Supplement to "forum papers" (4th Annual Meeting, Australian Transport Research Forum, 1978) - referred to on page 211.

OPTIMAL AIRLINE FARES, SERVICE QUALITY AND CONGESTION*

P.J. FORSYTH
LECTURER IN ECONOMICS
THE UNIVERSITY OF NEW SOUTH WALES

R.D. HOCKING
LECTURER IN ECONOMICS
UNIVERSITY OF MELBOURNE

ABSTRACT:

The problems of determining optimal fares, frequencies and load factors for an airline service are discussed. Some contributions of other authors to these problems are examined, and their limitations are pointed out. These contributions form a starting point for a simple, yet general, analysis. Two processes, leading to differing results, are presented. One is a form of economy of scale, and the other is a standard congestion process. The former leads to optimal prices being lower than such as would cover costs, and the latter leads to load factors being less than 100%. Problems of determining and applying price structures to maximize net benefits when these processes are present are discussed.

* Due to unforeseen circumstances, the authors were unable to make this paper available for publication prior to the forum.

OPTIMAL AIRLINE FARES, SERVICE QUALITY AND CONGESTION

INTRODUCTION

The problem of determining the most efficient airline fares and capacity levels is one that has attracted some attention of late, though as yet, no complete theoretical treatment of it has been given. Several writers in the Air Transport area, notably De Vany (1975) and Douglas and Miller (1974), have advanced the analysis considerably, especially in the empirical direction, but have stopped short of providing a general model. The problems that arise are quite similar to those in other areas which have received attention by writers such as Turvey and Mohring in their discussion of Optimal Bus Fares (1975). The results of the contributions mentioned above are suggestive of important policy conclusions, which have not been spelt out in so far as they are relevant to Air Transport Pricing.

We hope to provide a simple, but general analysis of the key considerations, and to indicate how optimal fares and capacity can be determined. While we do not provide any empirical estimates here, the paper has clear empirical implications. The complications of measuring some of the necessary parameters are indicated, and the way they affect the interpretation of existing studies is pointed out. There are difficulties in applying models such as these but they are not such as to prevent quite satisfactory approximations being made, given data availability. Furthermore, the relevance of a type of congestion to the pricing problem is indicated; its role has previously gone unrecognized.

In the following section the problems are described, as are the methods adopted by some authors to take account of them. The third section provides some simple models which are used to derive optimality rules. In the last section the implications of these results for empirical work and for practical pricing/capacity choice problems are examined, as are the implications of different price structures.

SERVICE QUALITY AND CONGESTION

Airline services must be provided in lumps of a given size, determined by the capacity of the aircraft making the flight. There is some, though not much, flexibility open to the airline in making different flights with different capacity aircraft, in order to tailor capacity supplied to demand. Passenger demand, viewed over a period, could be regarded as a continuous variable, in that it can exist for any time within the period. It is possible to provide service when each passenger desires it, but to do so would be prohibitively expensive, given economies of scale in aircraft capacity. Thus, over a period, a number of flights will be scheduled such that they satisfy passenger demand as well as possible.

The continuous nature of demand and the discrete nature of supply gives rise to the schedule problem, which has received considerable attention (e.g. see Miller, (1972)).

As more flights are scheduled, the deviation of the passengers' preferred from actual time of travel falls. Thus benefits to passengers increase with greater frequency of flights, quite apart from whether more people travel. This process is explicitly recognized and analysed by Douglas and Miller (1974) and De Vany (1975). We shall adopt the terminology of the former, who call this deviation between preferred and actual travel times "frequency delay".

A second way in which frequency may affect benefits to passengers is through what Douglas and Miller call "stochastic delay". Given the random nature of demand, both as to when a given number of passengers in a day will wish to travel, and as to the number each day wishing to fly, there will be a chance that travellers will not be able to fly on their preferred flight, but will have to accept some less preferred alternative. The difference between preferred and actual travel times is "stochastic delay". More flights relative to traffic will mean less chance of failing to obtain the preferred one, and the greater the number of flights, the smaller the interval between the actual flight and the preferred one. Thus stochastic delay falls with frequency. For this additional reason, greater frequencies will be valued positively by passengers.

Both these forms of "delay" can be measured in terms of minutes, though it must be recognized that often the "delay" involves passengers catching a flight earlier than their preferred travel time. One possible measure of the cost of delay is to regard the "delay" as being wasted time, and evaluate it using a valuation of time obtained from other sources. This is what Douglas and Miller do, though they recognize the deficiencies in the procedure. De Vany prefers to estimate the value of frequency directly, via trade offs between frequency and other variables, such as price, in empirically determined demand functions. We now look more closely at these procedures.

It is important to distinguish between rearrangement of a time pattern of activities, and wasting time in some unproductive activity (such as waiting at an airport lounge). The two types of delay, aggregated by Douglas and Miller as "schedule delay", involve elements of both these aspects. The two are inter-related, in as much as an inconvenient pattern of activities may lead to a wasting of time, as other activities cannot be rearranged to fill an unoccupied period. Douglas and Miller recognize that schedule delay is often primarily not wastage of time *per se* , and note that the interval between preferred and actual travel times can be put to some use (1974, p.85). Hence they consider that valuation of schedule delay at an implicit en route value of time for air travellers will provide an upper limit to the valuation (Douglas and Miller, 1974, p.86).

It is likely that this procedure will normally provide an upper limit, but it need not. It is possible that the costs of rearrangement of activities are very high, for some passengers, and are higher than the amount determined by the value of time. For example, if a flight leaves twenty minutes too late for convenience, it may be imperative to catch the earlier flight, which may leave five hours earlier. The schedule delay is not the difference between preferred and closest flight times, twenty minutes, but five hours. This is probably not an important qualification when frequencies are fairly low, since flights will be scheduled when most passengers find them convenient. It could be important for frequent flights. The time difference may be taken up as wasted time, and in addition the travel time may be inconvenient. For example, it may be necessary to catch an 8 a.m. flight, when an 8.30 a.m. flight is not available, and arrive half an hour before an appointment. This time may be completely wasted, and, in addition, 8 a.m. may be much more inconvenient than 8.30 a.m. as a starting time; activities prior to the flight may have to undergo costly rearrangement. For short schedule delays, which represent time which cannot be put to any good use, and some rearrangement of activities, estimation of schedule delay costs from time valuations could yield a lower, not an upper, bound.

To make further propositions about the way schedule delay cost varies, it is necessary to examine how its components vary in response to the relevant variables. As frequency increases schedule delay decreases, but it would do so at a decreasing rate - this is suggested by Douglas and Miller's results (Douglas and Miller, p.83). However, it is likely that per unit delay costs increase with increases in frequency, and decreases in length of average delays. The shorter the delay, the less it is able to be taken up in alternative activities - if it is quite short it may be simplest to regard it as wasted time. For long "delays" it is unlikely that they will be taken up as wasted time, though they could result in very inconvenient rearrangements of activities. On balance it is difficult to be certain, *a priori,* whether the value of an additional flight increases or decreases with greater numbers of flights, though as the effect on delay becomes very small, the cost increase would not compensate, and the value of an extra flight will ultimately fall.

Perhaps it is more important to determine the way schedule delay costs vary with the length of the flight. The longer the flight, the more costly in time and money it will be, and passengers will make a flight as part of a larger overall trip. The larger the trip, the greater the flexibility in rearranging activities around the travel time. For a brief, short distance trip, it may be important to reach a destination by a specific time of day, whereas for a long distance trip, even if specific appointments are made, the passenger will usually plan the trip to be longer and more flexible. This is partly because frequencies are usually lower, because the travel time is longer and in absolute terms, more variable, and because the nature of the trip is different. The success of charter operations indicates that many travellers are insensitive to schedule delay, especially when they are making a trip long in distance or duration. It is probably safe to assume that "delay" costs fall with the length of the journey. If this were not the case, the delay costs for low frequency, long distance international flights would be enormous, and would indicate substantially increased frequencies were desirable.

Important though the estimation of the value of frequency is, it is only part of the problem. The other part namely that of determining the optimal level of frequency and of passenger demand, is just as critical, yet it has attracted less attention. It is usual to solve such problems by setting up an optimising model, and using it to derive the conditions for efficiency. This task remains to be done. Several writers, such as Douglas and Miller, have indicated part solutions, but these can be misleading and inadequate.

In the Douglas and Miller model, prices are relevant only as a revenue raising device. The optimisation problem set out in this model is one of minimising schedule delay costs by altering frequency, for a fixed number of passengers. It is no surprise that their "optimality" requires that the costs of extra frequency be traded off against the costs of extra delays. Even if stochastic delays were zero, frequency delays would require a load factor of less than 100% in the optimum. Fares are simply levied so as to cover costs. The role of price affecting demand is ignored, as is the role of service quality. This is an important omission, since it rules out the possibility of determining an optimum frequency, and in the absence of stochastic delay, dropping prices so as to set load factors at 100%. If air travel demand were highly inelastic, this approach could be a good approximation. The elasticity, however, of demand is almost always quite high. The model cannot be taken as a guide to the solution of real price/capacity problems. It takes the analysis some steps along the path to a complete treatment, but stops well short of it.

Douglas and Miller's model is the most thoroughly worked out approach, but there are several other contributions which are addressed to the airline pricing problem. Mostly they are developed within the framework of the U.S. regulatory system, and are concerned with assessing its efficiency. Often efficiency is identified with what would have happened if there were competition unconstrained by regulation (e.g. see Keeler, 1972). This identification is made without reference to any optimising model. In an industry such as this, it does not follow that competition will, of its own accord, yield an optimum - this point is developed further in Sections 3 and 4. While several writers have made the points to be discussed here, we shall refer mainly to the contribution of De Vany (1975) who takes the analysis furthest.

This trade off between quality (lower delay) and price can be seen very easily in a regulatory framework. If a price is set, and there is no capacity constraint, capacity will expand until extra units add as much to costs as they do to revenue - this will be true under monopoly or competition. An increase in price leads to an increase in capacity and hence quality - this process has been observed in the airline industry, but it has been noted in others, such as motor fuel retailing (Edwards, 1962). It seems natural to consider the different levels of price, and see where cost reductions are met by reductions in quality, and to choose an optimum price/frequency solution in this way. Certainly this procedure seems to lead to an "optimum" price. But it is not the best possible price.

The error of this procedure is that it introduces a constraint to the solution - namely that revenues must cover costs. This restricts the number of possible price/frequency combinations, and excludes the optimum. It will be shown in Section 3 that it is not generally desirable that prices cover costs; optimality requires that firms make losses. This is so because frequency considerations introduce a type of economy of scale. This economy is dependent on industry size, not the size of an individual firm's output, and thus it cannot be internalised by the competitors. It is not possible to have competition with firms covering costs, and an optimal output level in the industry, for the same reasons as apply in the classic case of the national monopoly which is operated as a public utility to ensure optimality.

De Vany takes the analysis further, and tests whether, for the U.S. air market, frequency levels are optimal. To measure the value of frequency, he estimates a demand function with price and frequency as independent variables. The coefficients enable an estimate of the value, at the margin, of frequency increases. This enables an estimate of the benefits of an additional flight to be calculated, and to be compared with costs. Here the role of price and frequency in affecting demand are explicitly allowed for; a notable advance on the quality/price models. However the model is incomplete, as the effect of an additional flight on either delays (if load factors are less than 100%) or on numbers that can be carried (if existing flights are full) is not allowed for. Thus the marginal benefits are underestimated, and the optimal capacity would be larger. De Vany concludes that at existing (regulated) fares, routes which are competitive (more than two carriers) have the optimal frequencies. Given the way he has measured the marginal benefit, this will be true in a competitive situation, regardless of the price and load factor. This is discussed in Section 3.

It is difficult to characterise the state of knowledge, but some general propositions may be advanced. There is held to be a trade off between frequency and cost, and that this results in load factors being less than 100%. This is the type of view being accepted into more general writings on air transport economics (see Friedman, 1976, and Eads, 1975). This view is a simplification of the process. There will be a trade off, clearly observed, if there is regulation of price. If there is no regulation, and the objective is to determine optimal fares, capacity and load factors, there will be a trade off if there is a stochastic delay element, or more generally if there is stochastic demand, over the day or between different days. In practice, this trade off will almost always be present, because uncertainty will be present. The second proposition is that competition will be feasible and optimal without subsidy. This conclusion is challenged in the next section.

To obtain an adequate model, it is necessary to drop some unacceptable assumptions, made either explicitly or implicitly in other models. In the Douglas and Miller (1974) model, it is assumed that (a) the elasticity of demand is zero, (b) service quality does not affect demand and (c) optimality can be achieved in a competitive environment, with prices covering costs. In the De Vany model, assumptions (a) and (b) are dropped, though (c) is maintained. In addition, the estimate of the marginal value of a flight is only partial.

As a check on the plausibility of the results, we consider the optimality rules derived for an analogous problem, by Mohring (1972) and Turvey and Mohring (1975). The bus scheduling problem involves units of fixed capacity and greater quality with greater frequency. In addition, extra passengers impose congestion costs upon others. Turvey and Mohring conclude that optimal frequency requires that costs not be covered, when the service is provided under constant returns to scale. Prices to not have much to do with frequency directly, and load factors on buses are less than 100% because of the congestion aspects of additional passengers per bus. It is this congestion which yields the optimality conditions for pricing - prices should be set according to the additional congestion cost imposed by the passenger. It is to be noted that the solution to this problem, which is the result of setting up an explicit optimising model, is quite different to that proposed for airlines. In fact, the model and results of Section 3 will be consistent with the Turvey and Mohring analysis.

A MODEL OF OPTIMAL FREQUENCY, CONGESTION AND PRICE

It is easiest to appreciate the points made in the section above if a model which does not take account of stochastic delay is developed first. It is a simple matter then to make allowance for this form of congestion.

For the purposes here, it is adequate to use a consumers' surplus type maximand. Total benefits to consumers B, may be expressed as a function of the numbers of passengers, X, and of flights F. This latter term is introduced because frequency is assumed to affect benefits via reducing frequency delay. Thus, consumers' benefits can be expressed by (1).

$$B = B\ (X,\ F) \qquad (1)$$

In addition, some statements about the form of B( ) may be made. It is likely to be increasing at a decreasing rate in X and F, and that extra frequencies increase marginal benefits at a given X; i.e.

$$\frac{\partial B}{\partial X} \geq 0, \ \frac{\partial^2 B}{\partial X^2} \leq 0 \qquad (2a)$$

$$\frac{\partial B}{\partial F} \geq 0, \ \frac{\partial^2 B}{\partial X^2} \leq 0 \qquad (2b)$$

$$\frac{\partial^2 B}{\partial X \partial F} > 0 \qquad (2c)$$

For simplicity, assume that costs, C, depend on the number of flights only. It is easy to generalise to allow for passenger related costs. Thus

$$C = C\ (F) \qquad (3)$$

and $$\frac{\partial C}{\partial F} \geq 0, \ \frac{\partial^2 C}{\partial F^2} \geq 0 \qquad (4)$$

For convenience, assume that costs per flight are constant.

Optimality will require that B(X, F) -C (F) be maximized. Write this maximand as G (X, F) i.e.

$$G(X, F) = B(X, F) - C(F) \qquad (5)$$

It is very likely that, in practice, this function will be concave.

It is important to include a constraint on aircraft capacity - otherwise the numbers carried per flight could be arbitrarily large. Suppose that the seating capacity for each flight is given by m. Then the constraint will be

$$X/F \leq m. \qquad (6)$$

Write X/F as M(X,F). The maximization problem can be written out as

$$\text{Maximize} \quad G(X, F) \qquad (7)$$

$$\text{subject to} \quad M(X, F) \leq m \qquad (8)$$

$$X \geq 0 \qquad (9)$$

$$F \geq 0 \qquad (10)$$

The conditions for solving this can be written out as follows:

$$\left(\frac{\partial G}{\partial X} - \gamma \frac{\partial M}{\partial X}\right) \leq 0 \qquad (11)$$

$$\left(\frac{\partial G}{\partial F} - \gamma \frac{\partial M}{\partial F}\right) \leq 0 \qquad (12)$$

$$\left(\frac{\partial G}{\partial X} - \gamma \frac{\partial M}{\partial X}\right) X = 0 \qquad (13)$$

$$\left(\frac{\partial G}{\partial F} - \gamma \frac{\partial M}{\partial F}\right) F = 0 \qquad (14)$$

$$X \geq 0 \qquad (15)$$

$$F \geq 0 \qquad (16)$$

$$m - M(X, F) \geq 0 \qquad (17)$$

$$\gamma\,(m - M\,(X, F)) = 0 \qquad (18)$$

$$\gamma \geq 0 \qquad (19)$$

Where γ is the multiplier associated with the seating capacity constraint. It is to be interpreted as the value of increasing capacity by one unit, aggregated over all the flights.

Suppose that $X > 0$, $F > 0$. Then from (13) and (14) it will be the case that the terms in brackets will have to be equal to zero. If (17) is not satisfied by equality, then γ will be zero, and it will be necessary to set $\frac{\partial G}{\partial X}$ and $\frac{\partial G}{\partial F}$ equal to zero, i.e.

$$\frac{\partial G}{\partial X} = \frac{\partial B}{\partial X} = 0 \qquad (20)$$

and

$$\frac{\partial G}{\partial F} = \frac{\partial B}{\partial F} - \frac{\partial C}{\partial F} = 0 \qquad (21)$$

Condition (20) implies that the marginal benefit for passengers be set equal to zero, and (21) implies that the marginal cost of a flight be set equal to the benefits from greater frequency alone that it provides. Suppose that passengers are charged a fare p. They can be regarded as maximising consumers' surplus,

$$B\,(X, F) - p\,X \qquad (22)$$

The condition for (22) to be a maximum is

$$\frac{\partial B}{\partial X} - p = 0 \qquad (23)$$

If (23) is to be satisfied with (20), prices must be set at zero. This would be a case where aircraft are too large, and the required frequency cannot be achieved with full aircraft even at zero prices. This case can be dismissed for practical purposes.

It is almost certain that there will be a positive shadow price, γ, for seating capacity at the optimum. For (13) and (14) to be satisfied, it will be necessary to set

$$\frac{\partial B}{\partial X} - \gamma/F = 0 \qquad (24)$$

and $$\frac{\partial B}{\partial F} - \frac{\partial C}{\partial F} + \gamma m/F = 0 \tag{25}$$

Condition (24) states that passengers be charged the shadow price of capacity (for F flights) divided by the number of flights, F - i.e. the shadow price of a unit of capacity for one flight.

Condition (25) states that the marginal cost of a flight be set equal to the marginal benefit from increased frequency plus (on substituting from (24)) the marginal benefit from flights for passengers $\frac{\partial B}{\partial X}$, times the number of passengers on the flight, m. Recalling (23), it will also be true that

$$\frac{\partial B}{\partial F} + m\ p = \frac{\partial C}{\partial F} \tag{26}$$

This means that the revenue from the flight, m p, will be less than the cost of the flight, if $\frac{\partial B}{\partial F}$ is positive. This is equivalent to the case of additions to capacity when there are economies of scale (see Williamson (1966)).

In sum, it should be noted that when the benefits of frequency are explicitly considered, optimal load factors are 100%, except when aircraft are too large to be filled even at zero prices - a highly unlikely eventuality. The effect of considering frequency is to result in optimal prices not covering costs, when production costs exhibit constant returns to scale.

In a competitive situation, with constant costs, airlines will continue to schedule flights as long as they cover costs. One possibility is that the elasticity of demand is low, and that as extra flights are scheduled, they can only be covered by higher prices. This approximates the case studied by Douglas and Miller (1974). It is not a solution consistent with competition, however. If one airline provides many flights to reduce delays, but charges a high price and operates a low load factor, it can be successfully undercut by an operator with a high load factor. The benefits of the high frequency can be reaped by travellers on the original airline and on the new competitor. Hence this would be an unstable position.

It is much more likely that elasticity of demand exceeds unity, and lower prices yield higher revenues. If this is so, it will always be optimal and profitable to have full aircraft, since if aircraft are not full, extra revenue can be obtained at no extra cost. Airlines will add flights as long as marginal revenues exceed marginal costs. Extra revenue can be obtained from lowering price, or increasing quality, and stimulating demand indirectly. Competitive pressures will force prices down to such a level as can cover the marginal cost of a flight. At this price, airlines add flights such that the revenue from the additional travellers induced by the service improvement equal the costs of a flight. Writing revenue R, as a function of traffic and price, and traffic as a function of price and quality

$$R = p.x = p.\ f(p,F) \tag{27}$$

Then

$$\frac{\partial R}{\partial F} = p.\frac{\partial x}{\partial p}\ \ \frac{\partial p}{\partial F} \tag{28}$$

$$= x\ (\frac{p}{x}\ \ \frac{\partial x}{\partial p})\ \ \frac{\partial p}{\partial F} \tag{29}$$

$$= x\frac{\partial p}{\partial F}\ \ \xi_D \tag{30}$$

(where ξ_D represents the price elasticity of demand)

The term $\frac{\partial x}{\partial p}$ indicates the equivalent in price terms of the change in frequency. Another way of looking at it is as the value of the increased frequency at the margin. If the average equals the marginal value, $X\ \frac{\partial p}{\partial F}$ is equal to the total benefits from an additional flight reaped by all the original passengers. However, the term in (30) is a measure of the revenue from the extra travellers. In competition, this will be set equal to the marginal cost of the flight. This amount is less than the marginal benefits from the flight $(X\ \frac{\partial p}{\partial F} + \xi_D\ x\ \ \frac{\partial p}{\partial F} = \frac{\partial B}{\partial F} + m\ p)$.

It will be the case that the marginal cost of a flight is set equal to $\xi_D\ x\ \frac{\partial p}{\partial F}$ regardless of what price is actually set. For example, it is possible that p is set high, by regulation, and the load factor is low. It will still be the case that this is the measure of extra revenue from a flight. This result will generalize to the situation where stochastic delay is present. The amount $x\ \frac{\partial p}{\partial F}$ is the measure that De Vany uses as the marginal benefit from a flight. He finds that this is equal to the cost of an additional flight (De Vany, 1975, p.342, 343), and claims that this indicates efficiency. In fact, since his elasticity estimate is very close to unity, as indicated above, this will always occur in a competitive situation, regardless of whether price and load factors are optimal. To this extent, our model of competitive behaviour fits de Vany's empirical results.

Allowance for stochastic delay makes some difference to the maximization problem. The cost in financial terms of stochastic delay, D, can be written as depending upon the numbers of passengers and flights; it will be increasing in passengers and decreasing in flights.*

* An alternative would be to include this delay directly in the Benefit function.

$$D = D\ (X,\ F) \tag{31}$$

The objective will now be to maximize G(X, F) less D(X, F) - call this $G^1(X, F)$. The same constraints will be present.

Taking X, F as being positive, the maximization will yield conditions of the form

$$\frac{\partial G^1}{\partial X} - \gamma \frac{\partial M}{\partial X} = 0 \tag{32}$$

$$\frac{\partial G^1}{\partial F} - \gamma \frac{\partial M}{\partial F} = 0 \tag{33}$$

If the multiplier γ is zero, (32) will imply

$$\frac{\partial B}{\partial X} - \frac{\partial D}{\partial X} = 0 \tag{34}$$

For probable forms of D( ), delays will become very high with full aircraft. Thus in the optimum, $\gamma = 0$, and the marginal benefit of a passenger should be set equal to the marginal delay cost of a passenger. Suppose that a typical passenger faces average delay costs and cannot affect frequency. Passengers will behave as if they were maximizing (with $\overline{D(X,F)/X}$ as a constant).

$$B\ (X,\ F) - X.\overline{D(X,\ F)/X} - p.X \tag{35}$$

and to do so, they will set

$$\frac{\partial B}{\partial X} - p - \frac{D}{X} = 0 \tag{36}$$

Marginal benefits from a flight, less expected delay costs, are set equal to price. To achieve satisfaction of (34) it will be necessary to set a price p*, such that

$$p^* = \frac{\partial D}{\partial X} - \frac{D}{X} \tag{37}$$

That is, passengers should be charged the marginal delay cost, less that cost they face themselves. This is a standard result of congestion theory.

Satisfaction of (33), with $\gamma = 0$, entails that

$$\frac{\partial B}{\partial F} - \frac{\partial D}{\partial F} - \frac{\partial C}{\partial F} = 0 \tag{38}$$

This would mean that the marginal cost of a flight should be set equal to the benefits of reduced frequency delay, plus the reduction in stochastic delay cost.

It is most likely that, at the optimum, costs would not be covered. Suppose that the stochastic delay function is of the form

$$D + D(\frac{X}{F}) \qquad (39)$$

Then it can be shown that with prices as set in (37)

$$\frac{\partial D}{\partial F} = \frac{\partial C}{\partial F} = p^* \frac{X}{F} \qquad (40)$$

This is a result from congestion theory (see Mohring, 1976, pp. 23,24). Thus, as long as $\frac{\partial B}{\partial F} > 0$, $\frac{\partial D}{\partial F} = p.x$ will fall short of $\frac{\partial C}{\partial F}$. Thus, at the optimum, prices will not cover costs, because of the effects of flight frequency on frequency delay.

In fact, the delay cost function will not be as simple as that shown in (39). The level and cost of stochastic delay depend upon the relationship of demand to capacity, $\frac{X}{F}$ (which determines the probability of not obtaining the preferred flight) and on the amount of capacity offered, F (which determines the delay if the preferred flight is not obtained). A preferable form of function would be

$$D = D(\frac{X}{F}, \frac{1}{F}) \qquad (41)$$

Douglas and Miller (1974) use this form. Because of the additional term in F, it will now be true that $\frac{\partial C}{\partial F} + \frac{\partial D}{\partial F} > \frac{X}{F} . p^*$ (ignoring frequency delay) and costs will not be covered at the optimum.

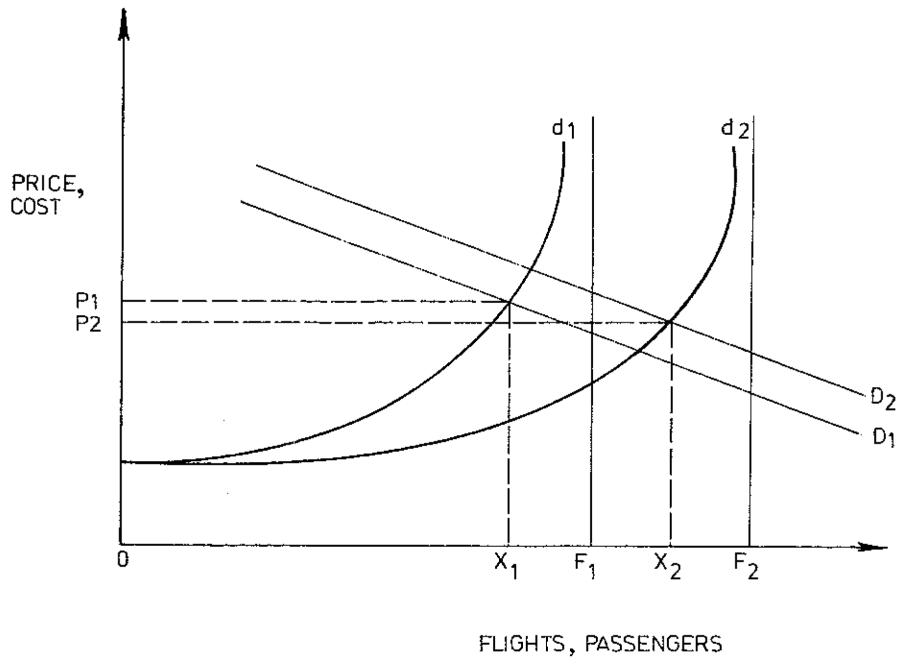
If there is competition, firms will add flights until the marginal revenue from that flight, $\frac{X}{F} . p$, $(= \xi_D \times \frac{\partial P}{\partial F})$ equals the marginal cost of its provision. Competition will force prices down, and no choice of price or frequency could increase profits. It will be the case that equation (37) will be satisfied: given the level of capacity offered, airlines will price it efficiently, but the wrong level of capacity is offered. Equation (38) will not be satisfied, indicating that improvements are possible with more capacity. Comparing this to the Douglas and Miller solution, it can be seen that they follow (38) for a given X, and charge prices higher than would satisfy (37). The price and frequency levels chosen would not be consistent with completely competitive behaviour.

The competitive solution yields too low a capacity as compared with the optimum. If more capacity is supplied, it will have to be at a lower price, and the R.H.S. of (37) will fall. Thus the optimum will have a lower load factor than the competitive solution, because the full benefits of extra capacity are taken into account, and the lower price will not be enough to cover costs.

While frequency delay, on its own, does not provide any reason why load factors should be less than 100%, it will be a force for lower load factors when it operates with stochastic delay. Suppose that (38) is satified, and the cost of frequency delay increases, i.e. at F*, $\frac{\partial B}{\partial F}$ increases. At the new equilibrium, with a higher F, $\frac{\partial D}{\partial F}$ will have fallen. This implies a lower load factor, and lower prices.

Of the models presented here, the more inclusive one is to be preferred. It combines the stochastic and frequency delay aspects and indicates that two, rather different processes are at work. One is a type of external economy, with additional flights creating benefits which cannot be reaped in full by competitive firms. Frequency delay represents this process only, and stochastic delay combines this with a congestion process, whereby congestion depends on the relationship of demand to capacity. It is this latter process that provides the rationale for load factors less than 100%. It is the other process that results in prices not covering costs at the optimum. In addition, this model highlights the need for the passenger to be paying the marginal delay costs resulting from his decision to travel. While this is a common condition in most congestion models, it has been overlooked in airline pricing problems.

The points made in the models above can be illustrated by means of a simple diagram. It is not necessary to show the case where stochastic delays are zero.



Two capacity levels are allowed for, F_1 and F_2, and corresponding to these are two marginal delay cost curves, d_1 and d_2 and two demand (marginal benefit) curves, D_1 and D_2. Average delay cost curves are not shown. The optimum with capacity of F_1 is with price p_1 - this price includes the average delay cost component. Marginal benefits for passengers are set equal to marginal delay costs.

When capacity is expanded to F_2, marginal delay costs fall, because the ratio of X to F falls for any X. Also service frequency increasing means that, on average and at the margin, benefits from travelling increase. There will be a new optimum at p_2. This price includes the cost of delay that the passenger must suffer.

Whether the extra flights are worthwhile depends on whether their extra cost is greater than additional benefits. These can be shown as the difference in the areas between D_1 and d_1 and D_2 and d_2. This difference in areas can be broken up into the difference between D_1 and D_2, to the left of X_1 (the benefits of greater frequency to the original passengers), plus the difference between d_1 and d_2 to the left of X_1 (the benefits of a greater probability of obtaining the preferred flight) plus the triangular area between D_2 and d_2 to the right of X_1 (the benefits to the extra passengers). Airlines will be able to recoup some, though not all of these benefits, in higher revenue.

It is possible that the demand curves, D, might cross the capacity line higher than the marginal delay cost curves, d, do. This is unlikely when we are dealing with an extended single period, not followed by different periods. Then, stochastic delay will tend to infinity as the load factor tends to 100%. However, typically, a busy period will be followed fairly soon after by a slack period, and marginal delay costs will not have time to mount to infinity, even if load factors reach 100%. If may be optimal for some busy periods to have very high load factors, if there are slack periods following.

SOME MODIFICATIONS AND PROBLEMS OF APPLICATION

To apply a model such as that of Section 3 it is obviously necessary to obtain estimates of the frequency and schedule delay functions. Once this is done, it is a straightforward matter to calculate optimal prices, load factors and delay levels. The greater the value of extra frequency, the lower the load factor will be, and the greater divergence between revenues and costs. The lower the stochastic delay cost, the greater the load factor called for.

This suggests the desirability of measures designed to reduce stochastic delay. This is recognised by Douglas and Miller (1974, pp.92,93), who explore the possibilities briefly. Several types of fares are suggested, including quotas of cheaper seats (part charters), standby fares, or seats which are booked ahead but where passengers are allocated to flights by the airline. The effect of these devices is to reduce stochastic delay, and given passengers a choice of service qualities and fares. To the extent that passengers are able to obtain service which conforms to their preferences, welfare is improved. If their trade-off between quality and price forms a continuum, then an infinity of prices and quality levels is desirable. While this is impractical, it is clear that as great a range of choice as is possible should be offered. It is possible to go further and suggest which options are best.

Two broad classes of price structure can be identified. One is where there are specific allocations of seats to specific classes, with the class associated with high delays having low fares - this is the part charter structure. The other is where empty seats, if available, are filled by lower fare passengers - this would include the standby fares, and also the case where the airline allocates the flight to the passenger. Both serve to increase choice, and to lower the costs of stochastic delay, by inducing those with the lower delay costs to bear more of the delay.

Of the two price structures, the part-charter structure is less efficient. This is because of the way delay costs are borne by passengers. For an individual, delay is encountered only if he fails to obtain a seat on a preferred flight. If the two classes of fare are separated only by delay and fare differences, it will be always sensible for the individual to try first for the lower fare, and if this is unavailable, try for the higher fare. If there is this substitution between fares possible, it will degenerate into a system of a lower price for the first block of seats sold. There will be some reduction in delay and delay cost, some reduction in revenue, and optimal load factors will increase slightly. For those willing to pay a higher price to avoid delay, there will be more capacity available, as those not willing to pay will wait till the next flight.

Airlines recognise the substitution possibilities, and often restrict the availability of the lower fare to a specific group. Often the fare is open only to those who book some time ahead, or the fare is tied to an inclusive tour. It may be that this segmentation increases the liklihood that the low delay cost users bear the greater delays, but there is no certainty that this will happen. It may be that early booking or inclusive tours are desirable in themselves, but they do not have any specific effect on the delay process. As a device for encouraging the low delay cost passengers to bear the delays, a system of dividing a flight into high and low load factor categories is likely to prove fairly ineffective.

To be effective in achieving this separation, the price structure must be directly related to the delay process. A standby fare, or early booking fare where passengers are allocated to flights would directly associate delays with the price offered. It would be possible to switch from one fare to another only if a substantial change in expected delay were accepted. Those passengers who encounter small delay costs would be induced to bear virtually all delays. Delays would not be reduced to zero for the higher fare passengers, though they would be reduced substantially.

With the stochastic delay cost being reduced for each load factor and frequency, the optimal load factor will increase and the optimal frequency will tend to decrease. The level of the optimal load factor depends on the level of stochastic delay costs. Normally these would not fall to zero, though they could fall considerably if the proportion opting for the lower fare is large enough to minimize delays for the high fare passengers, yet is small enough to include only those passengers with small per unit delay costs. It will still be true that in the optimum, the marginal delay cost imposed by a passenger equals the price he pays; the marginal delay cost imposed by a standby passenger is less than that imposed by the high fare passenger.

Other variations of the price structure could be devised which would increase net benefits. An obvious one in a situation of uncertainty is to make a distinction between regular and irregular passengers or between those whose demand correlates with high delays and those whose demand does not (a fare structure with standby fares already does this to an extent). However, the main benefits will be reaped by a system that distinguishes between delay cost categories. The high load factors and low fares of charter operators provide an example of the benefits from distinguishing delay cost categories. Similar load factors and average fares can be achieve on scheduled flights with the appropriate price structure. The exact price structure chosen will be important however. Some seemingly attractive structures, such as part charters, may yield only a fraction of the benefits that better designed structures will provide.

In order to set prices and frequencies, it is necessary to know the valuation that travellers put on frequency, and on stochastic delay. One approach is to value the two forms of "delay" from the value of time - as is done by Douglas and Miller. As has been pointed out, this need not yield an accurate estimate, since "delay" need not be time wasted. It is preferable to estimate the value of frequency more directly, as is done by De Vany (1975). The valuation obtained from comparing coefficients of price and frequency in a demand function will be essentially a marginal valuation. It is obtained from estimating the change in frequency which will have the same impact on demand as a change in price.

When the value of frequency is being estimated, cognizance will have to be given to the fact that it forms part of two processes. A change in frequency lowers delays directly, and through the reduction of load factors. The impact on demand will depend on the load factor at which the frequency change takes place. Thus a study which measured the effect of frequency would have to identify separately the direct effect and indirect effect, by including the load factor and frequency as arguments. For a service with given load factor and frequency, the impact of a change in frequency can be estimated (since demand changes, strictly it would be necessary to compare the old and new equilibria). The direct and indirect effects do not correspond to effects on frequency and stochastic delay, as the latter is affected directly and indirectly by frequency changes. The marginal valuation thus obtained is not a measure of the full value of the benefits of an additional flight, but with knowledge of the price elasticity of demand this can be calculated using (30).

Another problem is that of devising a test for estimating the marginal value of service frequency accurately. The marginal valuation will differ across flights, and thus testing will have to use a form which recognizes this possibility. The marginal value of a change in frequency $\frac{\partial B}{\partial F}$ is obtained from considering the value of a change in traffic, and the responsiveness of traffic to frequency. Thus

$$\frac{\partial B}{\partial F} = \frac{\partial B}{\partial X} \frac{\partial X}{\partial F} = p \frac{\partial X}{\partial F} \qquad (42)$$

Note that this does not equal the full benefit from a frequency change. Alternatively, we may write the elasticity of demand with respect to frequency ξ_F as

$$\xi_F = \frac{\partial X}{\partial F} \frac{F}{X} = \frac{1}{p} \frac{\partial B}{\partial F} \frac{F}{X} = \frac{\partial B}{\partial F} \frac{\partial F}{pX} \qquad (43)$$

The discussion in Section 2 was inconclusive about how $\frac{\partial B}{\partial F}$ would vary. It was suggested that the impact on schedule delay would decline with increases in frequency but that the unit cost of this delay increases. It is possible that the benefits of additional flights remains fairly constant over a range of frequencies, or they could fall. If the former were true routes with different densities and frequencies might be considered in the same equation. The important constraint then on testing is that $\frac{\partial B}{\partial F}$ or $p\frac{\partial X}{\partial F}$ should remain constant over the sample.

This would mean that when p is large $\frac{\partial X}{\partial F}$ would be small. Put another way, if $\frac{\partial B}{\partial F}$ is regarded as constant, the elasticity of demand with respect to frequency will rise with (F/P X), the reciprocal of the revenue per flight. As the revenue per flight rises, the elasticity should fall. This is contrary to the expectations of some, for example, Ellison and Stafford (1974, p.131), but it is not really inconsistent with their statistical results. It is inconsistent with fitting a model with a constant elasticity to data with differing flight revenues. Even if it were believed that $\frac{\partial B}{\partial F}$ fell significantly with increases in frequency, it is unlikely that the fall would offset the changes in revenue per flight which it would typically be associated with so as to result in a constant elasticity.

All of this stands to reason. An additional flight on a low density, long distance route may be just as valuable, in absolute terms, as an additional flight on a high density short distance route. But its effect on demand,in relative terms, might be a lot less. It is difficult to believe that a doubling of frequency on a long distance route will lead to 50% increase in patronage, but it is quite possible that a 10% increase in frequency may lead to a 5% increase in patronage on a short distance route.

In setting up a model to measure the marginal valuation of frequency changes, it is important that its form does not impose unrealistic constraints on the parameters. Where demand, x, is a linear function of frequency, F, the term $\frac{\partial X}{\partial F}$ will be constant, and this is improbable. In a model where elasticity with respect to frequency is constrained to be constant, there is an improbable restriction on the way $\frac{\partial B}{\partial F}$ must vary. An adequate model would be specified in a way that $\frac{\partial B}{\partial F}$ and $\frac{\partial X}{\partial F}$ are allowed to vary with changes in F.

CONCLUSIONS

The conclusions here are at variance with those of several of the recent papers on airline pricing. Since the optimal fare is likely to be below cost, it would not be adequate to allow frequencies and fares to be set by competition alone. Nor is it sufficient to have a competitive industry subject to regulatory controls, since it would not be feasible to induce competitors to supply the optimal capacity at the right price by means of controls such as those on price and capacity. The industry could be operated as a public utility like a national monopoly. This is not necessary, however, because the economy of scale is industry dependent, and not dependent on individual firms outputs, it is quite possible to have a stable outcome with a number of competitors, as long as there is a subsidy to reduce the price to the optimum level.

There may be reasons why prices should be greater than marginal cost, e.g. the industry may be subject, like others, to optimal commodity taxes, as proposed by Diamond and Mirrlees (1971) and others. This does not mean that it would then be optimal to just cover costs; such a result could occur only by chance. There remains a strong liklihood that the industry will have to be run at a financial loss.

Normally, load factors should be less than 100% - this results from the stochastic nature of demand, not because of the frequency problem. The technique of keeping demand constant while working out optimal frequencies has been shown to be highly misleading. When the effects of service quality and price on demand have been allowed for, the results change and become comparable to those for analogous problems, such as the bus frequency and pricing problem.

REFERENCES

De Vany, A. (1975), "The Effect of Price and Entry Regulation on Airline Output, Capacity and Efficiency", *Bell Journal of Economics,* Spring.

Diamond, P. and Mirrlees, J.A. (1971), "Optimum Taxation and Public Production", *A.E.R.,* March and June.

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

Eads, G.C. (1975), "Competition in the Domestic Trunk Airline Industry: Too Much or Too Little?", A. Phillips, (Ed.) *Promoting Competition in Regulated Markets,* (Washington, D.C.: Brookings Institution).

Edwards, H.R. (1962), *Competition and Monopoly in the British Soap Industry,* (London: Oxford University Press).

Ellison, A.P. and Stafford, E.M. (1974), *The Dynamics of the Civil Aviation Industry,* (Saxon House).

Friedman, J.J. (1976), *A New Air Transport Policy for the North Atlantic,* (New York: Atheneum).

Keeler, T.E. (1972) "Airline Regulation and Market Performance", *Bell Journal of Economics and Management Science,* Autumn.

Miller, III, J.C. (1972), "A Time of Day Model for Aircraft Scheduling", *Transportation Science,* August.

Mohring, H. (1972), "Optimisation and Scale Economies in Urban Bus Transportation", *American Economic Review,* September.

Mohring, H. (1976), *Transportation Economics,* (Cambridge, Mass: Ballinger)

Turvey, R. and Mohring, H. (1975), "Optimal Bus Fares", *Journal of Transport Economics and Policy,* September.

Williamson, O.E. (1966), "Peak Load Pricing and Optimal Capacity under Indivisibility Constraints", *A.E.R.*