OPTIMAL AIR FARES AND FLIGHT FREQUENCY AND MARKET RESULTS

By Christopher C. Findlay*

INTRODUCTION

The problem examined in this paper is the optimal pricing of, and investment in, air transport services. Air transport is regarded here as a jointly consumed but congestible service. The extent of congestion depends on the size of the facility, in other words capacity, and on the number of passenger trips. These elements of congestion can be summarised in a congestion, or delay, function; the specification of this delay function is described in detail below. The extent of delay can be regarded as an index of service quality, and consumer welfare will depend on this index as well as on the money fare. This view of air transport readily applies to scheduled services. Once-only charter flights could be thought of as a joint good, the joint products of which are trips in two directions and joint use of the flight capacity.

A similar view of transport services has been taken by other writers, and the problem of optimal pricing and investment\(^1\) in these conditions has received much attention. For example, road construction has been examined by Strotz (1965), Vickrey (1969), Dafermos and Sparrow (1971) and Keeler and Small (1977). Optimal bus fares and frequencies have been derived by Mohring (1972). Optimal taxi fares and service quality have been discussed by Douglas (1972), and optimality rules for air transport have been derived by de Neufville and Mira (1974), Forsyth and Hocking (1978) and Panzar (1979b). Douglas and Miller (1974) also discuss optimal service quality. But, as Forsyth and Hocking (1978) point out, Douglas and Miller used the appropriate optimal frequency condition at the current output but imposed a breakeven pricing rule, which need not be optimal. A common result is that passengers should be charged a fare equal to the marginal social cost of travel, and that optimal total output of the service could require non-zero profits. One of the aims of the work reported here is to review the derivation of the optimality rules for air transport and the source of the non-zero profit condition.

Different methods have been used to derive air transport optimality conditions. For example, Forsyth and Hocking (1978) and de Neufville and Mira (1974) use a consumer surplus type of maximand, in which consumer benefits are assumed to depend on trips and flight frequency. Panzar (1979b) uses a similar maximand which surplus is defined as the area below the demand curve and above the full price.

---

*Economics Department, Research School of Pacific Studies, Australian National University. I thank Peter Forsyth and the late Robin Hocking for bringing this topic to my attention. Also I thank E. Sierper, P. Swan and N. Vossen for helpful comments, and seminar participants for comments. An earlier draft of this paper was presented at the ANZAAE Jubilee Congress, Adelaide, May 1980.

\(^1\) There is also a large literature on optimal pricing of a given facility, e.g., Walters (1961).
of travel, which includes delay. A different approach is that of Strotz (1965), who
defines consumer utility as a function of trips, all other goods and trip quality, indexed
by delay, and then derives a Pareto optimal solution by maximising one person’s
utility subject to constraints on utilities of other people and resources. This method,
called here the Strotz parable, has also been used by Oakland (1972) and Kay
(1979). Its value is that it offers more information on the source of the non-zero profit
condition, because it makes explicit the form of the delay (or inconvenience) function.
This type of model has also been synthesised with models of optimal club size and
utilisation (Sandler and Tschirhart, 1980, 1981), and we refer again below to the
literature on clubs.

The plan of the rest of this paper is as follows. Section 2 presents the Strotz parable,
and Section 3 derives the optimality conditions for a special case of the utility
function. Section 4 contains a comparison of market results and optimality for the
special case. The results are illustrated in Section 5.

THE STROTZ PARABLE

Assume that there are I persons in the community, and the utility function of person i
is

\[ U^i = U^i(t^i, d, x^i) \quad i = I \ldots I \]  

(1)

where

- \( t^i \) = number of trips on the route of interest in the time period considered by
  person i
- \( x^i \) = consumption of all other commodities by person i
- \( d \) = average delay

and

\[ U^i_t > 0, \quad U^i_d < 0, \quad U^i_t > 0 \]
\[ U^i_t < 0, \quad U^i_d < 0, \quad U^i_{xx} < 0 \]  

(2)

where the subscripts denote partial derivatives.

Delay results when trip-making facilities are provided in lumpy units. There are two
types of delay, called frequency and stochastic delay (Douglas and Miller, 1974). The
first type occurs because passengers on average face a gap between their ideal and
available departure times. This gap is called frequency delay, and its average over all
travellers in the system increases as flight frequency decreases. The second type of
delay occurs because demand for travel at a particular time is stochastic. On some
days, a flight at a particular time will be fully booked, so passengers excluded will have
to wait for the next flight. The cost of stochastic delay depends on the gap between
flights and the probability of having to wait, which will increase with the average load
factor. Congestion of aircraft and complementary facilities can also increase waiting

---

\(^2\) The thesis by Dorman (1976), which I saw after the first draft of this paper was completed, also
uses the Strotz model to derive the optimality conditions; but Dorman’s interpretation of the results has
a different emphasis, some implications of which are noted below. He also uses a different model of
competition, and uses a technique different from that in Section 4 to compare optimality and market
results. These differences are noted below.
OPTIMAL AIR FARES AND FLIGHT FREQUENCY  
Christopher C. Findlay

times for ticket checks or baggage, and otherwise reduce the comfort of travel, and airspace near terminals will become more congested as flight frequency increases. These effects could be included in the delay function.\footnote{The case of frequency delay only is considered in detail by Forsyth and Hocking (1978), de Neufville and Mira (1974) and Panzar (1979a), and, for taxicabs, by Douglas (1972). De Neufville and Mira (1974) also consider the case where increasing frequency adds to delays through congestion of air space. Forsyth and Hocking (1978) also consider the case of both frequency and stochastic delay.}

The total number of passenger trips ($T$) and total output of other commodities ($X$) are, respectively:

$$T = \sum_{i} t_i$$

and

$$X = \sum_{i} x_i$$

The cost of providing $F$ flights is $C(F)$; so, if $R$ is the aggregate resource per period available to the community, a constraint is

$$X = X[R - C(F)] X' > O, X'' < O,$$

Where $X'$ is the first and $X''$ the second derivative of $X$.

It is assumed that flights are provided as lumpy units of capacity at a constant marginal, equal to average, cost, $k$.

The next step is to specify the determinants of average delay, $d$. Here it is assumed that

$$d = D(T,F)$$

where $D_r > O, D_F < O, D_{rr} > O, D_{FF} > O$ and $D_{RF} \leq O$. $F$ is included because of the presence of frequency delay, and the inclusion of $T$ implies there is stochastic delay as well. If units of capacity were perfectly divisible there would be no frequency on stochastic delay, though there could still be congestion of complementary facilities.

It might be argued that a relevant constraint is seating capacity on each flight. It is assumed that this is not binding on average, because, with full aircraft, stochastic delay costs will become very high. In other words, full loads on all flights cannot be efficient because they imply very high waiting costs.

The problem is to find a price ratio and a flight frequency that, in these conditions, will yield a Pareto optimum. The method of deriving the necessary conditions for an optimum is to fix utility levels of all but one person and maximise that person’s utility subject to constraints on utilities of other people, resources, and the nature of the delay function.

Let $c^j$ be the fixed utility level of person $j$ and form the Lagrangean ($L$).

$$L = \sum_{i} w^i [U_i(t_i, d, x_i) - c^i] + \alpha [d - D(T,F)] + \beta [X - X(R - kF)]$$

Assume $w^i = 1$ and $c^i = O$; then the other $w^i$ can be thought of as Lagrangean multipliers.
The necessary conditions (which are taken to be sufficient) for a Pareto optimum are the following:
\begin{align}
  w^i U^i_t - a D_T &= O, \quad i = 1 \ldots I \quad (8) \\
  \sum w^i U^i_d + a &= O \quad (9) \\
  w^i U^i_x + \beta &= O, \quad i = 1 \ldots I \quad (10) \\
  \beta x^i k - a D_F &= O \quad (11) \\
  d - D(T,F) &= O \quad (12) \\
  X - X(R - kF) &= O \quad (13) \\
  U^i(t^i, d, x^i) - c^i &= O, \quad i = 1 \ldots I - 1 \quad (14)
\end{align}

There are $3I + 3$ equations, equal to the number of unknowns $(t^i, x^i, a, \beta, d, F, w^i)$ in this problem; the number of individuals who travel is taken as fixed.\(^4\)

This set of conditions can be rearranged to indicate the characteristics of the optimum. First note that $w^i U^i_x$ is equal for all $i$. Then from (8), (9), (10) and (11)

\begin{align}
  \frac{U^i_t}{U^i_x} &= - \sum \frac{U^i_d}{U^i_x} D_T \quad (15) \\
  \beta x^i k &= \sum \frac{U^i_d}{U^i_x} D_F \quad (16)
\end{align}

Equations (15) and (16) are similar to the optimality conditions for joint goods. The joint goods, which enter all travellers' utility functions in equal amounts, are flight frequency ($F$) and the utilisation of existing flight facilities ($T$). Equation (15) says that $t^i$ should be increased, raising $T$, until the marginal benefit to the traveller equals the marginal costs imposed on other travellers. Equation (16) says that frequency should be increased until its marginal cost equals the sum of the marginal evaluations of reductions in delay.

In the Strotz model the facility is shared by people of differing tastes. In the clubs literature this would be called a "mixed club". It is plausible that forcing people of differing tastes to share the same facility could be sub-optimal; recently, Berglas and Pines (1981) claim to have shown that mixed transport clubs are not optimal. In a comment on this result, however, Sandler and Tschirhart (1981) show that it follows from a self-financing constraint imposed by Berglas and Pines. The self-financing constraint would be appropriate when "the club" is characterised by constant costs of capacity and a delay function homogeneous of degree zero in trips and flight frequency (see below), that is, by constant returns to scale. The intuitive explanation of the non-optimality of mixed clubs is then clear; under constant returns to scale it is

\(^4\) In the literature on clubs, the optimisation would have included $I$ as well; that is, club membership would have been a variable. Here it is assumed that the population is the membership.
OPTIMAL AIR FARES AND FLIGHT FREQUENCY

Christopher C. Findlay

possible to have a separate "club" to cater precisely for each person's tastes. Although
the assumption of constant returns is adopted below, it is not assumed that this
assumption is relevant over the whole range of possible output. Sandler and Tschirhart
also show that, in the type of problem considered here, mixing is optimal. Differing
tastes for service quality are accounted for by differing utilisation rates and total toll
payments. In terms of the first order conditions, the mixed club is optimal because the
evaluation of a unit of congestion is reflected in both the toll condition and the
capacity condition.

Combining (15) and (16) leads to (17):

\[
\frac{U_i^t}{U_x^t} = -X't^k \frac{D_T}{D_F}
\]  

(17)

The right-hand side of (17) is the same for all travellers, so the marginal rate of
substitution between trips and other goods will be equal for all. So a situation where
each person faces the same set of relative prices can lead to an optimal result. The next
step is to determine the optimum relative prices. The approach is to consider the
problem facing each person. Let

\[
P = \text{fare} \\
V = \text{price of other goods} \\
R^t = \text{income from sales of resources (the price of which is unity)} \\
O^t = \text{profits earned} \\
H^t = \text{lump sum tax paid}
\]

Then the individual's constraint is

\[
R^t + O^t - H^t = Pt^t + Vx^t
\]  

(18)

The individual's problem is to maximise

\[
U^t = U^t(t^t, d, x^t)
\]

subject to (18), using \( t^t \) and \( x^t \); an individual regards \( d \) as fixed. At the individual's optimum

\[
\frac{U_i^t}{U_x^t} = \frac{P}{V}
\]  

(19)

Combining (19) and (17), the optimal price ratio is

\[
\frac{P}{V} = -X't^k \frac{D_T}{D_F}
\]  

(20)

The next step is to determine the optimal number of flights. The average delay
function was defined in (6); now let\(^5\)

\[
a = D_T T + D_F F
\]  

(21)

\(^5\) \( a \) is not equal to the degree of homogeneity of delay function, since \( a = d.k \) where \( h \) is the degree of homogeneity.
If $D$ is homogeneous of degree zero, then $a = 0$, but generally it is expected that $a$ is non-positive. The reason is as follows (Mohring, 1972).

Ignoring airspace congestion, consider the time a traveller spends waiting for a flight. Assume that expected waiting time is equal to half the time between flight departures, which are evenly spaced over the day. Then suppose $T$ and $F$ are doubled. Times between departures will then be halved, so average delay will fall. As a result, $a < 0$; Strotz calls this a case of "favourable" returns to scale (Strotz, 1965, p. 135).

Expenditure on flights is assumed to equal sales revenue plus tax revenue ($H^i$ by person $i$).

$$kF = \sum H^i + P \sum i$$

(22)

The question is what is the appropriate total tax revenue. Assume that the other goods sector is competitive; so, assuming resources to be the numeraire,

$X'V = 1$ or $X' = 1/V$

and from (20)

$$P = -kD_F/D_F$$

(23)

To interpret this expression for price, fix delay. Then

$$\left(\frac{dT}{dF}\right)_d = -D_F/D_F$$

so

$$P = k\left(\frac{dT}{dF}\right)_d = -\bar{d}$$

(24)

which is the marginal cost of an extra trip, if service quality is held constant. Then from (21)

$$kF = (ka/D_p) - kT(D_F/D_p)$$

$$= (ka/D_p) + PT$$

(25)

Since $D_F < 0$, if $a < 0$, flight expenditures will exceed sales receipts at the optimum, even though cost per flight ($k$) is constant.

The results can be interpreted by considering some special cases. For instance if delay depended only on $F$ and not on $T$, then $F$ could be thought of as a public good and the optimality condition would be (16). Revenue equal to the cost of the marginal flight could be raised by taxing each consumer an amount equal to his marginal evaluation of an extra flight. The trip fare should be equal to the marginal cost of a passenger, which in this case is zero. The outcome can be thought of as a travel club, members of which pay a varying joining fee but travel on the route at no extra cost.

The other special case arises where flights do not enter the delay function but delays increase with $T$. This is a case of a consumption external diseconomy: (15) applies and a congestion tax should be levied on travellers. This case could be thought of as a community of people; the greater the number who fly, the more congested flight paths become.

The model presented above is a combination of these two cases (as noted by Forsyth and Hocking (1978, p. 13)). Though revenue is raised from a congestion tax, it may not be sufficient to cover the costs of the optimum number of flights.
OPTIMAL AIR FARES AND FLIGHT FREQUENCY

Christopher C. Findlay

A SPECIAL UTILITY FUNCTION

Equation (1) defined a general utility function. A special case (suggested by Kay, 1979) can be derived by assuming $D(T,F)$ is a unit cost function which has the same value for everybody. The utility function can then be written

$$U^i = U[i, x_i - tD(T,F)]$$

(1b)

Individuals in this specification are not identical, because their preferences for trips can vary; but, when a trip is made, they each face the same average delay cost, in terms of good $x$. In that case, the individual's and, subsequently, the market demand for trips can be written as a function of the "full fare" $(P/V + D)$. Substituting from (1b) into (15) and (16), the optimal toll $(P/D)$ in this case is $TD_F$, which is the marginal social cost of the extra trip, and when frequency is optimal

$$X'k = -TD_F$$

(26)

so the marginal cost of an extra flight must equal the marginal gain. Total revenue will be

$$T(P/V) = T^2 D_F$$

and total cost will be

$$Fk/V = -TFD_F$$

Thus the subsidy required will be, in units of the private good,

$$-T(FD_F + TD_F) = -Ta$$

or, since $T = -kX'/D_F$, the subsidy equals $ka/D_F$ (in units of the numeraire); this was the result from Section 2. If $a = 0$, no subsidy is required.

In a model which emphasised the location of flights in time (analogous to the location of ice-cream sellers around a lake) Panzar (1979b) derived the conditions for zero profits at the optimum. The Panzar conditions were that the value of frequency delay is zero and the direct⁴ marginal effect of a change in frequency on stochastic delay (through a reduction in the gap between flights) is zero. There is still delay due to congestion in these conditions, but the optimal toll is such that tax revenue covers the cost of providing flights. This corresponds to the case where delay depends only on the load factor of each flight, which implies that the delay function is homogeneous of degree zero $(a = 0)$. Hence the Panzar condition and that derived here, for zero profits at the optimum, are equivalent⁷.

Telser's (1969, 1971) frequency-delay-only model is a special case of Panzar's (1979b) model. The case of frequency delay only corresponds to the characteristics of

---

⁴ The delay function $D(T,F)$ can also be written $E(T,F,F)$; then $E_F = E_1/F$ and $E_x = E_u - TE_x/F$, so $a = TE_x + FE_x = FE_x$ where $E_x$ is the direct effect of frequency on delay (the indirect effect is via the load factor), so $a = 0$ when $E_x = 0$. In that case delay depends only on $T/F$.

⁷ At this point it is appropriate to summarise some differences between Dorman's (1976) analysis and that used here. Dorman does not emphasise the separate optimal toll and frequency conditions, does not offer the congestion toll interpretation of the optimal toll condition, and does not consider the self-financing characteristics of optimality. Dorman does derive the same general optimality rule (equation (20) above), but in discussing marginal cost pricing he uses an equation like (24).
spatial markets, a point evident in Telser’s (1969) model of this problem. But Demsetz (1971) pointed out that Telser (1969) had ignored differential land rent, arising because of differing distances of consumers from the point of purchase (a factory in the Telser model). Subsequently, Southey (1974) showed that the frequency of factories would be optimal when their owners could capture these differential rents. The application of these results to air transport is limited. It is difficult to imagine how airlines would capture the surpluses of consumers located close to flight departure time. Southey’s optimality result applies only in the case limited to frequency delay; so, when frequency also has a direct effect on stochastic delay, frequency will still be too low.

MARKET RESULTS

The next step is to compare the optimal results with those generated by types of markets. Assume initially that there is just one firm providing all flights, and take capacity as fixed; these assumptions will later be relaxed. The firm is a profit maximiser and its problem is to choose a price \( P \).\(^8\) The pricing rule is to set marginal revenue to zero, as shown in (27):

\[
\frac{dT_P}{dP} = T + P \frac{dT}{dP} = 0
\]

(27)

An expression for \( P \) can be obtained from (27) by noting (Kay, 1979):

\[
\frac{dT}{dP} = \left( \frac{\partial T}{\partial P} \right)_D + \left( \frac{\partial T}{\partial D} \right)_D \frac{dT}{dP}
\]

Then, by multiplying both sides by \( P/T \) and using (27), it can be shown that

\[
P = \frac{-TD_T (\partial P/\partial D)_T}{1 - (1/e)}
\]

where

\[
e = - \left( \frac{\partial T}{\partial P} \right)_D \frac{P}{T} > 0
\]

and

\[
\left( \frac{\partial P}{\partial D} \right)_T = - \left( \frac{\partial T}{\partial D} \right)_P D \left( \frac{\partial T}{\partial P} \right)_D
\]

\(^8\) This price is actually relative to the price of the private good; but, in this section and the next, the private good is taken as the numeraire. This means \( k \) is redefined as the cost of the extra flight in terms of the new numeraire.
OPTIMAL AIR FARES AND FLIGHT FREQUENCY

Christopher C. Findlay

In the special case, the demand of an individual can be written

$$t^i = t^i(P + D)$$

and then for all $i$

$$\left( \frac{\partial P}{\partial D} \right)_T = -1$$

Therefore

$$P = TD_T/[1 - (1/e)]$$ (28)

and this price exceeds $TD_T$.

The pricing rule (28) shows that the relationship between the price and the optimal toll ($TD_T$) depends on $e$, which is the price elasticity of demand for units of constant quality. The higher $e$ is, the closer the price will be to the optimal toll.

The elasticity ($e$) will vary with the structure of the market. In competition, where there is free entry, $e$ will be very high. For example, if the firm tried to raise price without changing the level of delay, all passengers would be worse off and would immediately switch to an entrant offering the old mix of price and quality.

In this model, the individual has the same reaction to a rise in the money price, given $d$, as to a rise in $d$, given $P$. Thus a high value of ($e$), the response to a change in the money price, also implies a high response to a change in $d$, given the money price. These results imply that in competition the firm faces a very elastic full price ($P + d$) demand curve at the going full price. In other words, the competitive firm must take the full price as given for the optimal money toll to be generated.

The model of competition discussed so far has referred to the behaviour of a single firm supplying all flights, but decentralised decision makers can also reach the optimal toll and utilisation. Let each flight correspond to one firm. Each firm takes the full price of travel as given. No firm has any locational advantage over another, so total traffic, which is fixed since the full price is fixed, will be divided equally between all firms. Each firm acts in the belief that it can choose its own combination of money fare and passengers, with the aim of maximising revenue, as long as it takes the full price as given. It can be shown that the money fare will then equal the optimal toll\(^9\) when the target load factor is achieved. If there are excess seats, the full price will fall and competitive firms will readjust their money fares.

There are other models of competition in the literature. Some (for example De Vany (1975) and Douglas and Miller (1974, ch. 4)) refer to regulation where the money fare is fixed. Panzar (1979b) develops a model where firms take their rivals’ money fare as fixed (and shows that the competitive result will not be optimal). In contrast, in the model used here, firms take the full price of travel as given. The

\(^9\) The firm’s problem is to choose $P$ and $t$ (average passengers per flight) to maximise $Pt$ subject to $P + E = C$ where $C$ is a constant and $P + E$ is the full price. The Lagrangean ($L$) and first order conditions are

$$L = Pt + a(P + E - C) \quad t + a = 0 \quad P + aE_t = 0$$

from which $P = tE_t$ and, since $t = T/F$ and $D_T = E_t/F$, then $P = TD_T$. 

57
problem is now to explain why this should occur in competition. One answer is that in this model demand depends on the full price, so the analogy with competitive behaviour in markets where there is no variation in quality is that competitive firms here should take the full price as given.

More support is available for this view of the nature of competition in airline markets. The following paragraphs will show that, when firms take their rivals’ money price as given, zero profits may be the result, but that this outcome is unlikely to be an equilibrium. This result suggests that the model in which money price is taken as given is not a useful description of competitive behaviour.

Consider the full price of travel

\[ P^* = P + d(t) \]

where \( P \) is the money fare and \( t \) is passengers per flight. Each flight corresponds to a different firm. Full price contours can be drawn in \( (P,t) \) space; their slope will be negative and will become steeper as \( t \) increases.\(^{10}\) Some contours are illustrated in Figure 1.

Assume initially that all firms are charging the same money price \( P_1 \), so total traffic is equally divided between all flights (\( t_1 \) each). The full price is \( P^* \). The next step is to add a rectangular hyperbola as an iso-revenue curve \( (R = Pt) \). When firms take the full price as given they will choose \( P \) and \( t \) to maximise revenue subject to that constraint. In Figure 1 the solution to this problem will be to move to point \( B \). Suppose that revenue corresponding to contour marked \( R_4 \) equals cost per flight; then at \( B \) profits will be positive. Entry will now occur, lowering the full price, so that optimal fares have to be recomputed by each firm. This process continues until zero profit equilibrium is reached at point \( C \).\(^{11}\)

This process can be compared with that in which firms take their rivals’ money fares as given. For example, let the starting point be \( A \); then one firm considers cutting its money fare. It assumes other firms will hold their money fares constant at \( P_1 \), but it knows that after all adjustments the full price of travel must be equal on all flights. The implication is that the cutter will perceive his set of opportunities to lie along a line like \( AD \). This line is steeper than the slope of the full price contour. The reason is that, as the price cutter lowers the money fare and attracts new customers, loads on rival flights will fall, so lowering their full prices and reducing the attractiveness of switching to the cutter. The price cutter will choose a point on this line where revenue is maximised (point \( D \)). He will expect other firms to be shifted to \( E \), as the full price is equal at \( E \) and \( D \). Since the opportunity line \( AD \) is steeper near \( A \) than the full price contour, it is expected that, given the shape of the iso-revenue contours, the solution \( D \) will be at higher money fare than \( B \).\(^{12}\)

\[ \frac{dP}{dt} \bigg|_{P^*} = -d' < 0 \]
\[ \frac{d^2P}{dt^2} \bigg|_{P^*} = -d'' < 0 \]

\(^{10}\) For simplicity this description omits one part of the adjustment process. In the shift from \( A \) to \( B \) firms’ expectations may not be fulfilled and the full fare may fall, leading to another adjustment in money fares even before entry occurs. The question whether entry leads to optimal frequency is examined below.

\(^{11}\) This result is more formally derived by Panzar (1979b).
In competitive markets all firms adopt the same strategy and try to move to $D$. If their expectations are not realised they recompute their money fares. The characteristic of the solution will be tangency of an iso-revenue contour and an opportunity line like $AD$. If profits exist at this outcome entry will occur, and the final solution will lie on the zero profit iso-revenue line. But, because of the use of lines like $AD$, this zero profit solution need not lie at point $C$. In other words, only by chance will the second competitive mechanism minimise the full cost of travel, given zero profit. In that case the final solution will not be an equilibrium; all firms could be undercut by another offering a lower full price, and the new competitor can still earn non-negative profit.
The conclusion that competitive utilisation of capacity is optimal is an old one (see Knight, 1924). An explanation is as follows. Suppose that access to the air transport facility is free: then utilisation will expand to the point where the last passenger is indifferent whether he makes a trip or not. His decision will be based on the average delay which is faced by the individual traveller. But this solution is inefficient, because a traveller can be removed from the system, and that traveller will be no worse off but congestion will be reduced, so other travellers will be ready to pay a premium to remain. The premium is a measure of the gain. Now take away another traveller and raise the charge to the remainder. This change will be an improvement as long as total revenue is higher. Thus passengers should be removed as long as the aggregate premium paid by the remainder goes on increasing. The optimum occurs at the point where revenue is maximised, which is precisely the competitive solution. Effectively, private ownership of the facility causes the congestion externality to be internalised; or, as Kay (1979, p. 604) says, “the profit maximising position would be identical if congestion costs were borne by the supplier rather than by consumers”.

Utilisation of existing capacity will be optimal in competition; but profits may be earned, which will attract entry of new firms, so capacity will expand. As capacity expands, the pricing rule will be adjusted so that utilisation continues to be optimal. Entry continues until profits are zero, that is,

\[ PT(P,D(T,F)) - kF = 0 \]  

(29)

Substituting for \( P \) in (29) it can be shown, using the definition of \( a \) (from (21));

\[ k = aT/F - TD_F \]  

(30)

Hence, if \( a = 0 \), competitive frequency will be optimal (see (26)). When \( a \) is negative \( k \) will be less than \( -TD_F \) and, assuming that \( D_{ve} \) is positive so that \( -TD_F \) decreases with \( F \), competitive \( F \) is less than optimal. The important result is that the coincidence of optimality and competition depends on the characteristics of the delay function.

The monopolist’s choice of \( P \) and \( F \) will be characterised by the following conditions:

\[ dTP/dP = 0 \]

\[ PT_F - k = 0 \]

The monopolist’s price will be (28) above, and, since entry restriction implies that the value of \( e \) facing the monopolist will be low, it exceeds the optimal toll and output will be too low. But it can be shown that, from the condition for profit maximising frequency,

\[ k = -TD_F \]

so that, given \( P \), the monopolist chooses the optimal frequency, whatever the value of \( a \). These profit-maximising conditions have been used by De Vany (1975) and by Olson and Trapani (1981) to derive loci of price and trips and of price and capacity, which can then be used to predict the effects of regulatory rules which have been assumed to apply to US domestic civil aviation.

In summary, air transport is provided in lumpy units (flights) which can be thought of as jointly shared but congestible facilities. The optimal money fare is therefore a congestion toll. Increased capacity creates benefits for all travellers by reducing
OPTIMAL AIR FARES AND FLIGHT FREQUENCY

Christopher C. Findlay

congestion and, in the general case, revenue raised by the congestion tax may not be sufficient to cover the cost of optimal capacity, so a subsidy may be required. In a special case where the degree of homogeneity of the delay function is zero, no subsidy is required.

When all consumers have identical evaluations of delay, competitive firms in air transport markets were shown to charge the optimal toll, given capacity. Competitive frequency will be optimal when the conditions for a zero profit optimum are met. This condition can be expressed in three equivalent statements:

1. the delay function is homogeneous of degree zero in frequency and total trips;
2. delay depends only on load factors;
3. there is no direct effect of changes in frequency on stochastic delay, and frequency delay is zero.

THE STROTZ PARABLE ILLUSTRATED

Let $Q$ be passengers per flight. Then, in the case where only one flight is available (so $Q = T$) total cost ($C$) will be

$$ C = k + dQ $$

Average cost ($A$) and marginal cost ($M$) are then

$$ A = \frac{k}{Q} + d $$

$$ M = d + QD_T $$

These functions are depicted in Figure 2, which also shows average variable cost ($V$), which is equal to $d$.

The relationship between these functions is as follows. $M$ cuts the $S$ axis at $d_{r}$, which is equal to frequency delay imposed by the existing number of flights. As $Q$ rises, congestion increases, so $d$ rises, as is shown by the rising $V$ curve. The gap between the $M$ and $V$ curves is $QD_{r}$, the marginal cost imposed by an extra traveller. The slope of $A$ is given by $-k/Q^2 + D_T$, which is zero when $QD_{r} = k/Q$, that is, when $M$ cuts $A$ at $A$'s minimum. At that point ($C$), the distance $BD$ is equal to $d$ and $BC$ is equal to both $k/Q$ and $QD_{r}$.

The next step is to derive the long-run average cost function. This envelope can be derived by solving the problem

$$ \min kF = Td $$

where $T = T$

and in that case

$$ k = -TD_{F} $$

The envelope curve ($LA$) and the corresponding marginal curve ($LM$) are then

$$ LA = d - FD_{F} $$

$$ LM = d + TD_{F} $$
and

$$LA - LM = -(FD_r + TD_r)$$

$$= -a$$

If $D$ is homogeneous of degree zero, then $a = 0$. When $a$ is negative, the $LA$ curve always lies above the $LM$ curve, which means long-run average cost falls as $T$ increases. The explanation of this result is that $LA$ is the envelope of a series of short-run curves. The short-run curve for a particular frequency is the horizontal sum of $A$ curves for a single flight (like that in Figure 2). But as $F$ increases, the curve for one flight falls because the $d_o$ intercept decreases as frequency delay decreases. When $a = 0$ this effect is stronger than the congestion effect of a larger number of travellers, and therefore the envelope curve has a downward slope.

The optimal toll is equal to the marginal cost which the extra trip imposes on other travellers. In this model, the optimal toll is therefore $TD_r$. So the $LM$ curve is equal to the average delay plus the optimal toll, and can be interpreted as the full price schedule when the optimal toll is charged.
The $LA$ and $LM$ curves plus the demand curve ($D$) are drawn in Figure 3. Demand is assumed, in this special case, to be a function of the full price of travel ($P + d$). Optimal $T$ (denoted $T^*$) occurs where $D$ cuts $LM$. Optimal $F$ corresponding to $T^*$ can be obtained from the $LA$ curve, as illustrated in Figure 3. Given $T^*$, the optimal number of flights occurs when $LM$ cuts $M$ at $E$. The optimal toll is $EG$, equal to $TD_f$, which is less than $k/Q$, equal, in Figure 3, to the difference ($HG$) between $A$ and $V$. Thus revenue from the optimal toll will not cover the resource costs of the optimal number of flights. The unit subsidy required will be $EH$, the gap between $LA$ and $LM$, which was shown above to be equal to $-a$. Hence the total subsidy is $-aT$; but, since $T$ equals $-k/D_p$, the total subsidy is $ka/D_p$, as shown in Section 3.

As $F$ and $T$ increase it may be that the degree of homogeneity approaches zero. In that case $TD_f = -FD_f$ and $LA$ becomes flat and equal to $LM$. Therefore

$$T(TD_f) = kF$$

---

13 This diagram is similar to Figure 1 in Mohring (1972). Some differences are that Mohring did not include a demand curve in his figure. Mohring discussed congestion, but it is a type different from that discussed here; in Mohring's bus system, stopping to pick up an extra passenger imposes delay costs on people already on the bus. Mohring says (in footnote 7, p. 594) that he ignores the stochastic nature of demand.
Since the left-hand side is total revenue, profits at the optimum will be zero.

The competitive case is illustrated in Figure 4. Competitive output is $T_c$, competitive price is equal to $BC$, and the full fare is $BT_c$. Optimal output is $T^*$, which exceeds competitive output and implies a higher frequency. The competitive solution minimises average cost, given capacity; in other words, competitive utilisation of the available facility is optimal. If the homogeneity of the delay function is zero, $LA$ will correspond to the locus of the minimum points of the $A$ curve, which are the competitive solutions. In this case, competitive frequency will be optimal.\(^{14}\)

**CONCLUSION**

There are two main results of this paper. Firstly, it confirmed the possibility that optimal air fares and flight frequency could produce non-zero profits. The source of this divergence was found to be the presence of two externalities: a negative congestion externality generated by the extra trip and a positive external effect generated by an extra flight. Revenue can be raised from an optimal congestion tax, but it may not be sufficient to cover the costs of the optimal number of flights.

The second result is that it is also possible that competition can provide the optimal number of flights. This occurs when the delay function is homogeneous of degree zero. This conclusion has a number of interpretations. For example, it implies frequency delay of zero, and changes in frequency have no effect on stochastic delay. In other words, it means that delay depends only on the load factor. The relevance of this case in international air transport and its implications for regulation are discussed in Findlay (1982).

The final point is that the Strotz model can be extended in a number of ways. For example, two types of travellers can be introduced into the model. Different types of travellers can be assumed to have different evaluations of delay but can rank themselves by buying different types of tickets, such as standby and normal tickets. Optimal fares and capacity under this ticket system can be derived from the model. Instead of referring to different types of ticket or of travellers, separate delay functions could refer to trips in different directions or at different times of the year. Thus, the model is also applicable to questions of peak-load and back-haul pricing.\(^{15}\)

\(^{14}\) This is another convenient point to make comparison between Dorman's (1976) results and those obtained here. Dorman uses a series of iso-revenue, iso-cost (of capacity) and iso-full price contours in $(P, d)$ space to compare the results of competition and optimality, rather than the more familiar diagrams used here. In Dorman's model of competition a traveller first chooses a firm, then one of that firm's flights, whereas here a traveller only chooses among flights. He justifies his assumption on the grounds of search costs. He derives a result analogous to the case of a declining $LA$ curve when $a$ is negative; but, with his specification of the nature of competition, he is then led to conclude this case implies a natural monopoly. Subsequently, he argues, a zero value of the parameter $a$ is a necessary condition for the market to be competitive. On the contrary, it is shown here that the market can still be competitive even though $LA$ is not flat. A zero degree of homogeneity of the delay function is actually a necessary condition for competitive frequency to be optimal. If search costs are important they could be incorporated, but as another dimension of service quality. Dorman limits his discussion of competition to the case where a single firm provides all flights.

\(^{15}\) Notes on these applications of the model are available on request.
REFERENCES


