-1)}{1+\varepsilon\eta} \frac{1}{\hat{s}} \left[\overline{CS}^{I}(\hat{s}) - \overline{CS}^{I'}(\hat{s})\right].$$

If $\varepsilon > 1$, $\overline{MCS}^{I}(\hat{s}) - \overline{MCS}^{I'}(\hat{s})$ and $\overline{CS}^{I}(\hat{s}) - \overline{CS}^{I'}(\hat{s})$ have the same sign. If $\varepsilon < 1$, they have the opposite sign. Furthermore, by (2.24)

$$\overline{MCS}^{I}(\hat{s}) = \frac{\eta}{1+\varepsilon\eta} \frac{1}{\hat{s}} \int_d \int_m \bar{n}_{dm} \left(\bar{p}^{u}_{dm}\right)^{1-\varepsilon} dM(m|d)\, d\Gamma(d).$$

Differentiating and using (A2.3)

$$\frac{d\overline{MCS}^{I}(\hat{s})}{d\hat{s}} \frac{\hat{s}}{\overline{MCS}^{I}(\hat{s})} = - \frac{1+\eta}{1+\varepsilon\eta}.$$

Since the elasticity of marginal consumers' surplus is constant, the ranking of $\overline{MCS}^{I}(\hat{s})$ and $\overline{MCS}^{I'}(\hat{s})$ is independent of $\hat{s}$, so that the relative efficiency of the two information regimes can be established by examining *any* level of design capacity. QED.

APPENDIX 3

PROOF OF PROPOSITION 3

From (2.24) in the text

$$\overline{MCS}^I(\hat{s}) = \frac{\eta}{1+\varepsilon\eta}\frac{1}{\hat{s}}\int_d\int_m \bar{n}_{dm}\left(\bar{p}^u_{dm}\right)^{1-\varepsilon} dM(m|d)\ d\Gamma(d).$$

This reduces in the case of zero information to

$$\overline{MCS}^Z(\hat{s}) = \frac{\eta}{1+\varepsilon\eta}\frac{1}{\hat{s}}\int_d \bar{n}_d\left(\bar{p}^u_d\right)^{1-\varepsilon} d\Gamma(d).$$

It follows that

$$\overline{MCS}^I(\hat{s}) - \overline{MCS}^Z(\hat{s}) = \int_d\left(\overline{MCS}^I_d(\hat{s}) - \overline{MCS}^Z_d(\hat{s})\right)d\Gamma(d), \tag{A3.1}$$

where

$$\overline{MCS}^I_d(\hat{s}) - \overline{MCS}^Z_d(\hat{s}) = \frac{\eta}{1+\varepsilon\eta}\frac{1}{\hat{s}}\left[\int_m \bar{n}_{dm}\left(\bar{p}^u_{dm}\right)^{1-\varepsilon} dM(m|d) - \bar{n}_d\left(\bar{p}^u_d\right)^{1-\varepsilon}\right]. \tag{A3.2}$$

Now $\bar{n}_{dm} = \int_n\int_\sigma n dF_{dm}(n,\sigma),$ (A3.3)

$$\bar{n}_d = \int_n\int_\sigma n dF_d(n,\sigma), \tag{A3.4}$$

and from (2.19)

$$\bar{p}^u_{dm} = \left[\frac{\delta(\hat{s})^{-\eta}}{\bar{n}_{dm}}\int_n\int_\sigma n\left(\frac{n}{\sigma}\right)^\eta dF_{dm}(n,\sigma)\right]^{\frac{1}{1+\varepsilon\eta}}, \tag{A3.5}$$

$$\bar{p}^u_d = \left[\frac{\delta(\hat{s})^{-\eta}}{\bar{n}_d}\int_n\int_\sigma n\left(\frac{n}{\sigma}\right)^\eta dF_d(n,\sigma)\right]^{\frac{1}{1+\varepsilon\eta}}. \tag{A3.6}$$

Substituting (A3.3) - (A3.6) into (A3.2) and using the identity

$$dF_d(n,\sigma) = \int_m dF_{dm}(n,\sigma)\ dM(m|d)$$

we have

$$\overline{MCS}_d^I(\hat{s}) - \overline{MCS}_d^Z(\hat{s}) \lesseqgtr \int_m \left[\int_n \int_\sigma n dF_{dm}(n,\sigma) \right]^{\frac{\varepsilon(1+\eta)}{1+\varepsilon\eta}} \left[\int_n \int_\sigma \left(\frac{n}{\sigma}\right)^{\eta} n dF_{dm}(n,\sigma) \right]^{\frac{1-\varepsilon}{1+\varepsilon\eta}} dM(m|d)$$

$$- \left[\int_m \int_\sigma \int_n n dF_{dm}(n,\sigma) dM(m|d) \right]^{\frac{\varepsilon(1+\eta)}{1+\varepsilon\eta}} \left[\int_m \int_\sigma \int_n \left(\frac{n}{\sigma}\right)^{\eta} n dF_{dm}(n,\sigma) dM(m|d) \right]^{\frac{1-\varepsilon}{1+\varepsilon\eta}}. \qquad \text{(A3.7)}$$

The RHS of (A3.7) has the form

$$\int \int x^{\alpha} y^{1-\alpha} dF(x,y) - (\bar{x})^{\alpha} (\bar{y})^{1-\alpha}$$

where

$$x \equiv \int_n \int_\sigma n \; dF_{dm}(n,\sigma) \quad , \; \bar{x} \equiv \int_m x \; dM(m|d),$$

$$y \equiv \int_n \int_\sigma \left(\frac{n}{\sigma}\right)^{\eta} n \; dF_{dm}(n,\sigma) \quad , \; \bar{y} \equiv \int_m y \; dM(m|d),$$

$$\alpha \equiv \frac{\varepsilon(1+\eta)}{1+\varepsilon\eta},$$

and the integration with respect to x and y is taken along the curve $(x(m), y(m))$.

Applying Hölder's inequality as in Appendix 1 we have

$$\overline{MCS}_d^I(\hat{s}) \gtreqless \overline{MCS}_d^Z(\hat{s}) \qquad \text{as} \quad \varepsilon \gtreqless 1.$$

Combining this with (A3.1) and Proposition 2 yields

$$\overline{CS}^I(\hat{s}) \geq \overline{CS}^Z(\hat{s}),$$

and

$$\hat{s}_*^I \gtreqless \hat{s}_*^Z \quad \text{as} \quad \varepsilon \gtreqless 1. \qquad \text{QED.}$$

APPENDIX 4

PROOF OF PROPOSITION 4

We prove that $\rho(t)$ is constant for $t \in (t_0, t_n)$ and weakly decreasing for $t \in (t_n, t_r)$. The proof begins with two lemmas.

Lemma 1 $\phi(t)\rho(t) \le 1$

Proof A necessary condition for there to be no queue is $r(t) \le s$, or $\rho(t) \le 1/\phi$. By definition, this is true of $\phi(t)$, hence $\phi(t)\rho(t) \le 1$. QED.

Lemma 2 $[\phi(t)\rho(t) - 1]\dot{\phi}(t) \le 0$.

Proof Immediate if $\phi(t)\rho(t) = 1$. By Lemma 2 we need only consider $\phi(t)\rho(t) < 1$. By definition, $\phi(t)$ is the largest ϕ such that there is no queue at t, so with $\rho(t) < 1/\phi(t)$ the queue must be positive just before t. Let $\hat{t}$ be the last time when the queue was zero. By (3.1) in the text

$$\int_{\hat{t}}^{t} \phi(t)\rho(\tau)d\tau = t - \hat{t}.$$

Differentiating,

$$\dot{\phi}(t) = \frac{1-\phi(t)\rho(t)}{\int_{\hat{t}}^{t} \rho(\tau)d\tau} > 0,$$

and hence $[\phi(t)\rho(t) - 1]\dot{\phi}(t) < 0$. QED.

We now consider each of the 3 arrival intervals in turn.

(a) $t \in (t_0, t_n)$

By (3.14), Z(t) defined in (3.15) is constant. Differentiating, we find

$$\dot{\rho}(t) \int_{\phi(t)}^{\phi_M} \phi dJ(\phi) = (\phi(t)\rho(t) - 1)\frac{dJ(\phi)}{d\phi} \dot{\phi}(t) \qquad (A4.1)$$

By Lemma 2, the RHS of (A4.1) is nonpositive. The integral term on the LHS is strictly positive if $J(\phi(t)) < 1$, which is the case since otherwise there would never be a queue at $t > t_0$, and arrival at t would be preferable to arrival at t_0. Hence $\dot{\rho}(t) \le 0$, R(t) is concave and $\phi(t) = (t-t_0)/R(t)$. (A4.1) thus has a solution $\dot{\rho}(t) = 0$, which establishes that the arrival rate is constant for $t \in (t_0, t_n)$. QED.

(b) $t \in (t_n, t^*)$

Differentiating (3.17) and cancelling terms:

$$\left\{(\alpha-\beta)\int_{\phi(t)}^{\phi^*(t)} \phi dJ(\phi) + (\alpha+\gamma)\int_{\phi^*(t)}^{\phi_M} \phi dJ(\phi)\right\}\dot{\rho}(t) = \tag{A4.2}$$

$$(\alpha-\beta)[\phi(t)\rho(t)-1]\dot{\phi}(t)\frac{dJ(\phi(t))}{d\phi} + (\beta+\gamma)\phi^*(t)\rho(t)\dot{\phi}^*(t)\frac{dJ(\phi^*(t))}{d\phi}.$$

The term in braces on the LHS of (A4.2) is strictly positive if $J(\phi(t)) < 1$, which is the case by the reasoning given for $t \in (t_0, t_n)$.

$[\phi(t)\rho(t)-1]\dot{\phi}(t) \le 0$ by Lemma 2, so the first term on the RHS is nonpositive. The second term on the RHS is nonpositive iff $\dot{\phi}^*(t) \le 0$. By definition of $\phi^*(t)$, $t + \frac{Q(t,\phi^*(t))}{s} = t^*$. Differentiating and setting $\phi = \phi^*(t)$ one has

$$\dot{\phi}^*(t) = -\frac{\phi^*(t)\rho(t)}{\partial[Q(t,\phi)/s]/\partial\phi^*} < 0.$$

The RHS of (A4.2) is thus nonpositive, which establishes that $\dot{\rho}(t) \le 0$ for $t \in (t_n, t^*)$.

(c) $t \in (t^*, t_r)$

Differentiating (3.21):

$$\left\{ \int_{\phi(t)}^{\phi_M} \phi dJ(\phi) \right\} \dot{\rho}(t) = [\phi(t)\rho(t)-1]\dot{\phi}(t) \frac{dJ(\phi(t))}{d\phi}. \qquad (A4.3)$$

Since the LHS term in braces of (A4.3) is positive and the RHS is nonpositive

$$\dot{\rho}(t) \leq 0. \quad \text{QED.}$$

By arriving after t^* a user is always served late, so the late service effect operating through changes in $\phi^*(t)$ is absent, leaving only the queueing effect operating through changes in $\phi(t)$.

APPENDIX 5

PROOF OF LEMMA 1

Using (3.20) and (3.22), and the fact that with $t > t_r$, $R(t) = 1$ and hence $\phi(t) = t - t_0$, expected costs for arrival after t_r are

$$C(t) = \gamma\left\{\int_0^{t-t_0} (t-t^*)dJ(\phi) + \int_{t-t_0}^{\phi_M} (t_0+\phi-t^*)dJ(\phi)\right\} + \alpha\int_{t-t_0}^{\phi_M} (t_0+\phi-t)dJ(\phi). \quad (A5.1)$$

If t_r is indeed the last arrival time, then $C(t)$ must be nondecreasing after t_r. Since $C(t)$ is a continuous function of t it suffices to consider $\dot{C}(t)$ after t_r. From (3.21) and (3.15) the left-hand derivative (which is zero by construction) is:

$$\lim_{t\uparrow t_r} \dot{C}(t) = \gamma(1+\lim_{t\uparrow t_r} Z(t)) + \alpha \lim_{t\uparrow t_r} Z(t)$$

$$= \gamma - (\alpha+\gamma)(1-\lim_{t\uparrow t_r} J(t-t_0)) + (\alpha+\gamma)\lim_{t\uparrow t_r} \int_{t^*-t_0}^{\phi_M} \phi\rho(t)dJ(\phi) = 0. \quad (A5.2)$$

From (A5.1), the right-hand derivative is

$$\dot{C}(t) = \gamma - (\alpha+\gamma)(1-J(t_r-t_0)). \quad (A5.3)$$

Since $\dot{C}(t)$ is nonincreasing in t, a necessary and sufficient condition for equilibrium is

$$\lim_{t\downarrow t_r} \dot{C}(t) = \gamma - (\alpha+\gamma)(1-J(t_r-t_0)) \geq 0$$

or

$$J(t_r-t_0) \geq \frac{\alpha}{\alpha+\gamma}. \quad (A5.4)$$

But since the last term in (A5.2) is nonnegative

$$\lim_{t\uparrow t_r} J(t_r-t_0) \leq \frac{\alpha}{\alpha+\gamma}. \quad (A5.5)$$

Thus

$$t_r = t_0 + J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}, \tag{A5.6}$$

which is (3.28) in the text. A solution for t_0 and t_r now follows directly. From (3.13)

$$C(t_0) = \beta(t^*-t_0). \tag{A5.7}$$

From (3.20) and (3.22)

$$C(t_r) = \gamma(t_r-t^*) + (\alpha+\gamma)\int_{t_r-t_0}^{\phi_M}(t_0 + \phi - t_r)dJ(\phi)$$

$$= \gamma(t_r-t^*) - (\alpha+\gamma)(t_r-t_0)(1-J(t_r-t_0)) + (\alpha+\gamma)\int_{t_r-t_0}^{\phi_M}\phi dJ(\phi)$$

$$= -\gamma(t^*-t_0) + (\alpha+\gamma)\int_{J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}}^{\phi_M}\phi dJ(\phi) \quad \text{(using (A5.6))}.$$

Setting $C(t_0) = C(t_r) = C$:

$$C = \beta(t^*-t_0) = \beta\frac{\alpha+\gamma}{\beta+\gamma}\int_{J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}}^{\phi_M}\phi dJ(\phi), \tag{A5.8}$$

which is (3.29), and immediately yields (3.27). The initial premise that $t_r > t^*$ is satisfied provided:

$$t_r-t_0 > t^*-t_0 \Leftrightarrow J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\} > \frac{\alpha+\gamma}{\beta+\gamma}\int_{J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}}^{\phi_M}\phi dJ(\phi),$$

which is condition (3.26). This completes the proof of Lemma 1.

APPENDIX 6

PROOF OF LEMMA 2

If $t_r = t^*$

$$\phi(t_r) = \phi(t^*) = \phi^*(t_r) = \phi^*(t^*) = t^*-t_0 \text{ and } R(t^*) = 1. \tag{A6.1}$$

As in Appendix 5 a necessary and sufficient condition for equilibrium is (A5.4):

$$J(t_r-t_0) \geq \frac{\alpha}{\alpha+\gamma}. \tag{A6.2}$$

Using (A6.1) and (3.16) one also has

$$C(t_r) = C(t^*) = (\alpha+\gamma) \int_{t^*-t_0}^{\phi_M} (t_0+\phi-t^*)dJ(\phi). \tag{A6.3}$$

As before, $C(t_0) = \beta(t^*-t_0)$, which is (3.32). Equating $C(t_0)$ and $C(t^*)$ yields (3.31):

$$J(t^*-t_0) + \frac{1}{t^*-t_0} \int_{t^*-t_0}^{\phi_M} \phi dJ(\phi) - \frac{\alpha+\beta+\gamma}{\alpha+\gamma} = 0. \tag{A6.4}$$

Now the LHS of (A6.4) is strictly decreasing in t^*-t_0. Since $J(\phi)$ is monotonically increasing, (A6.2) is satisfied if the LHS of (A6.4) is nonnegative with $J(t^*-t_0) = \frac{\alpha}{\alpha+\gamma}$:

$$\frac{1}{J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}} \int_{J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}}^{\phi_M} \phi dJ(\phi) - \frac{\beta+\gamma}{\alpha+\gamma} \geq 0, \quad \text{or}$$

$$J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\} \leq \frac{\alpha+\gamma}{\beta+\gamma} \int_{J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}}^{\phi_M} \phi dJ(\phi), \tag{A6.5}$$

which is condition (3.30).

Page 59 blank

APPENDIX 7

PROOF OF PROPOSITION 2′

The proof of Prop. 2 for the static model given in Appendix 2 carries overly largely unchanged to the dynamic model. Expected consumers' surplus is still given by (2.23). The only difference is that expected user cost, given by (2.19), must be replaced by its dynamic counterpart. To reduce the notational burden, subscripts d and m denoting the day and message, and superscript u denoting the user are suppressed. By Prop. 5 there are two possibilities: $t_r > t^*$ and $t_r = t^*$. These must be considered separately.

(a) $t_r > t^*$

From (3.25) and (3.29)

$$\bar{p} = C = \beta(t^* - t_0) = \beta\frac{\alpha+\gamma}{\beta+\gamma}\int_{J^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}}^{\phi_M} \phi dJ(\phi).$$

Define $\theta \equiv \frac{n}{\sigma}$. Then $\phi \equiv \frac{N}{s} = \frac{n(\bar{p})^{-\varepsilon}}{\sigma\hat{s}} = \theta\frac{(\bar{p})^{-\varepsilon}}{\hat{s}}$. Let J_θ be the c.d.f. of θ from users' perspective. Then

$$\bar{p} = \beta\frac{\alpha+\gamma}{\beta+\gamma}\int_{J_\theta^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}} \theta\frac{(\bar{p})^{-\varepsilon}}{\hat{s}} dJ_\theta(\theta) = \left\{\delta\frac{\alpha+\gamma}{\gamma}\frac{1}{\hat{s}}\int_{J_\theta^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}} \theta dJ_\theta(\theta)\right\}^{\frac{1}{1+\varepsilon}}. \quad (A7.1)$$

Furthermore, $\frac{d\bar{p}}{d\hat{s}}\frac{\hat{s}}{\bar{p}} = -\frac{1}{1+\varepsilon}$, which establishes that (A2.3) and (A2.4) in Appendix 2 hold with $\eta = 1$.

(b) $t_r = t^*$

In this case

$$\bar{p} = \beta(t^* - t_0) = \beta\phi' \qquad \text{(A7.2)}$$

with ϕ' defined implicitly by (3.31)

$$J(\phi') + \frac{1}{\phi'}\int_{\phi'}^{\phi_M} \phi dJ(\phi) - \frac{\alpha+\beta+\gamma}{\alpha+\gamma} = 0. \qquad \text{(A7.3)}$$

Making the change of variable from ϕ to θ and defining $\theta' \equiv (p)^{\varepsilon}\hat{s}\phi'$, (A7.2) and (A7.3) can be rewritten

$$\bar{p} = \left\{\frac{\beta\theta'}{\hat{s}}\right\}^{\frac{1}{1+\varepsilon}}, \qquad \text{(A7.4)}$$

$$J(\theta') + \frac{1}{\theta'}\int_{\theta'}^{\phi_M} \theta dJ(\theta) - \frac{\alpha+\beta+\gamma}{\alpha+\gamma} = 0. \qquad \text{(A7.5)}$$

Since the distribution of ϕ is assumed to be independent of $\hat{s}$, so is the distribution of θ, and θ'. Thus, from (A7.4) $\frac{d\bar{p}}{d\hat{s}}\frac{\hat{s}}{\bar{p}} = -\frac{1}{1+\varepsilon}$, and (A2.3) and (A2.4) again hold with $\eta = 1$.

APPENDIX 8

PROOF OF PROPOSITION 3′

The proof follows closely the proof of Prop. 3 in Appendix 3.

With full information and $\eta = 1$, (2.24) yields

$$\overline{MCS}^F(\hat{s}) = \frac{1}{1+\varepsilon}\frac{1}{\hat{s}}\int_d\int_n\int_\sigma n\, p^{1-\varepsilon}\, dF_d(n,\sigma)\, d\Gamma(d),$$

Substituting for p with (2.10)

$$\overline{MCS}^F(\hat{s}) = \frac{1}{1+\varepsilon}\frac{1}{\hat{s}}\left\{\frac{\delta}{\hat{s}}\right\}^{\frac{1-\varepsilon}{1+\varepsilon}}\int_d\left\{\bar{n}_d\int_\theta \theta^{\frac{1-\varepsilon}{1+\varepsilon}}\, dJ_{d\theta}(\theta)\right\} d\Gamma(d).$$

With zero information and $\eta = 1$, (2.24) yields

$$\overline{MCS}^Z(\hat{s}) = \frac{1}{1+\varepsilon}\frac{1}{\hat{s}}\int_d \bar{n}_d\left(\bar{p}_d\right)^{1-\varepsilon} d\Gamma(d).$$

Then

$$\overline{MCS}^F(\hat{s}) - \overline{MCS}^Z(\hat{s}) = \int_d\left(\overline{MCS}^F_d(\hat{s}) - \overline{MCS}^Z_d(\hat{s})\right)d\Gamma(d), \tag{A8.1}$$

where

$$\overline{MCS}^F_d(\hat{s}) - \overline{MCS}^Z_d(\hat{s}) = \bar{n}_d\left[\left\{\frac{\delta}{\hat{s}}\right\}^{\frac{1-\varepsilon}{1+\varepsilon}}\int_\theta \theta^{\frac{1-\varepsilon}{1+\varepsilon}}\, dJ_{d\theta}(\theta) - \left(\bar{p}_d\right)^{1-\varepsilon}\right].$$

(a) $t_r > t^*$

From (A7.1)

$$\bar{p}_d = \left\{\delta\,\frac{\alpha+\gamma}{\gamma}\,\frac{1}{\hat{s}}\int_{J_{d\theta}^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}} \theta dJ_{d\theta}(\theta)\right\}^{\frac{1}{1+\varepsilon}}. \tag{A8.2}$$

Thus

$$\overline{MCS}_d^F(\hat{s}) - \overline{MCS}_d^Z(\hat{s}) \lesseqqgtr \int_\theta \theta^{\frac{1-\varepsilon}{1+\varepsilon}} dJ_{d\theta}(\theta) - \left\{ \frac{\alpha+\gamma}{\gamma} \int_{J_{d\theta}^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}} \theta dJ_{d\theta}(\theta) \right\}^{\frac{1-\varepsilon}{1+\varepsilon}} . \qquad (A8.3)$$

If $\varepsilon < 1$

$$\overline{MCS}_d^F(\hat{s}) - \overline{MCS}_d^Z(\hat{s}) \lesseqqgtr \left\{ \int_\theta \theta^{\frac{1-\varepsilon}{1+\varepsilon}} dJ_{d\theta}(\theta) \right\}^{\frac{1-\varepsilon}{1+\varepsilon}} - \frac{\alpha+\gamma}{\gamma} \int_{J_{d\theta}^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}} \theta dJ_{d\theta}(\theta).$$

But

$$\left\{ \int_\theta \theta^{\frac{1-\varepsilon}{1+\varepsilon}} dJ_{d\theta}(\theta) \right\}^{\frac{1-\varepsilon}{1+\varepsilon}} \leq \int_\theta \theta \, dJ_{d\theta}(\theta) \leq \frac{\alpha+\gamma}{\gamma} \int_{J_{d\theta}^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}} \theta dJ_{d\theta}(\theta) \text{ for } \varepsilon > 0,$$

where the first inequality follows by Hardy, Littlewood and Polya (HLP) (1934, Prop. 2.9.1), and the second because the last expression is a mean with the left-hand tail of the distribution truncated. Thus, if $\varepsilon < 1$

$$\overline{MCS}_d^F(\hat{s}) < \overline{MCS}_d^Z(\hat{s}) \quad \forall \ d,$$

and therefore

$$\overline{MCS}^F(\hat{s}) < \overline{MCS}^Z(\hat{s}) \quad \text{by (A8.1)},$$

$$\overline{CS}^F(\hat{s}) > \overline{CS}^Z(\hat{s}) \quad \text{by Prop. 2},$$

$$\hat{s}_*^F < \hat{s}_*^Z.$$

If $\varepsilon > 1$

$$\overline{MCS}_d^F(\hat{s}) - \overline{MCS}_d^Z(\hat{s}) \lesseqgtr \frac{\alpha+\gamma}{\gamma} \int_{J_{d\theta}^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}} \theta dJ_{d\theta}(\theta) - \left\{\int \theta^{\frac{1-\varepsilon}{1+\varepsilon}} dJ_{d\theta}(\theta)\right\}^{\frac{1-\varepsilon}{1+\varepsilon}}.$$

The proof concludes as for $\varepsilon < 1$.

(b) $t_r = t^*$

By (A7.4) and (A7.5)

$$\bar{p}_d = \left\{\frac{\beta\theta_d'}{\hat{s}}\right\}^{\frac{1}{1+\varepsilon}}, \qquad \text{(A8.4)}$$

with

$$J_{d\theta}(\theta_d') + \frac{1}{\theta_d'} \int_{\theta_d'}^{\phi_M} \theta dJ_{d\theta}(\theta) - \frac{\alpha+\beta+\gamma}{\alpha+\gamma} = 0. \qquad \text{(A8.5)}$$

Now

$$\frac{\beta+\gamma}{\gamma} - \frac{1}{\theta_d'}\int_0^\infty \theta dJ_{d\theta}(\theta) = 1 + \frac{\beta}{\gamma} - \left\{1 + \frac{\beta}{\alpha+\gamma} - J_{d\theta}(\theta_d') + \frac{1}{\theta_d'}\int_0^{\theta_d'} \theta dJ_{d\theta}(\theta)\right\}$$

$$= \frac{\beta\alpha}{\gamma(\alpha+\gamma)} + J_{d\theta}(\theta_d') - \frac{1}{\theta_d'}\int_0^{\theta_d'} \theta dJ_{d\theta}(\theta) > 0. \qquad \text{(A8.6)}$$

If $\varepsilon < 1$

$$\overline{MCS}^F_d(\hat{s}) - \overline{MCS}^Z_d(\hat{s}) \gtreqless \left\{ \int_\theta \theta^{\frac{1-\varepsilon}{1+\varepsilon}} dJ_{d\theta}(\theta) \right\}^{\frac{1+\varepsilon}{1-\varepsilon}} - \frac{\beta+\gamma}{\gamma} \theta'_d .$$

But

$$\left\{ \int_0^\infty \theta^{\frac{1-\varepsilon}{1+\varepsilon}} dJ_{d\theta}(\theta) \right\}^{\frac{1+\varepsilon}{1-\varepsilon}} \leq \int_\theta \theta \, dJ_{d\theta}(\theta) \leq \frac{\beta+\gamma}{\gamma} \theta'_d ,$$

where the first inequality follows again by HLP (Prop. 2.9.1) and the second from (A8.6). Therefore $\overline{MCS}^F_d(\hat{s}) < \overline{MCS}^Z_d(\hat{s})$, and the proof concludes as for $t_r > t^*$. The proof for $\varepsilon > 1$ is analogous.

APPENDIX 9

In this appendix we show that for the bivariate distribution of capacity given in (3.33), $(\bar{p})^{1-\varepsilon}$ varies with π as shown in Figure 5.

The first step is to determine when $t_r > t^*$. The condition given in (3.26) is $\tilde{\phi} > \frac{\gamma}{\beta+\gamma}\hat{\phi}$, or in terms of θ

$$J_\theta^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\} > \frac{\alpha+\gamma}{\beta+\gamma}\int_{J_\theta^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}} \theta dJ_{d\theta}(\theta). \tag{A9.1}$$

If $\pi > \frac{\gamma}{\alpha+\gamma}$ then $J_\theta^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\} = \frac{n}{\sigma}$ and (A9.1) is always satisfied. If $\pi \le \frac{\gamma}{\alpha+\gamma}$ then $J_\theta^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\} = n$, and (A9.1) reduces to

$$\pi < \pi_c \equiv \frac{\beta}{\alpha+\gamma}\,\frac{\sigma}{1-\sigma}. \tag{A9.2}$$

Thus, $t_r > t^*$ if $\pi \in [0, \pi_c) \cup (\frac{\gamma}{\alpha+\gamma}, 1]$. The complementary interval $[\pi_c, \frac{\gamma}{\alpha+\gamma}]$ within which $t_r = t^*$ is nonempty if $\sigma < \frac{\gamma}{\beta+\gamma}$, a condition we assume holds.

Now for $t_r > t^*$ we have by (A8.2)

$$\bar{p} = \left\{\delta\,\frac{\alpha+\gamma}{\gamma}\,\frac{1}{\hat{s}}\int_{J_\theta^{-1}\left\{\frac{\alpha}{\alpha+\gamma}\right\}} \theta dJ_\theta(\theta)\right\}^{\frac{1}{1+\varepsilon}}. \tag{A9.3}$$

If $\pi < \pi_c$, (A9.3) reduces to

$$\bar{p} = \left\{\frac{\delta n}{\hat{s}\sigma}\right\}^{\frac{1}{1+\varepsilon}}\left\{1 - \frac{(1-\pi)(\alpha+\gamma)-\alpha}{\gamma}(1-\sigma)\right\}^{\frac{1}{1+\varepsilon}}, \tag{A9.4}$$

and if $\pi > \frac{\gamma}{\alpha+\gamma}$

$$\bar{p} = \left\{\frac{\delta n}{\hat{s}\sigma}\right\}^{\frac{1}{1+\varepsilon}}. \tag{A9.5}$$

If $\pi \in [\pi_c, \frac{\gamma}{\alpha+\gamma}]$, $t_r = t^*$ and by (A8.4) and (A8.5)

$$\bar{p} = \left\{\frac{\beta\theta'}{\hat{s}}\right\}^{\frac{1}{1+\varepsilon}},$$

where

$$J_\theta(\theta') + \frac{1}{\theta'}\int_{\theta'} \theta dJ_\theta(\theta) - \frac{\alpha+\beta+\gamma}{\alpha+\gamma} = 0.$$

(A9.3) reduces to

$$\bar{p} = \left\{\frac{\delta n}{\hat{s}\sigma}\right\}^{\frac{1}{1+\varepsilon}}\left\{\frac{(\beta+\gamma)(\alpha+\gamma)}{\gamma[\beta+(\alpha+\gamma)\pi]}\pi\right\}^{\frac{1}{1+\varepsilon}}. \tag{A9.6}$$

Figure 3 follows by plotting (A9.4), (A9.6) and (A9.5) over the three respective intervals $[0,\pi_c)$, $[\pi_c, \frac{\gamma}{\alpha+\gamma}]$ and $(\frac{\gamma}{\alpha+\gamma}, 1]$. It is straightforward to establish

$$\lim_{\pi\downarrow\pi_c} \frac{d(\bar{p})^{1-\varepsilon}}{d\pi}\Big|(A9.6) \gtrless \lim_{\pi\uparrow\pi_c} \frac{d(\bar{p})^{1-\varepsilon}}{d\pi}\Big|(A9.4) \quad \text{as} \quad \varepsilon \lessgtr 1,$$

so that the curves in Figure 3 are kinked at $\pi = \pi_c$ as shown.

ENDNOTES

[1]In Nims (1981, p.202).

[2]Henderson (1977, 1981), Hendrickson and Kocur (1981), Hurdle (1981), Fargier (1983), Mahmassani and Herman (1984), Newell (1987), de Palma and Arnott (1990), Braid (1987, 1989), Arnott *et al*. (1988a, 1990a, 1990b, 1991a, 1991b, 1991c).

[3]Among these can be mentioned the Autoguide project conducted in London, the Comprehensive Automobile Traffic Control (CACS) study carried out by MITI in Japan, the ALI-SCOUT Destination Guidance System in (formerly) West Germany, the European PROMETHEUS and DRIVE projects, and the U.S. ETAK system which is being tested in the San Francisco and Los Angeles areas. Descriptions of these projects, and of the RGS technologies, are found in Boyce (1988), OECD (1988) and Hoffmann (1991).

[4]Peak-load pricing is more common in private sector facilities.

[5]As shown by Wilson (1989), efficient rationing can sometimes be effected by priority service pricing without need for either a spot market or for users to be informed in advance of consumption. Priority pricing is more suited to loss systems such as electricity service, where supply can be interrupted instantaneously and temporarily, than delay systems, where individuals often incur costs of commitment to usage before learning what state has been realized, and where the set of individuals using the facility and their relative preferences for service reliability are less stable over time.

[6]The computer model used by Smith and Krutilla (1976) to simulate movements of users through recreational areas has been described as a "traffic simulation model" (Cicchetti and Smith (1976, p.199)). Flow congestion occurs on highways when vehicle densities exceed a critical level, whereas queueing occurs at bottlenecks. Flow congestion occurs on ski slopes and nature trails when hiking and backpacking parties meet or overtake each other. At entry points and other locations where usage is concentrated, bottlenecking and delays can occur. Long waits may also develop where access is rationed first-come-first-serve by authorities, or by the capacity of parking lots.

[7]For example, state wildlife management agencies maintain data bases on hunting and fishing success rates by region. The U.S. National Weather Service keeps extensive historical data on wind patterns, temperatures, cloud formations and rainfall that can be used by recreationists to make long-range travel plans.

[8]These two views corresond roughly to 'market uncertainty' and 'event uncertainty' as defined by Hirshleifer and Riley (1979, p.1377).

[9]In most of the peak-load pricing literature it is assumed that the quality of service is constant up to 'capacity', and that output beyond this is impossible. The formulation here, in which the quality of service degrades smoothly with the level of usage, is consistent with Panzar (1976) and Burness and Patrick (1990).

[10]This point is raised by d'Ouville and McDonald (1990).

11 As Bowden (1985) has noted, accurate forecasts may be impossible even if potential attendants are interviewed repeatedly about their intentions.

[12]The form of the message system is similar to that considered by Nelson and Winter (1964) and Marschak and Miyasawa (1968). If demand or capacity is serially correlated, users can update the respective probability distributions using the recent history of realized states. Autocorrelation will not be treated explicitly here.

[13]Proof: $d\hat{F}_d(n,\sigma) = \int_m d\hat{F}_{dm}(n,\sigma)\, dM(m|d) = \int_m dF_{dm}(n,\sigma)\, dM(m|d)$

$$= \int_m dF_d(n,\sigma)\, dM(m|d,n,\sigma) = dF_d(n,\sigma) \int_m dM(m|d,n,\sigma) = dF_d(n,\sigma),$$

where the second equality follows from (2.4) and the third from (2.5), While unbiasedness would appear to be a *sine qua non* of a message system, it does not always hold in practice. For example, it has been observed that weather forecasters often over-forecast precipitation. Broadus and Solow (1988) suggest that this could be because forecasters apply an asymmetric loss function to their forecast errors.

14 See, for example, Cooper (1981, p.57). This point initially escaped us, which led to errors in an earlier version of the paper (Arnott *et al.* (1988b)).

[15]This is at best an approximation if capacity comes in discrete units (*e.g.* traffic lanes) and if each unit is working either fully or not at all. This problem is bypassed here by assuming capacity is a continuous variable.

[16]In the case of electricity demand, which has strong and regular cycles, it is optimal to have a 'diverse' technology (Chao (1983)). The base load is served by capacity with high fixed costs but low marginal cost, whereas demand peaks are met with capacity having low fixed but high marginal costs. The same may be true of other facilities. Our assumption of a constant marginal capacity cost is for simplicity.

17 We are grateful to Professor Stan Teply of the University of Alberta Civil Engineering Department for clarifying the rules used by traffic engineers for road capacity design.

18 Use of such a rule of thumb might be justified by data collection and processing costs. However, data on the complete distribution of traffic flow is often assembled (*viz.* HCM, Section 2).

19 Rules analogous to the nth busiest hour of the HCM are used in the design of parking facilities: see Frantzeskakis (1982, p.22) and Smith (1983, p.441). Analogous rules are also used to choose reserve levels for storable outputs. For example, British Gas holds sufficient gas reserves to meet demand in a cold winter occurring once in 50 years (Cannon (1987)). Similarly, water utilities may construct sufficient reservoir capacity to meet a once-in-50-years drought (Crew and Kleindorfer (1986, p.260)). In cases such as these, where only extreme values of the distribution matter, decision rules based on a particular fractal of the distribution may be justifiable.

20 Work by McFadden (1974), Pucher and Rothenberg (1976) and Small (1983) suggests that the elasticity is about 0.2.

21 A similar result was derived by Kay (1979), who considered optimal capacity under demand uncertainty for an electric utility which sets welfare-maximizing rates paid conditional on service. Kay showed (p.611) that, with isoelastic demand, optimal capacity under uncertainty is greater than riskless capacity if, but not only if, $\varepsilon < 1$. Capacity may be lower under uncertainty if $\varepsilon > 1$, if the load characteristics of incremental demand and poor and the costs of denied service modest.

[22]This result may be compared with that of Brown and Johnson (1969), who considered a public utility which sets price before demand intensity is known. In their model, output is adjusted to meet demand at the set price, but only up to the capacity limit. If demand exceeds capacity, rationing necessarily occurs since capacity is fixed. (Thus, theirs is a loss system, whereas ours is a delay system.) On the assumption of a linear demand curve, and that supply is rationed to users with the highest willingness to pay, Brown and Johnson showed that optimal capacity is unambiguously increased by demand uncertainty. Visscher (1973) later showed that with random rationing, or rationing to users with the lowest willingness to pay, optimal capacity may be lower with demand variability.

Prop. 1 may also be compared with Kraus (1982), who considered a situation in which travel demand is constant but known only imperfectly to a planner. On the assumption of a unitary demand elasticity he showed that optimal road capacity is increased by the planner's uncertainty. By contrast, d'Ouville and McDonald (1990) assumed that individual travel demand is stochastic. On the assumption that individual travel costs are a quadratic function of the number of drivers they show that optimal capacity is increased by demand variability.

[23]In Section 2.2 it was assumed that $F_d(n,\sigma)$ is independent of $\hat{s}$. We now assume this is true of $F_{dm}(n,\sigma)$ and $M(m|d)$.

[24]Queueing models has been used to consider various aspects of peak-period traffic congestion; e.g. the morning rush hour (Hendrickson and Kocur (1981)), the afternoon rush hour (Fargier (1983)), heterogeneous drivers (Cohen (1987), Newell (1987), Arnott *et al.* (1988)), tolling (Arnott (1990a, 1991b)) and simple networks (Braid (1987), Arnott *et al.* (1990b)).

[25]If there are no such restrictions, the whole road is in effect the bottleneck. If there is more than one bottleneck, but only one entry and exit point, then with pure queueing congestion the flow capacity is the capacity of the smallest bottleneck.

[26]It is more common in the traffic engineering and transport economics literatures to assume flow rather than queueing congestion. However, several recent empirical studies have found that travel speed on freeways declines only slightly with flow until capacity is approached, and that the discharge rate of vehicles from a queue is equal to or only slightly below free-flow capacity (see, for example, Hurdle and Datta (1983), Hurdle and Solomon (1986), Banks (1990) and Hall and Hall (1990)). This is consistent with the properties of the bottleneck model that travel time not spent in queues is constant, and that the maximum service rate of a bottleneck is the same with and without a queue.

27 The term 'schedule *delay* cost', which is standard in the literature, is perhaps misleading in that it refers to the cost of being early as well as late. An insightful discussion of schedule delay in air travel is found in Douglas and Miller (1974); for a welfare analysis see Panzar (1979).

In the case of commuting to work it is natural to express time-of-use preferences in terms of arrival time, when use of the road *ends*. In other contexts, *e.g.* the afternoon commute and perhaps long-distance telephone calling, preferences may be more strongly associated with the time use is initiated.

The assumption that everyone has the same t^* is for simplicity. While the shape of the distribution of t^* affects equilibrium schedule delay costs, it does not affect the evolution of the queue provided the distribution is not too spread out; see Vickrey (1969), Newell (1987) and Arnott *et al.* (1988a).

28 The schedule delay cost function (3.2) is (once) differentiable at the point of zero schedule delay iff $\eta > 1$. If individuals are indifferent as to time of use within some time interval, and incur linear schedule delay costs outside it, the nonlinear specification with $\eta > 1$ may serve as a better approximation than $\eta = 1$. Small (1982) found that some commuters in his sample did indeed experience such a threshold effect for lateness.

29 The case $\alpha \le \beta$, for which an equilibrium can exist only when users arrive *en masse*, is discussed in Arnott *et al.* (1985).

30 The full information equilibrium of the dynamic model can be treated using a static model whatever the form of the SDC function. Let $D(t-t^*)$ be an arbitrary SDC function. In equilibrium, $C = D(t_0-t^*) = D(t_0+\frac{N}{s}-t^*)$, which can be solved for t_0 and an equilibrium cost function $\bar{C}(\frac{N}{s})$.

31 Pure strategies are more realistic than mixed strategies if individuals prefer a routine (conditional on the day and message) such as to arrive at work early every day.

32 $$J(\phi) = \Pr\{\frac{N}{s} \le \phi\} = \Pr\{n \le (\bar{p})^{\varepsilon}\phi\hat{s}\sigma\} = \int_{\sigma}\int_{n=0}^{(\bar{p})^{\varepsilon}\phi\hat{s}\sigma} dF^{\upsilon}(n,\sigma) = \int_{\sigma}\int_{n=0}^{(\bar{p})^{\varepsilon}\phi\hat{s}\sigma} \frac{N}{\bar{N}}\, dF(n,\sigma)$$

$$= \int_{\sigma}\int_{n=0}^{(\bar{p})^{\varepsilon}\phi\hat{s}\sigma} n(\bar{p})^{-\varepsilon}dF(n,\sigma) \Big/ \int_{n}\int_{\sigma} n(\bar{p})^{-\varepsilon}dF(n,\sigma) = \frac{1}{\bar{n}}\int_{\sigma}\int_{n=0}^{(\bar{p})^{\varepsilon}\phi\hat{s}\sigma} n\, dF(n,\sigma).$$

[33]In Figure 4 it is assumed that $t_r > t$, although as will be shown $t_r = t^*$ is possible.

[34]It can be shown that $\rho(t)$ is concave for any convex SDC function. While the proof for the general case is in fact quicker than for the linear case it is not possible to carry through the subsequent calculations for the general case; moreover, the intermediate steps for the linear function are required to solve it.

[35]Restoring the d and m subscripts we have for day d

$$\int_m \hat{\phi}_{dm}\, dM(m|d) = \int_m \frac{\alpha+\gamma}{\gamma} \int_{J^{-1}_{dm}\left\{\frac{\alpha}{\alpha+\gamma}\right\}}^{\phi_M} \phi\, dJ_{dm}(\phi)\, dM(m|d)$$

$$> \int_m \int_0^{\phi_M} \phi\, dJ_{dm}(\phi)\, dM(m|d) = \int_0^{\phi_M} \phi \int_m dJ_{dm}(\phi)\, dM(m|d) = \int_0^{\phi_M} \phi\, dJ_d(\phi).$$

[36]Small (1982, Table 2, model 1) reports estimates of $\beta/\alpha = 0.61$ and $\gamma/\alpha = 2.38$, which yield $\frac{\gamma}{\beta+\gamma} = 0.796$. The behavior shown in Figure 5 thus results if more than about 20% of design capacity is lost in a capacity reduction.

[37]The nonconcavity here is sharper than that identified by Radner and Stiglitz (1984), where the value of information *gross* of cost is convex but nonnegative.

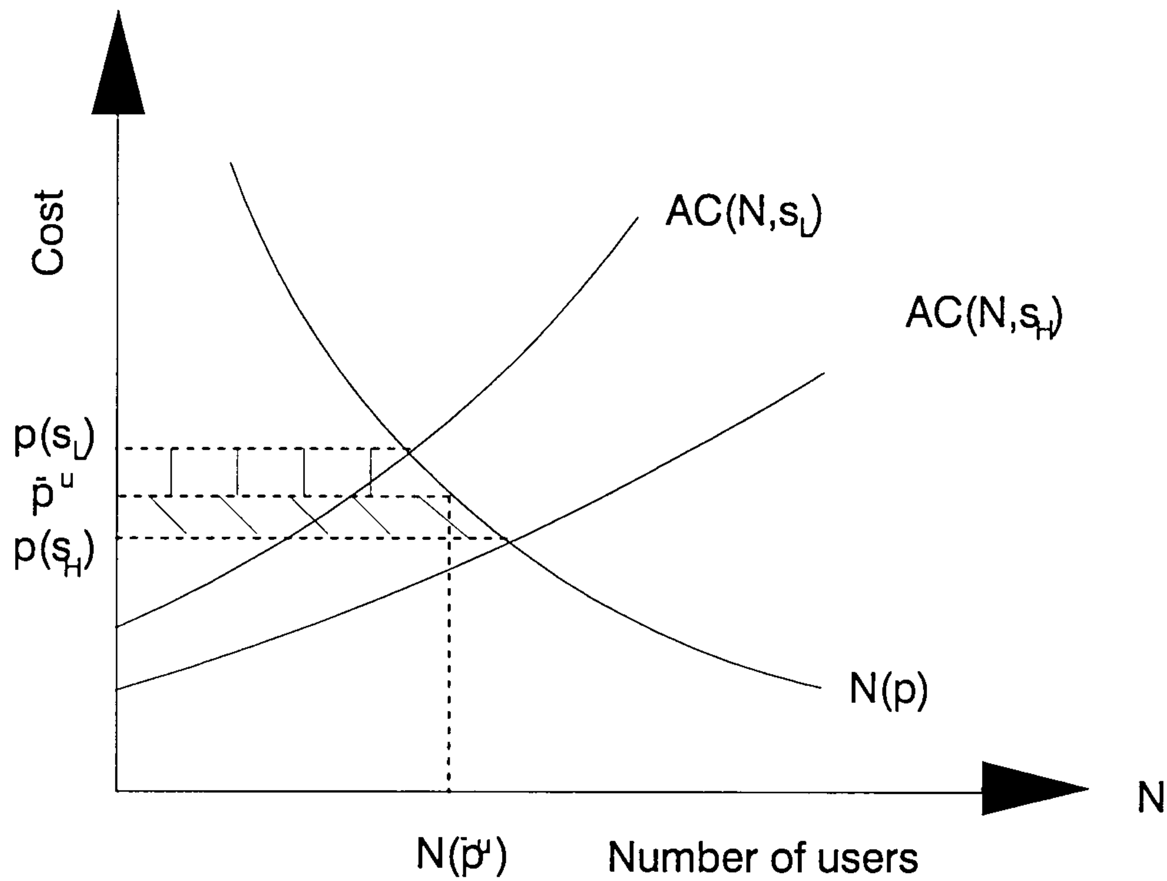
[38]If $\varepsilon > 0$, and $|\pi-\pi_c|$ is small but positive, it is possible that both weakly informative and fully informative message systems improve welfare, but a moderately informative system reduces it. The value of information can thus be multiple-peaked.

[39]A conceptual analysis of multiple congested recreational areas has been undertaken by McConnell and Sutinen (1984). We have extended the bottleneck queueing model to a travel corridor with two routes in parallel in Arnott *et al.* (1990b), and to two routes and heterogeneous users in Arnott *et al.* (1991c).

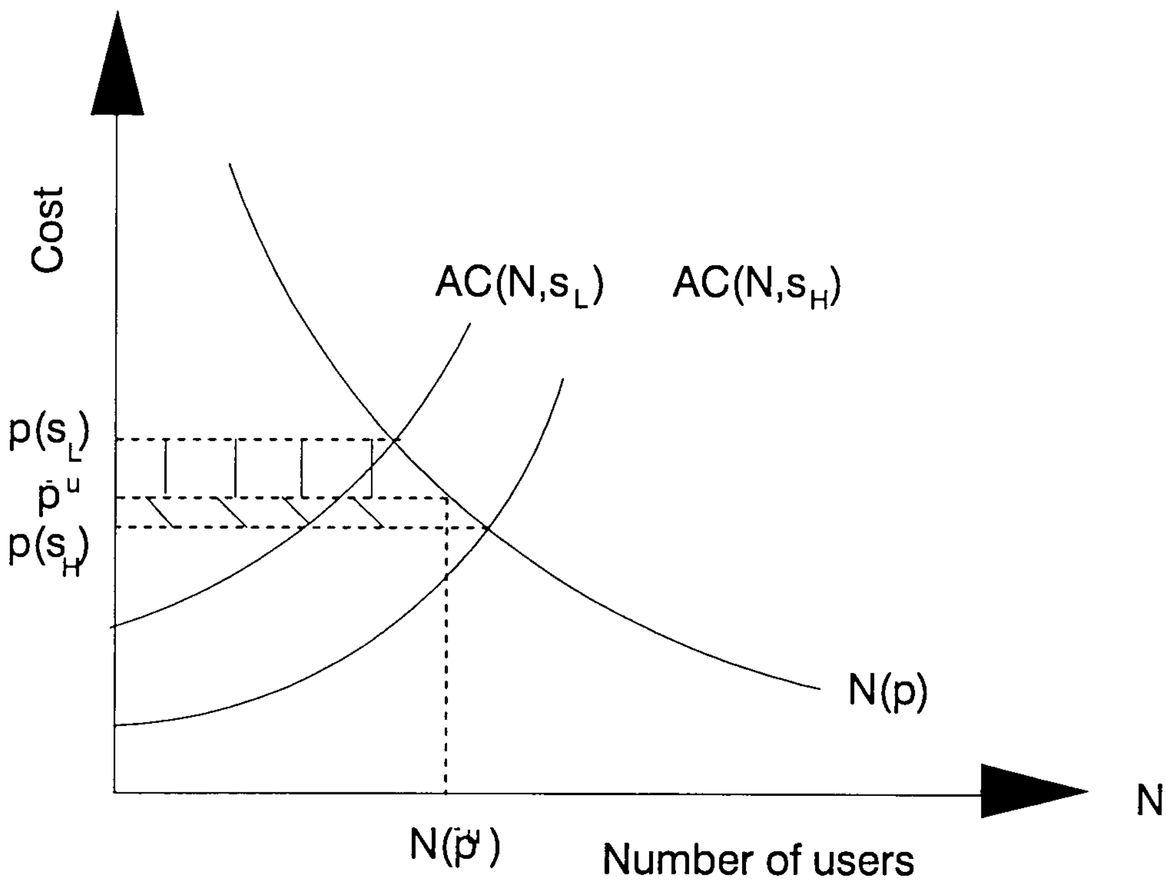
FIGURE 1

Welfare comparison of full information and zero information

(a) Homogeneous user cost, constant elasticity demand



(b) Nonhomogeneous User Cost



(c) Highly Convex Demand

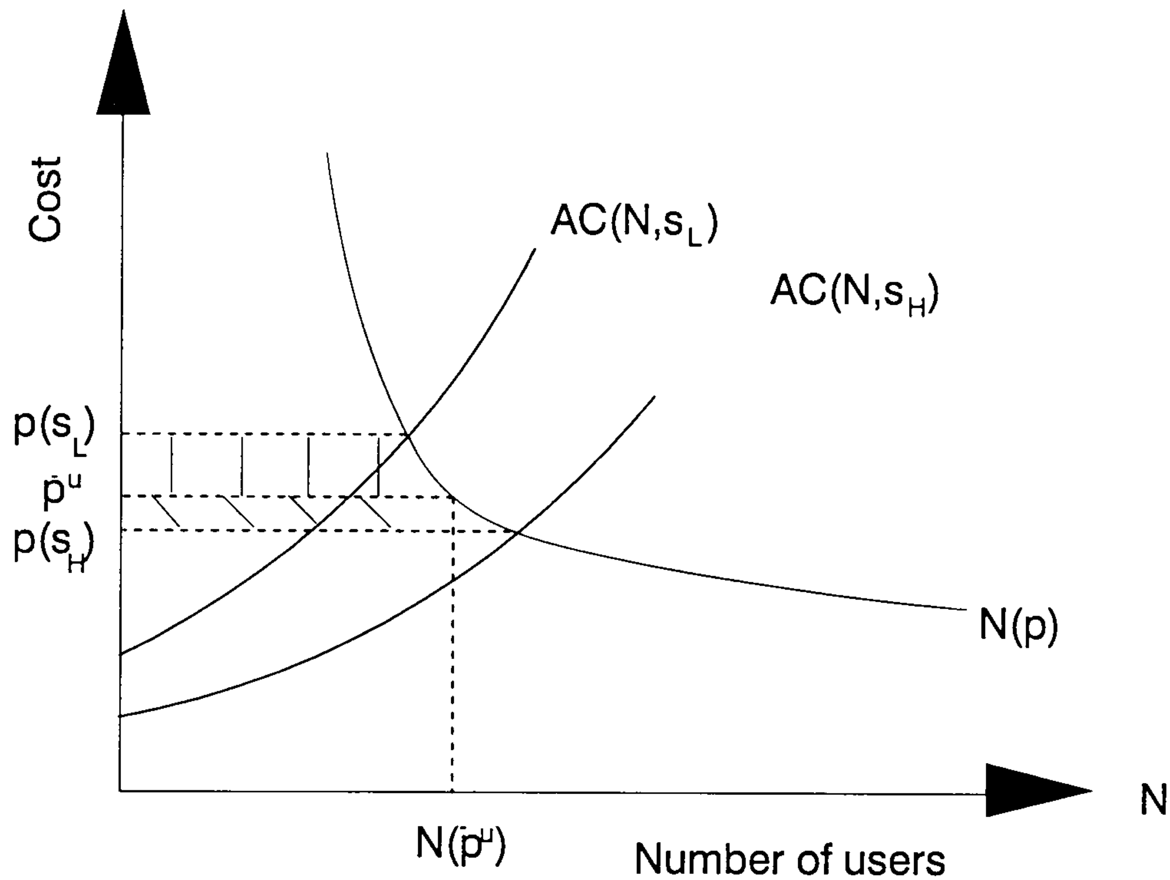
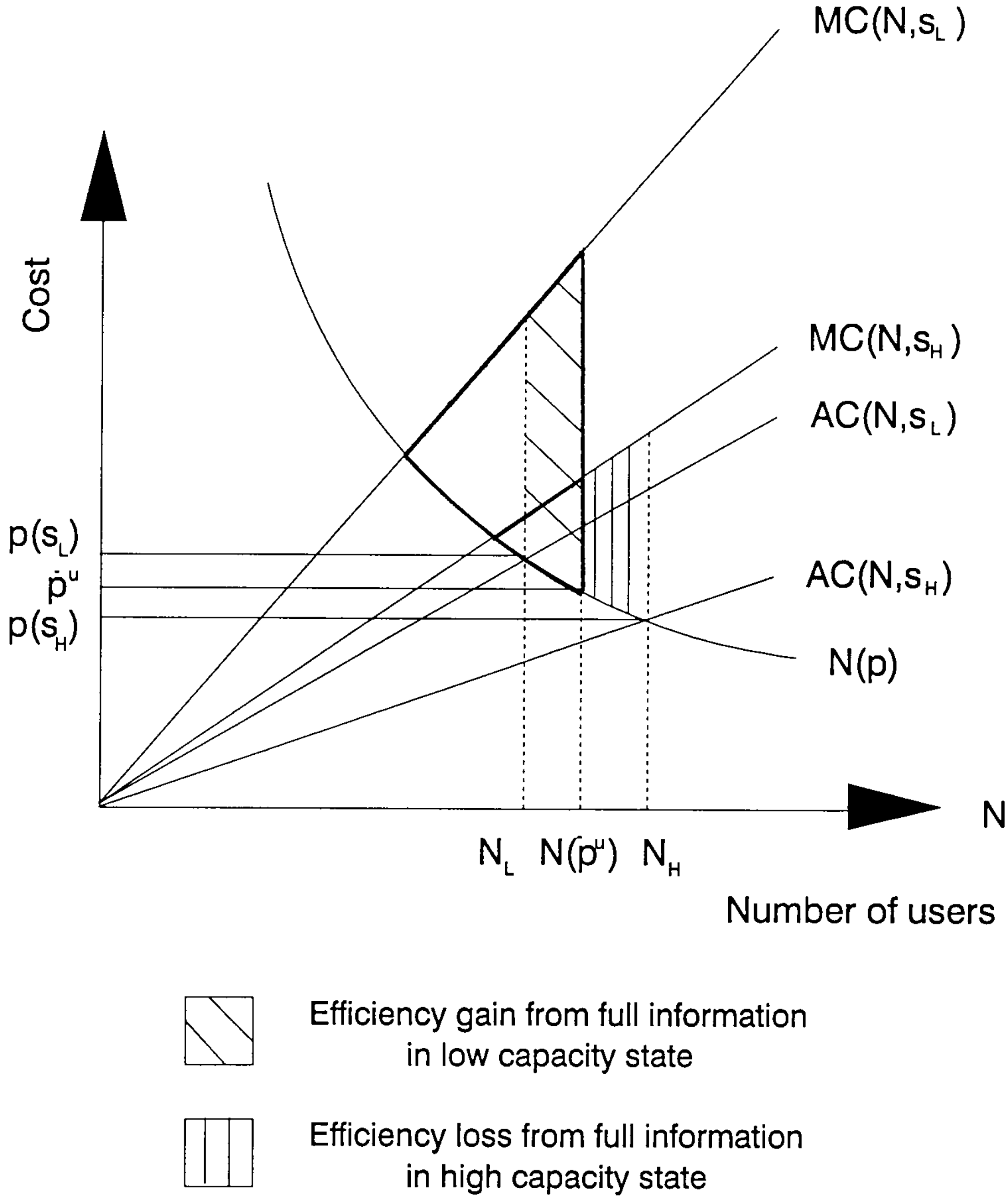


FIGURE 2

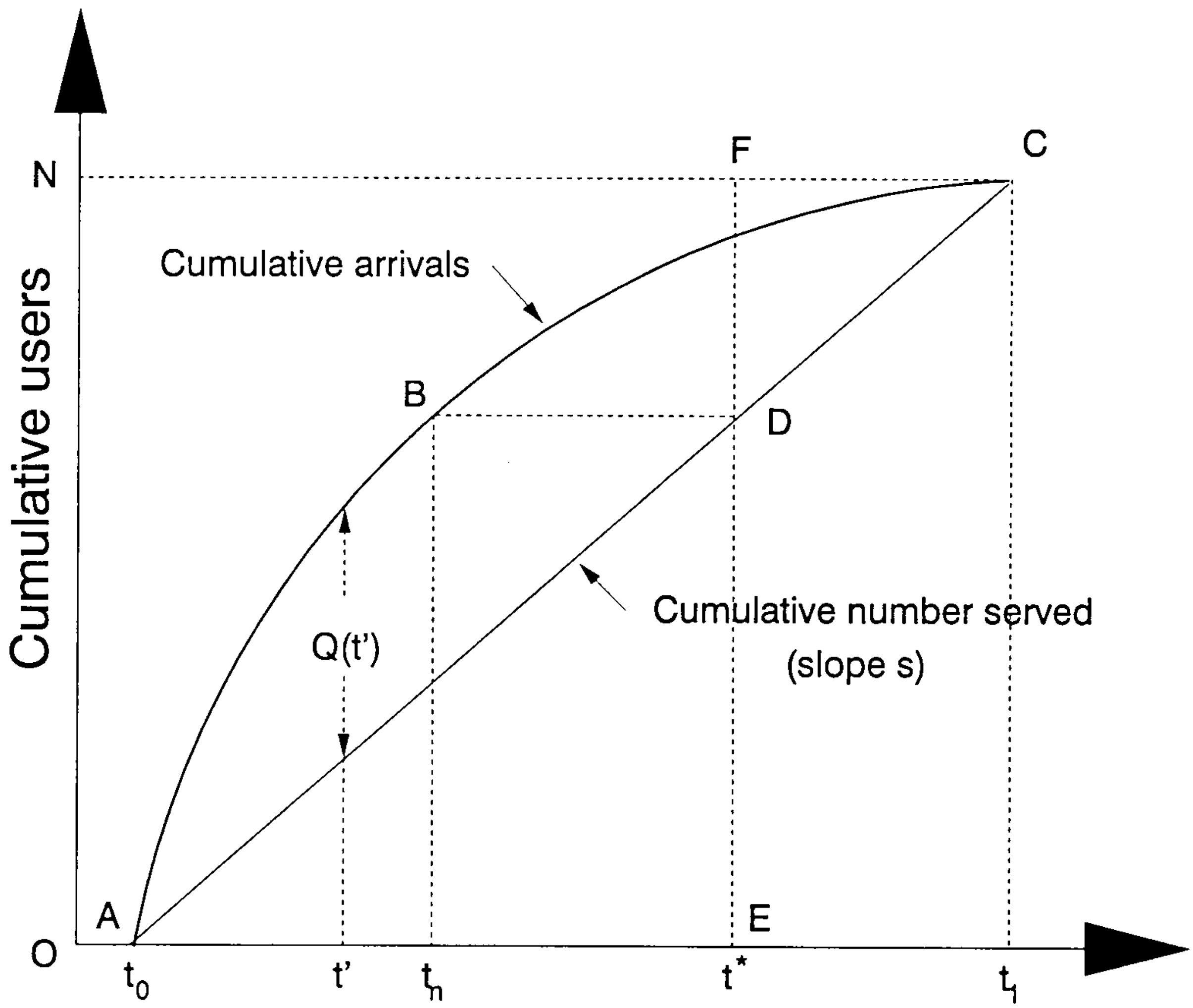


MC(N,s_L)
MC(N,s_H)
AC(N,s_L)
AC(N,s_H)
N(p)
Cost
p(s_L)
p^u
p(s_H)
N
N_L
N(p^u)
N_H
Number of users
Efficiency gain from full information
in low capacity state
Efficiency loss from full information
in high capacity state

FIGURE 3

Dynamic Equilibrium with Full Information

($\eta > 1$)



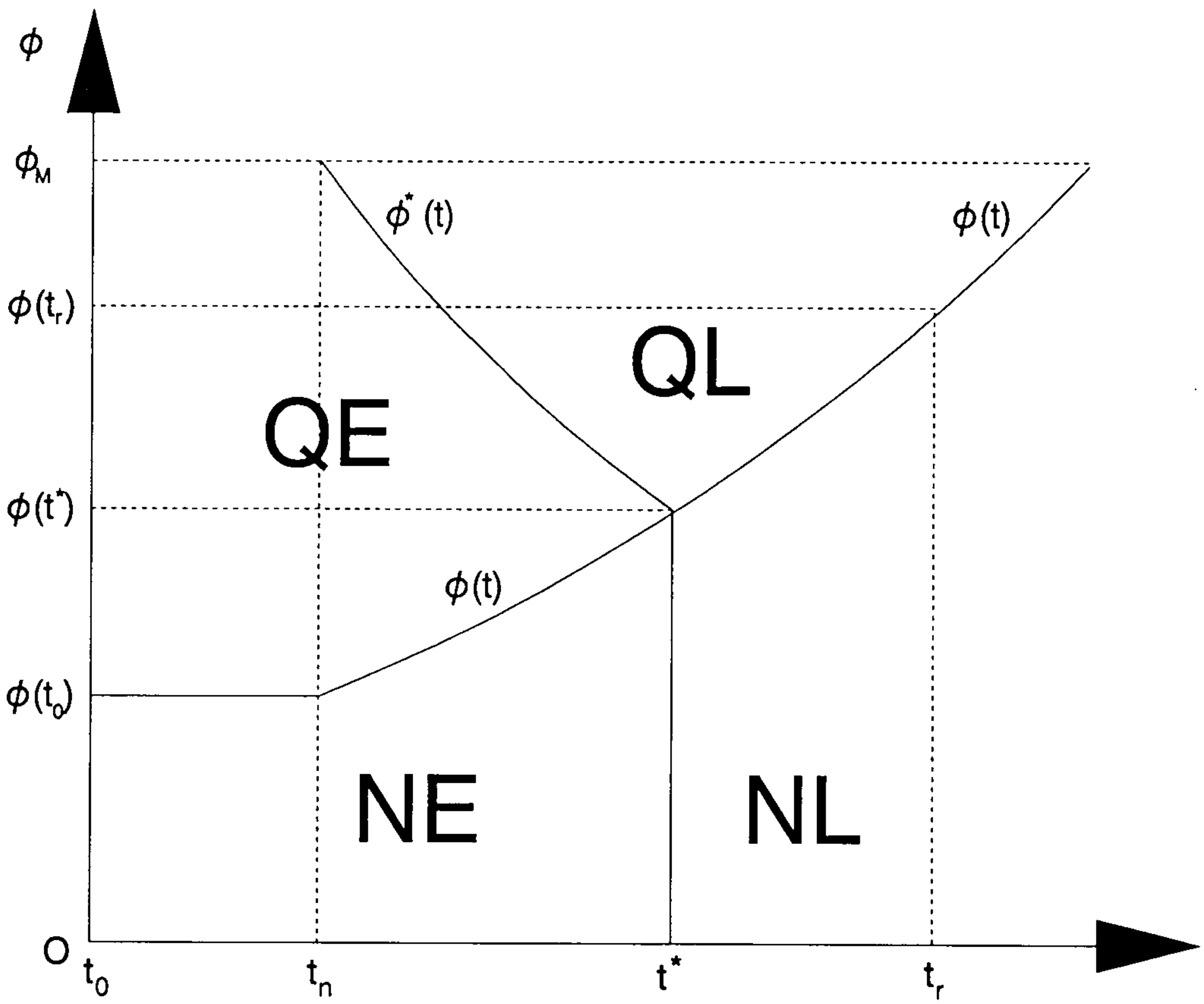
Total queueing time = ABCA

Total time early = ADEA

Total time late = CDFC

FIGURE 4

Queue Encountered and Time of Service as a Function of Arrival Time



Key: NE: No queueing, served early

QE: Queueing, served early

NL: No queueing, served late

QL: Queueing, served late

FIGURE 5
Welfare Effect of Imperfect Information
(Demand fixed, capacity stochastic)

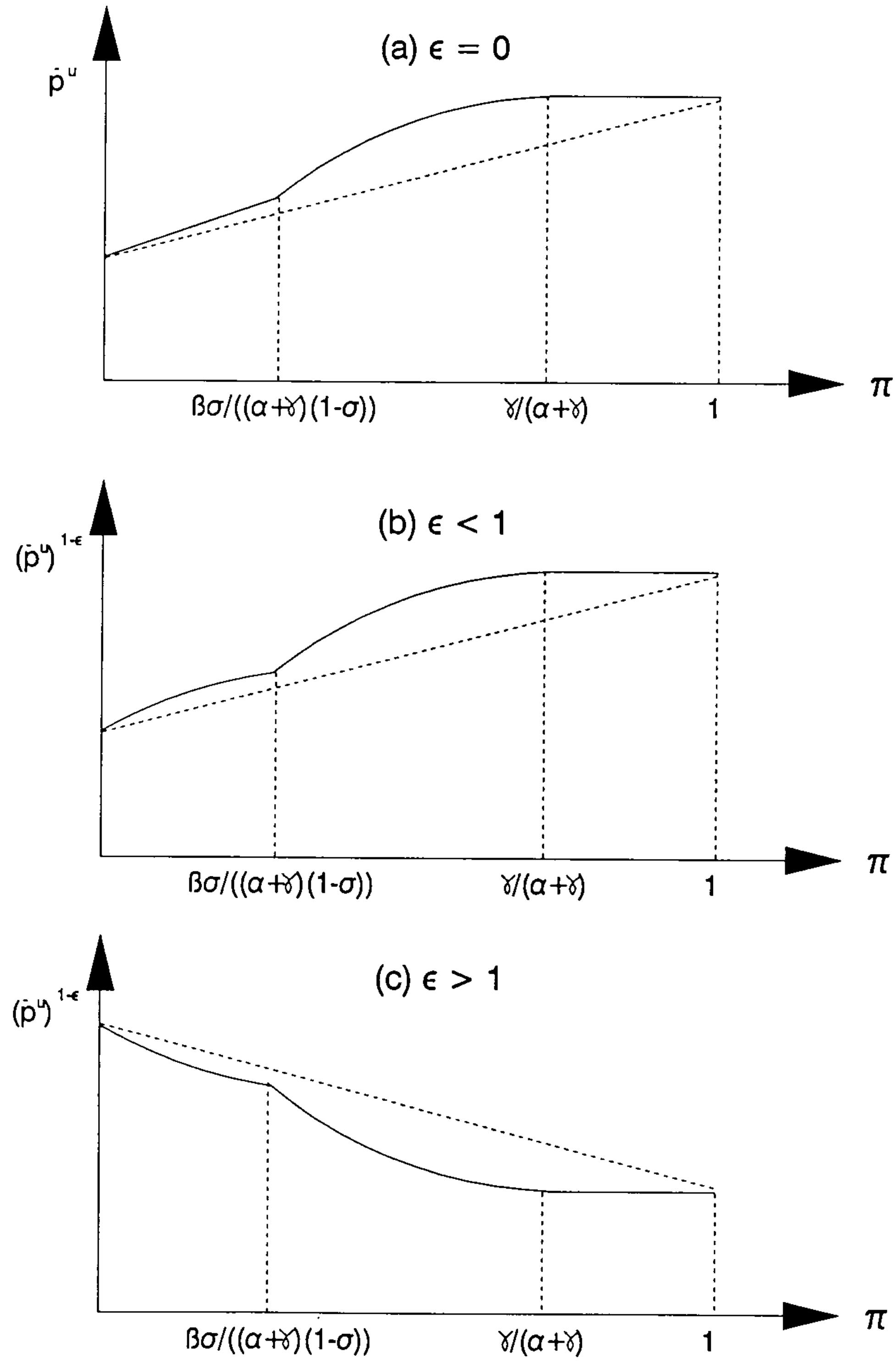
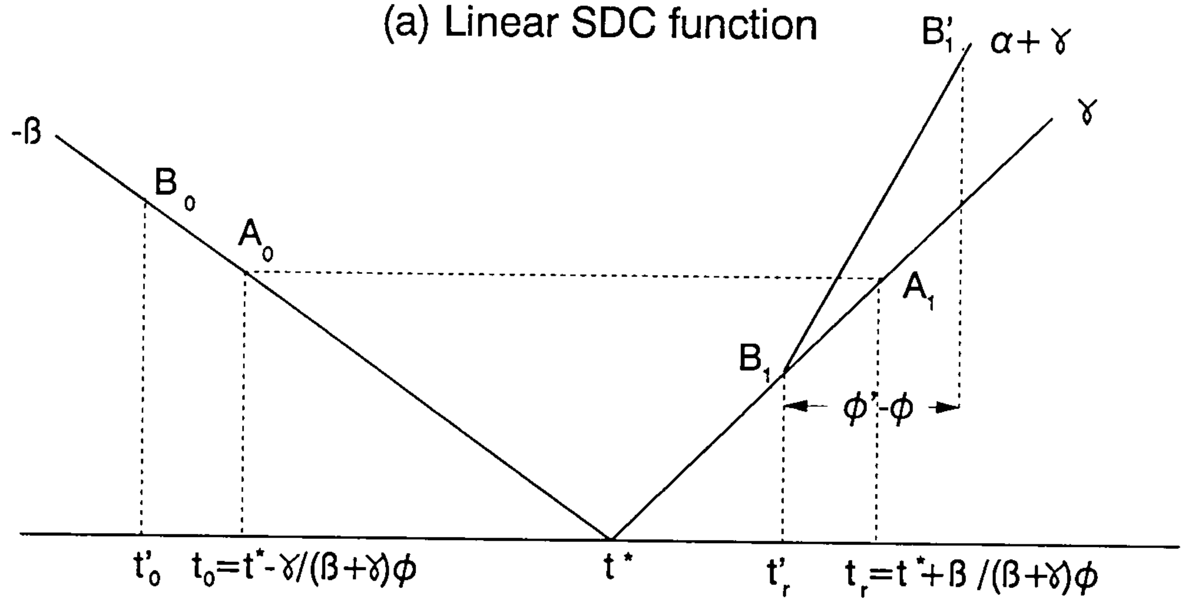


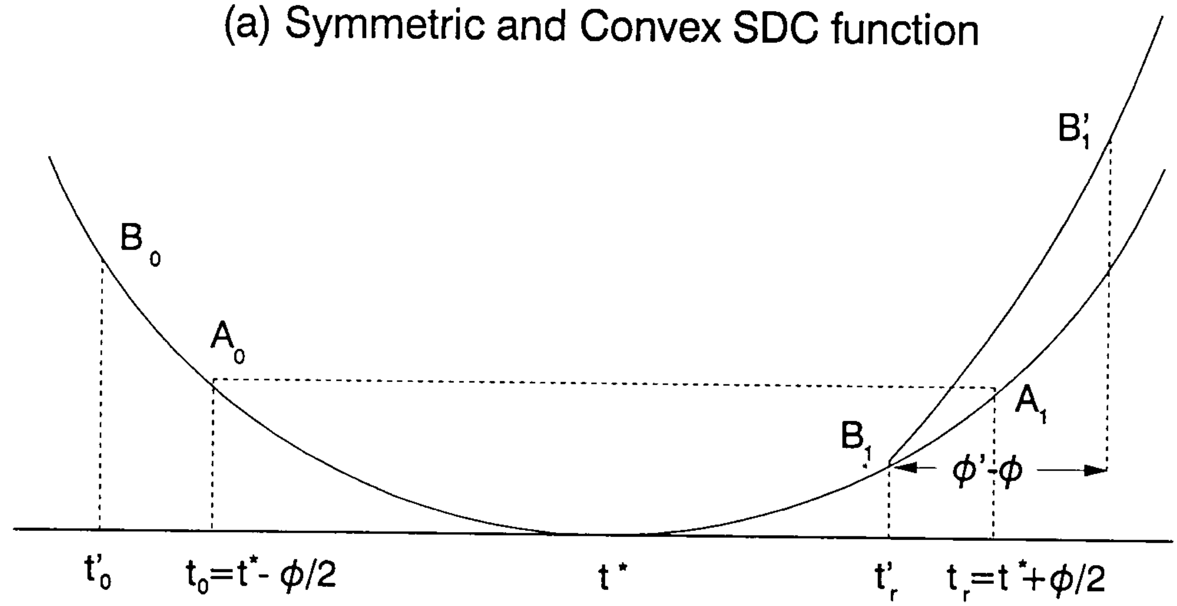
FIGURE 6

Time of Usage with Two-point ϕ Distribution in Zero Information Equilibrium

(a) Linear SDC function



(a) Symmetric and Convex SDC function



REFERENCES

Abkowitz, M. (1981a), "Understanding the Effect of Transit Service Reliability on Work Travel Behavior", Transporation Research Record 794 (1981), 33-41.

Abkowitz, M. (1981b), "An Analysis of the Commuter Departure Time Decision", Transportation 10, 283-297.

Alfa, A. and D. Minh (1979), "A Stochastic Model for the Temporal Distribution of Traffic Demand - The Peak Hour Problem", Transportation Science 13, 315-324.

Arnott, R., A. de Palma and R. Lindsey (1985), "Economics of a Bottleneck", Queen's University, Institute for Economic Research, discussion paper #636.

Arnott, R., A. de Palma and R. Lindsey (1988a), "Schedule Delay and Departure Time Decisions with Heterogeneous Commuters", Transportation Research Record 1197, 56-67.

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

Arnott, R., A. de Palma and R. Lindsey (1990a), "Economics of a Bottleneck", Journal of Urban Economics 27, 111-130.

Arnott, R., A. de Palma and R. Lindsey (1990b), "Departure Time and Route Choice for Routes in Parallel", Transportation Research B 24B(3), 209-228.

Arnott, R., A. de Palma and R. Lindsey (1991a), "Does Providing Information to Drivers Reduce Traffic Congestion?", Transportation Research A 25A(5), 309-318.

Arnott, R., A. de Palma and R. Lindsey (1991b), "A Structural Model of Peak-Load Congestion: A Traffic Bottleneck with Elastic Demand", American Economic Review, forthcoming.

Arnott, R., A. de Palma and R. Lindsey (1991c), "Route Choice with Heterogeneous Drivers and Group-Specific Congestion Costs", Regional Science and Urban Economics, forthcoming.

Banks, J.H. (1990), "Flow Processes at a Freeway Bottleneck", paper presented at the 69th Annual Meeting of the Transportation Research Board, Washington, D.C.

Ben-Akiva, M., A. de Palma and P. Kanaroglou (1986), "Dynamic Model of Peak Period Traffic Congestion with Elastic Arrival Rates", Transportation Science 20, 164-181.

Ben-Akiva, M., A. de Palma and I. Kaysi (1991), "Dynamic Network Models and Driver Information Systems", Transportation Research A 25A(5), 251-266.

Bowden, R.J. (1985), "Reflexive Forecasting of 'One-off' Events", mimeo.

Boyce, D. (1988), "Route Guidance Systems for Improving Urban Travel and Location Choices", Transportation Research A 22(A), 275-281.

Braid, R.M. (1987), "Peak-Load Pricing of a Transport Facility with an Unpriced Substitute", mimeo.

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

Broadus, J.M. and A.R. Solow (1988), "Towards a Policy for Global Climate Change", Association of Environmental and Resource Economics Newsletter 8(2), November, 3-8.

Brown, G. Jr. and M. B. Johnson (1969), "Public Utility Pricing and Output under Risk", American Economic Review 59 (March), 119- 128.

Burness, H.S. and R.H. Patrick (1990), "Peak Load Pricing with Continuous Interdependent Demand", Dept. of Mineral Economics, Colorado School of Mines.

Camerer, C., G. Loewenstein and M. Weber (1989), "The Curse of Knowledge in Economic Settings: An Experimental Analysis", Journal of Political Economy 97(5), 1232-1260.

Cannon, C.M. (1987) "Peak-pricing and Self-Rationing of Gas", Energy Economics 9(2), 99-103.

Chao, Hung-Po (1983), "Peak Load Pricing and Capacity Planning with Demand and Supply Uncertainty", Bell Journal of Economics 14(1), 179-190.

Cicchetti, C.J. and V. Kerry Smith (1976) "The Measurement of Individual Congestion Costs: An Economic Application to Wilderness Recreation", in Steven A.Y. Lin (ed.) Theory and Measurement of Economic Externalities, New York: Academic Press, 183-200.

Coate, S. and J.C. Panzar (1988), "Public Utility Pricing and Capacity Choice under Risk: A Rational Expectations Approach", mimeo.

Cohen, Y. (1987), "Commuter Welfare under Peak-Period Congestion Tolls: Who Gains and Who Loses?", International Journal of Transport Economics 14, 238-266.

Cooper, R.B. (1981), Introduction to Queueing Theory 2nd ed. New York: North-Holland.

Cosslett, S. (1977), "The Trip Timing Decision for Travel to Work by Automobile" in D. McFadden et al. (eds.), Demand Model Estimation and Validation, Vol V. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley.

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

de Palma, A. and R. Arnott (1990), "The Temporal use of a Telephone Line", Information Economics and Policy 4, 155-174.

de Palma, A., M. Ben-Akiva, C. Lefèvre and N. Litinas (1983), "Stochastic Equilibrium Model of Peak Period Traffic Congestion", Transportation Science 17, 430-453.

de Palma, A., C. Lefèvre and M. Ben-Akiva (1987) "A Dynamic Model of Peak Period Traffic Flows and Delays in a Corridor", International Journal of Computers and Mathematics with Applications 14, 201-223.

De Vany, A. (1976), "Uncertainty, Waiting Time, and Capacity Utilization: A Stochastic Theory of Product Quality", Journal of Political Economy 84(3), 1976, 523-541.

De Vany, A. and T. Saving (1977), "Product Quality, Uncertainty, and Regulation: The Trucking Industry", American Economic Review 67, 583-594.

De Vany, A. and T. Saving (1980), "Competition and Highway Pricing for Stochastic Traffic", Journal of Business 53 (January), 45-60.

Douglas, G. and J. Miller (1974), Economic Regulation of Domestic Air Transport: Theory and Policy, Washington: Brookings.

d'Ouville, E.L. and J.F. McDonald (1990), "Effects of Demand Uncertainty on Optimal Capacity and Congestion Tolls for Urban Highways", Journal of Urban Economics 28, 63-70.

Drèze, J. (1987), "A Paradox in Information Theory", Chap. 4 in Essays on Economic Decisions under Uncertainty, Cambridge University Press.

Fargier, P.H. (1983), "Effects of the Choice of Departure Time on Road Traffic Congestion: Theoretical Approach", in Proceedings of the Eighth International Symposium on Transportation and Traffic Theory, University of Toronto Press, 223-263.

Frantzeskakas, J. (1982), "Capacity Analysis and Divergency of Off-street Parking facilities", Transportation Quarterly 36, 21-43.

Gal-Or, C. (1988), "The Advantages of Imprecise Information", Rand Journal 19(2), 266-275.

Gardes, Y. and A.D. May (1990), "Traffic Modeling to Evaluate Potential Benefits of Advanced Traffic Management and In-Vehicle Information Systems in a Freeway/Arterial Corridor", ITS, U.C. Berkeley, PATH Research report UCB-ITS-PRR-90-5.

Hall, F.L. and L.M. Hall (1990), "Capacity and Speed Flow Analysis of the QEW in Ontario", paper presented at the 69th Annual Meeting of the Transportation Research Board, Washington, D.C.

Haltiwanger, J. and M. Waldman (1985), "Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity", American Economic Review 75, 326-340.

Hamerslag, R. and E.C. van Berkum (1991), "Effectiveness of Information Systems in Networks with and Without Congestion", paper presented at the 70th Annual Meeting of the TRB, Washington, D.C.

Hardy, G.H., J.E. Littlewood and G. Polya (1934), Inequalities, Cambridge.

Henderson, J.V. (1977), Economic Theory and the Cities, Chap. 8, Academic Press, New York.

Henderson, J.V. (1981), "The Economics of Staggered Work Hours", Journal of Urban Economics 9, 349-364.

Hendrickson, C. and G. Kocur (1981) "Schedule Delay and Departure Time Decisions in a Deterministic Model", Transportation Science 15(1), 62-77.

Hendrickson, C. and E. Plank (1984), "The Flexibility of Departure Times for Work Trips", Transportation Research 18A(1), 25-36.

Highway Research Board (1985) Highway Capacity Manual (National Research Council, U.S.: Washington, D.C. 1960).

Hirshleifer, J. (1971), "The Private and Social Value of Information and the Reward to Inventive Activity", American Economic Review 61, 561-574.

Hirshleifer, J. and J.G. Riley (1971), "The Analytics of Uncertainty and Information - An Expository Survey", Journal of Economic Literature 42, 1375-1421.

Hoffmann, G. (1991), "Up-to-the-minute Information as We Drive - How it can Help Road Users and Traffic Management", Transport Reviews 11(1), 41-61.

Hurdle, V. (1981), "Equilibrium Flows on Urban Freeways", Transportation Science, 15, 255-293.

Hurdle, V.F. and P.K. Datta (1983), "Speeds and Flows on an Urban Freeway: Some Measurements and a Hypothesis", Transportation Research Record, 905, 127-137.

Hurdle, V.F. and D.Y. Solomon (1986), "Service Functions for Urban Freeways: An Empirical Study", Transportation Science 20, 153-163.

Institute of Transportation Engineers (1982), Transportation and Traffic Engineering Handbook 2nd ed., Englewood Cliffs: Prentice-Hall.

Ju, R-S, A.R. Cook and T.H. Maze (1987), "Techniques for Managing Freeway Traffic Congestion", Transportation Quarterly 41(4), 519-537.

Kay, J. (1979), "Uncertainty, Congestion and Public Utility Pricing", Review of Economic Studies 46, 601-611.

Koutsopoulos, H.N. and T. Lotan (1989), "Effectiveness of Motorist Information Systems in Roadway Traffic Congestion", Department of Civil Engineering, MIT. Paper presented at the IEEE Vehicle Navigation and Information Systems Conference, Toronto.

Kraus, M. (1982), "Highway Pricing and Capacity Choice under Uncertain Demand", Journal of Urban Economics 12, 122-128.

Lindley, J.A. (1987), "Urban Freeway Congestion: Quantification of the Problem and Effectiveness of Potential Solutions", ITE Journal 57(1), 27-32.

Mahmassani, H. and R. Herman (1984), "Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials", Transportation Science, 18, 362-384.

Mahmassani, H. and R. Jayakrishnan (1991), "System Performance and User Response under Real-time Information in a Congested Traffic Corridor", Transportation Research 25A(5), 293-308.

Marschak, J. (1954), "Towards an Economic Theory of Organization and Information", in R.M. Thrall, C.H. Coombs and R.L. Davis (eds), Decision Processes, New York: Wiley.

Marschak, J. and K. Miyasawa (1968), "Economic Comparability of Information Systems", International Economic Review 9(2), 137-174.

McConnell, K.E. and J.G. Sutinen (1984) "A Conceptual Analysis of Congested Recreation Sites", in V.K. Smith and A.D. White (eds.) Advances in Applied Micro-Economics Vol.3, Greenwich: JAI Press, 9-36.

McFadden, D. (1974), "The Measurement of Urban Travel Demand", Journal of Public Economics 3, 303-328.

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

Moore, A.J., P.P. Jovanis and F.S. Koppelman (1984), "Modeling the Choice of Work Schedule with Flexible Work Hours", Transportation Science 18(2), 141-164.

Naor, P. (1969), "The Regulation of Queue Size by Levying Tolls", Econometrica 37(1), 15-23.

Nelson, R.R. and S.G. Winter (Jr.) (1964), "A Case Study in the Economics of Information and Coordination: The Weather Forecasting System", Quarterly Journal of Economics 78, 420-441.

Newbery, D.M.G. (1990), "Pricing and Congestion: Economic Principles Relevant to Pricing Roads", Oxford Review of Economic Policy 6(2), 22-38.

Newell, G. (1987) "The Morning Commute for Nonidentical Travelers", Transportation Science 21(2), 74-88.

Nims, J.F. (1981), The Harper Anthology of Poetry, New York: Harper and Row.

OECD (1988), "Road Transport Research: Route Guidance and In-car Communication

Systems", Paris.

Panzar, J. (1976), "A Neoclassical Approach to Peak Load Pricing", Bell Journal of Economics 7, 521-530.

Panzar, J. (1979), "Equilibrium and Welfare in Unregulated Airline Market", American Economic Review (Papers and Proceedings), 69, 92-95.

Pucher, J. and J. Rothenberg (1976), "Pricing in Urban Transportation: A Survey of Empirical Evidence on the Elasticity of Travel Demand", mimeo.

Radner, R. and J.E. Stiglitz (1984), "A Nonconcavity in the Value of Information", in M. Boyer and R.E. Kihlstrom (eds.), Bayesian Models in Economic Theory, Amsterdam: North-Holland, 33-52.

Rothschild, M. and J.E. Stiglitz (1970), "Increasing Risk: I: A Definition", Journal of Economic Theory 2, 225-243.

Sandler, T. and J. Tschirhart (1980), "The Theory of Clubs: A Survey", Journal of Economic Literature 43(4), 1481-1521.

Saving, T.R. and A.S. De Vany (1981), "Uncertain Markets, Reliability and Peak-Load Pricing", Southern Economic Journal 47(4), 908-923.

Shechter, M. and R.C. Lucas (1978), Simulation of Recreational Use for Park and Wilderness Management, Baltimore: Johns Hopkins University Press for Resources for the Future.

Small, K.A. (1982), "The Scheduling of Consumer Activities: Work Trips", American Economic Review 72(3), 467-479.

Small, K.A. (1983), "The Incidence of Congestion Tolls on Urban Highways", Journal of Urban Economics 13, 90-111.

Smith, W.S. (1983), "Automobile Parking Trends", Transportation Quarterly 37, 431-452.

Smith, V.K. and J.V. Krutilla (1976), Structure and Properties of a Wilderness Travel Simulator: An Application to the Spanish Peaks Area, Baltimore: Johns Hopkins University Press for Resources for the Future.

Steiner, P. (1957), "Peak Loads and Efficient Pricing", Quarterly Journal of Economics 71, 585-610.

Sullivan, E.C. and S. Wong (1989), "Development of a Dynamic Equilibrium Assignment Procedure for Network-Level Analysis of New Technology", ITS, U.C. Berkeley, PATH Research Report UCB-ITS-PRR-89-4.

Transportation Research Board (1985), Special Report 209: Highway Capacity Manual TRB, National Research Council, Washington, D.C.

Tsuji, H., R. Takahashi, H. Kawashima and Y. Yamamoto (1985), "A Stochastic Approach for Estimating the Effectiveness of a Route Guidance Systems and its Related Parameters", Transportation Science 19(4), 333-351.

Vickrey, W.S. (1969), "Congestion Theory and Transport Investment", American Economic Review (Papers and Proceedings), 59, 251-261.

Visscher, M. (1973), "Welfare-Maximizing Price and Output with Stochastic Demand", American Economic Review 63, 224-229.

Walsh, R.G. (1986), Recreation Economic Decisions: Comparing Benefits and Costs, State College, PA: Venture Pub. Inc.

Walters, A.A. (1987), "Congestion", in The New Palgrave: A Dictionary of Economics, New York: Macmillan, Vol. 1, 570-573.

Williamson, O. (1966), "Peak Load Pricing and Optimal Capacity under Indivisibility Constraints", American Economic Review 56, 810-827.

Wilson, R. (1989), "Efficient and Competitive Rationing", Econometrica 57(1), 1-40.