Theory and Methodology

Competition in a deregulated air transportation market

Nicole Adler

School of Business Administration, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel

Received 6 March 2000; accepted 14 May 2000

Abstract

Under deregulation, airlines developed hub-and-spoke (HS) networks enabling them to aggregate demand, increase frequency, reduce airfares and prevent entry into the marketplace. This research evaluates airline profit based on micro-economic theory of behaviour under deregulation. Through a two-stage Nash best-response game, equilibria in the air transportation industry is sought to evaluate the most profitable HS network for an airline to survive in a deregulated environment. In the first stage of the game, an integer linear program aids in generating potential networks. In the second stage, a nonlinear mathematical program maximizes profits for each airline, based on the networks chosen by all participants. The variables of the mathematical program include frequency, plane size and airfares. In an illustrative example, both monopoly and duopoly solutions are attainable as a function of demand. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Airlines; Air transportation; Scheduling; Location; Game theory

1. Introduction

In this paper, we are attempting to evaluate the most profitable hub-and-spoke (HS) network for an airline in a competitive environment. A series of mathematical programs have been developed for this purpose, which are used within a game-theoretic framework. Until the early 1990s, airline competition and airline network strategies were generally treated as separate subjects in the literature. Several papers have since been written using HS networks. Hansen (1990) developed an \( n \)-player, non-cooperative game in which the airline’s sole strategy set is frequency of service. The set of simplifying assumptions included fixed airfares, adequate capacity, inelastic demand to price and service level and consideration of nonstop and one-stop services only. Hong and Harker (1992) developed a two-stage, game-theoretic representation of an air traffic mechanism for slot allocation. Using a quasi-variational inequality technique to solve a Cournot–Nash model, they prove the existence of a unique solution. Dobson and Lederer (1993) developed a mathematical program to study the competitive choice of flight schedules and route prices by airlines operating in a pure HS (i.e., single hub) system. Assumptions in their model include a single aircraft size, one class of customers, no traffic originating at or destined...
to the hub airport, airline variable costs dependent on flying time alone and zero variable passenger costs. An additional assumption requires that duopolists serve the identical set of spoke cities using the same hub. Hendricks et al. (1997) wrote a theoretical paper discussing equilibria in airline networks. They study duopolies using a two-stage game in which two carriers simultaneously chose their networks and then compete for travellers. They argue that when carriers compete aggressively (e.g., Bertrand-like behaviour), monopoly is an equilibrium outcome and no duopoly equilibrium exists if both carriers choose HS networks. They further attest that duopoly equilibria exist either in non-HS networks or when HS-network carriers do not price aggressively and there are a sufficient number of nodes.

More recent operations research papers include surveys, such as that of Campbell (1994). Pirkul and Schilling (1998) develop a Lagrangian relaxation model for designing a single allocation HS system, whilst Marianov et al. (1999) use an algorithmic approach to deal with direct competition. Their one or two hub model allows an airline to capture customers if the competitor can provide a shorter distance (or time) from the origin to the destination. However, the model is entirely cost-based and does not consider pricing policies, thus allowing an airline with a shorter distance flight from airport $i$ to airport $j$ to charge a substantially higher price and still capture the entire relevant market. Furthermore, the issue of frequency and quality of service is not considered. Erby et al. (2000) solve the capacitated, multiple allocation, hub location problem using a multiple integer linear programming technique. They “locally” solve large problems quickly due to streamlined formulations and superior algorithms, but real world problems are more complicated. This research attempts to combine the two fields and account for economic realities, whilst using operational research techniques to solve the question of the most appropriate HS network in a competitive, deregulated environment.

Modelling assumptions are discussed in Section 2. In a game theoretic environment, each airline chooses a network including hub(s) and their connections to the spoke airports. This is attained through an integer linear program specified in Section 3. Subsequently, learning about competitors’ decisions, airlines attempt to maximize profits using a nonlinear mathematical program as developed in Section 3. The game and its rules are described in detail in Section 4 and an illustration of this process is discussed in Section 5. A summary of the paper and conclusions can be found in Section 6.

2. Assumptions

It is assumed that a HS network reduces total costs to the airline due to economies of scope and scale and that passengers will be prepared to travel over two or three legs if necessary. The US air transportation market today shows that these assumptions are reasonable. The model requires airlines to choose a HS system with one or two potential hubs. In reality, an airline may choose a specific HS network and add a few direct connections where demand justifies such a decision. Inclusion of these arcs would not add to the complexity of the solution procedure described. It is further assumed that an airline is entirely free to choose its most preferable network and purchase the rights to land and take-off as required. With the slow disappearance of the flag-carrier phenomenon, this assumption is deemed reasonable.

Passenger utility is defined as a function of frequency, willingness-to-pay for a direct flight and airfare. The utility of a traveller flying via a hub is based on the minimum frequency along one of the legs of the route, since this will cause the most important restriction on choice. The utility function also ensures that passengers forced to travel via one or two hubs will pay less than those travelling directly. Whilst this clearly does not include all possible factors affecting a passenger’s decision to travel, it does ensure tractability of the problem outlined. Maximum total origin–destination (O–D) passenger demand is assumed to be known and the carrier’s strategy, therefore, affects the number of passengers carried but not the total volume of demand. Sensitivity analysis can be used to understand the effects of such an increase or
decrease on maximal demand. A multinomial logit model’s utility function is used to describe travelers’ preferences to determine the market share of an airline, when compared to its competitors, as discussed in Alamdari and Black (1992).

Two fares are used as decision variables, one for business travellers and the other for non-business, from node $i$ to node $j$, per airline. This results in two prices per O–D per airline, which can be increased if a more complex analysis of airfares is required. The fare from origin $i$ to destination $j$ is assumed to be the same as that of $j$ to $i$ per airline. This reduces the number of airfare decision variables to $1/2n(n-1)$ for computational tractability but the assumption can be easily removed if necessary. Airlines have developed more complicated pricing strategies based on yield management technology, but this is ignored due to the paper’s purely strategic perspective. No interlining is permitted, thus each traveller type chooses one airline or not to fly based on price, frequency and value-of-time. In the real world, airlines attempt to differentiate their product to avoid interlining through pricing, minimum connecting times and marketing techniques such as frequent-flyer programs. The likelihood of lost baggage and missed connections also increases with interlining, thus encouraging travellers to avoid this option.

Operating costs of the carrier are assumed to be a function of frequency and include parameters for scale and network economies. This assumption is reasonable if the majority of flights are of a similar length (as in Western Europe, for example). Fixed airline costs such as the purchase of aircraft and accrued capital costs are ignored. In addition, plane sizes are limited so that the same size aircraft is used in both directions on a single path, a reasonable assumption given that fleet size variation is normally restricted.

Finally, airlines choose their strategic network and decision variables in two stages to facilitate the game-theoretic approach. First, all airlines choose their networks and whether or not to offer services concurrently, then they choose frequencies, plane sizes and airfares simultaneously, based on the knowledge of all the airlines’ choices during the first stage.

3. Mathematical model formulations

In this section, an integer linear programming approach is used to generate potential networks, the profitability of which will be subsequently evaluated by a nonlinear mathematical programming model. There are many possible methods of producing a connected HS network, the most direct of which is to simply connect spoke nodes to a chosen set of hub nodes according to a minimum distance criterion using integer linear programming such as

\[
\text{Minimise} (\varepsilon_1 + \varepsilon_2) \phi + \sum_{j=1}^{n} \left[ a_{ij} x_{ij} + a_{ij} x_{ij} \right] \quad (1)
\]

subject to

\[
x_{ij} + x_{ij} = 1 \quad \forall j, \ j \neq i_1, \ i_2, \quad (2)
\]

\[
\sum_{j=1}^{n} x_{ij} - \sum_{j=1}^{n} x_{ij} = \varepsilon_1 - \varepsilon_2, \quad (3)
\]

\[
e_i \geq 0, \quad i = 1, 2, \quad (4)
\]

\[
x_{ij}, x_{ij} \in \{0, 1\} \quad \forall j, \ j \neq i_1, i_2, \quad (5)
\]

where $i_1$ is hub number 1; $i_2$ hub number 2; $\varepsilon_1 - \varepsilon_2$ a variable measuring the balance of a solution; $a_{ij}$ the distance between nodes $i$ and $j$ and $x_i$ is the zero–one variable, where 0 represents no connection and 1 represents a single connection.

For $\phi$ equal to zero, the above ILP minimizes distance and may result in an almost pure HS system if one of the hubs is geographically further away from all other nodes. Since the hubs are supposed to represent the “centre” of the network, with all other nodes acting as spokes, a second solution, whereby both hubs have a reasonable number of connections, may also be considered. In addition, a single hub may be unable to satisfy the demand, given the severe congestion of large airports around the world. This is the rationale behind Eq. (3), which enables a more “balanced” network if necessary. For a large enough $\phi$, the network will be completely balanced, with approximately half the nodes connected to one hub and the remainder connected to the second hub. Values of $\phi$ can be selected according to the
reduced cost of the dual, where \( \phi \) equals zero. Alternative formulations minimizing total passenger travel distance or total number of travellers flying over more than one-leg journeys, will not be considered in this paper.

The mathematical program is based on a micro-economic model for an airline acting under deregulation (see Adler, 1999). Decision variables include airfare per O–D trip, traveller type and airline \( (p_{ija}) \), aircraft size per leg per airline \( (PS_{hi(k,a)}) \) and frequency per directed leg per airline \( (f_{ia}) \). The nonlinear mathematical program is also solved per airline per network chosen. An airline’s profit function is maximized subject to three sets of constraints. The profit function is based on an airline’s revenue and cost functions. The revenue function includes receipts based on airfares, maximal demand and the airline’s market share, which in turn is based on travellers’ utility functions. The airline’s operating cost function is assumed to be a constant elasticity-of-substitution Cobb–Douglas function for a set \( Arc(a) \) of all existing legs in airline \( a \)’s specific network configuration. That is

\[
C(f) = \left[ \sum_{k \in Arc} (f_k)\gamma \right]^\beta, \quad z, \beta > 0,
\]

where \( k \) is leg index, \( z \) the parameter reflecting scale economies and \( \beta \) the parameter reflecting network economies.

This class of function is general enough to capture the cost of operating different types of networks with a varying number of routes and layout patterns. Star and Stinchcombe (1992) in their economic analysis of HS and related systems argue that these systems are optimal under a variety of cost and demand configurations, typically demonstrating large, pervasive economies of scale. The cost function in Eq. (6) is monotonically increasing in frequency and exhibits increasing returns to scale if \( x\beta < 1 \). Thus, as frequency increases, the additional cost per flight decreases.

McShane and Windle (1989) reported in a US-based study that a 1% increase in hubbing increases average costs by 0.11%, strongly indicating that \( \beta < 1 \). Increasing frequency in a HS system enables the airline to deter entry into the market-place by carriers attempting to provide a direct service between \( i \) and \( j \), where this route does not yet exist.

Additional costs result from payments made by the airline in the form of airport landing and passenger charges. Landing charges (LC) paid to the arrival airport are based on maximum take-off weight defined by airplane type. Passenger charges (PC) are paid to the departure airport. Passenger charges include the full tariff paid at the first departure airport and a transfer charge paid at subsequent hubs, when the passenger is carried on two or more legs. This pricing system is in line with present rules for most international airports. Since there are many different types of charges, the LC and PC can be modified to include other relevant charges such as handling, night and noise charges and will be referred to as landing-related and passenger-related charges. The model proposed can also be used to evaluate the effects of the selected HS system on hub airports. Airport quality is beyond the scope of this paper and is discussed in detail in Adler and Golany (2000).

A multinomial logit (MNL) model can be used to compute airline market share, given maximum demand levels. Detailed evaluation of the MNL model is provided in Ben-Akiva and Lerman (1985). Alamdari and Black (1992) discuss the use of the logit model in evaluating the influence of liberalization on passenger demand. They argue that “simple all or nothing models that assume the cheapest airline is chosen by all the passengers are not suitable for determining airlines’ market share. Passenger demand is influenced by a combination of fare and the many attributes that make up the quality of service provided”.

The utility of a single passenger represents preferences concerning frequency, non-stop service and price. Model parameters are drawn from Adler and Berechman (2001), and the utility function is defined as follows:

\[
V_{ija} = \begin{cases} 
\lambda_w w_{ija} \left( \min_{a \in R_{ija}} f_{ia} \right)^\gamma - p_{ija}, & \text{travelling from } i \text{ to } j \text{ via airline } a, \\
0, & \text{travelling via a different mode or not travelling,}
\end{cases}
\]
where \( i, j \in N \) are the node indices; \( s \in \{ b, nb \} \) the type of traveller, either business (\( b \)) or non-business (\( nb \)); \( \gamma \), the parameter reflecting frequency elasticity of demand for traveller type \( s \); \( \delta \), the parameter representing willingness of traveller type \( s \) to pay for a direct flight;

\[
R_{ija} = \{ k | k \in Arc, \text{ leg } k \text{ belongs to route from node } i \text{ to node } j \text{ for airline } a \}
\]

\[
w_{ija} = \begin{cases} 
1 + \delta_s & \text{for direct travel,} \\
1 & \text{for indirect travel,} 
\end{cases}
\]

and \( p_{ija} \) is the airfare from airport \( i \) to airport \( j \) for traveller type \( s \) on airline \( a \).

This utility function can then be introduced into the logit model to calculate airline \( a \) market share for the O–D market \((i,j)\) per traveller type as follows:

\[
MS_{ija} = \frac{e^{(\lambda_{ija}(\min_{k \in R_{ija}} f_{KA})^{\gamma} - p_{ija})}}{1 + \sum_{d \in D} e^{(\lambda_{ija}(\min_{k \in R_{ija}} f_{KA})^{\gamma} - p_{ija})}}.
\]

The addition of 1 in the denominator of the market share function provides the traveller with the additional choice of not travelling, or at least not flying, consequently attaining zero utility. According to Hong and Harker (1992), the inclusion of the “no purchase” option captures the price elasticity of total demand and prevents airlines from over-charging. Numerous utility functions have been included in various papers, in an attempt to capture all passenger decision rules (see Alamdari and Black, 1992). Three parameters have been selected in this paper: airfare, service frequency and a value-of-time parameter. The first two variables were considered relatively important in most studies. Minimum service frequency can capture more than surface understanding of a HS network. It is a strategic variable enabling an airline to use product differentiation as a market entry deterrent. Furthermore, the level of frequency captures carrier costs directly and flying time indirectly. Value-of-time can be interpreted as the additional amount the consumer is willing to pay for a direct, non-stop service. It establishes whether an airline has an advantage over competitors with respect to network combination and the choice of hubs. Only passengers travelling directly pay the additional expense, an immediate result of the airline’s choice of hub.

The mathematical program’s objective function for airline \( a \) can then be stated as follows:

\[
\text{Max } \pi_a = \sum_{p, f, PS} \sum_{s} \left\{ d_{ija} MS_{ija} \left( \frac{p_{ija}}{p_{ija}^*} \right) \right\} - \eta \left( \sum_{k \in R_{ija}} f_{ka}^{a} \right)^{\beta} - \sum_{k \in R_{ija}} \left( a_1 + a_2 P_{SP}^{SEAT} L_{Ct} f_{ka} \right)
\]

subject to

\[
L_{ka} \leq L_{SP}^{SEAT} \leq U_{ka} \quad \forall k \in Arc(a), \ a \in A,
\]

\[
\sum_{a} f_{ka} \leq C_i \quad \forall k \in Arc(a), \ a \in A,
\]

\[
f_{ka} > 0 \quad \forall k \in Arc(a), \ a \in A,
\]

where \( d_{ija} \) is the traveller type \( s \) demand from node \( i \) to node \( j \) (passengers per time period); \( c(k,a) \) the commencement node for leg \( k \) of airline \( a \); \( e(k,a) \) the end node for leg \( k \) of airline \( a \); \( h(k,a) \) the aircraft size index for leg \( k \) of airline \( a \); \( P_{C_{i}}^{n} \) the non-transfer passenger related charge from departing node \( i \); \( P_{C_{i}} \) the transfer passenger related charge from departing node \( i \); \( L_{Ct} \) the landing charge per plane ton at node \( i \); and \( MTOW \) is the maximum aircraft take-off weight (in tons).

Constraint set (9) allows each airline to choose its own fleet size and structure independently, per undirected leg, where \( U_{ka} \) and \( L_{ka} \) represent upper and lower bounds on the number of seats, respectively. Constraint set (10) restricts total aircraft movement to airport maximum runway capacity, where \( C_i \) represents the maximum aircraft movement capacity of node \( i \) (ACM/time.
period). Constraint set (11) ensures that all airports are connected to the network. Eq. (12) ensures that the airfare from airport $i$ to airport $j$ is the same in the opposite direction.

If $\pi_a$ is translated into a continuous and differentiable function, the mathematical program can be solved using standard optimization techniques, such as partial differentiation with respect to frequency, aircraft size and prices. The conjugate–gradient projection algorithm in Goldfarb (1969) can be used to solve the nonlinear objective function and linear constraints per airline. However, due to the nonlinearity of the objective function, a global optimal solution is not necessarily found.

4. Game theoretical model of explicit competition

The competitive model can be defined as a multi-airline, non-cooperative, two-stage game. A game is defined by a set of players, each of whom uses alternative strategies to maximize his or her own pay-off function, the value of which depends on each of the players simultaneous actions (Von Neuman and Morgenstern, 1943). The game described here consists of a set $A$ of airlines with a set of strategies which include a HS network, frequency of service, aircraft size and airfares. In the first stage, airlines simultaneously choose a HS network and whether or not to participate. Once HS connections are set, they cannot be changed within the game. Furthermore, an airline can choose not to connect a spoke through low frequency, but the spoke cannot be attached to a different hub once the first stage has been completed. In the second stage, after the network choice is revealed, each airline attempts to maximize profit through choice of service frequency, aircraft sizes and airfares, given the decisions of all other airlines, as shown in Fig. 1.

Airlines are the only players in the game for several reasons, one being the difficulty of generating an airport strategy set, because they are frequently government-owned. In addition, airport landing and passenger related charges which account for approximately 7% of airline total costs do not seem to affect the airlines’ network choice decision (Gillen and Lall, 1997). Airport capacity, for instance is a decision based on political and environmental considerations by relevant government bodies. Furthermore, the time lag between a decision and project execution are too long to be included within the scope of this model. In addition, passengers are only accounted for indirectly in this game through a utility function which is used to assess airfares, airline market share and airline pay-off functions.

Given explicit competition, the following algorithm is proposed to obtain sub-game perfect Nash equilibria for this non-cooperative game. The ILP defines an initial airline strategy. A specific sequence of airlines is chosen and the mathematical program is solved for the first airline in the sequence, then for the second airline and so on. Once an initial solution has been found for all airlines, the program starts again with the first airline in the sequence.

This iterative process continues until one of the following occurs:

1. A second stage sub-game Nash solution is found and an iteration fails to provide any

---

**Fig. 1.** The rules of the airline game.
improvement or solution change for the set of airlines. This could result in a monopoly, duopoly, oligopoly or competitive market depending on whether all airlines have positive frequencies and serve the market in some capacity, or one or more airlines take over the market forcing the remaining airlines to lose market share and profits, and/or to go bankrupt.

2. A quasi-equilibrium is achieved, either because the program cycles around two or more possible solutions without converging or the majority of decision variables achieve convergence, although a few remain divergent.

3. No equilibrium exists and no convergence is attained.

A Nash equilibrium can be defined as a strategy profile in which each player’s choice solution is as good a response to the other airlines’ choices as any other strategy available to that player (Kreps, 1990). Sufficient conditions for a Nash equilibrium to exist require each player strategy set to be bounded, convex and closed, the pay-off function to be concave with respect to players’ strategy sets and all pay-off functions to be continuous over the players’ strategy sets. In this case, each player’s strategy set is bounded, convex and closed, and pay-off functions are continuous, however the profit function is not concave. Therefore, a sub-game equilibrium does not necessarily exist. In addition, if an equilibrium is found, it cannot be guaranteed to be unique. An alternative approach is to run the entire process using several different initial strategy profiles and airline sequences.

The existence of a sub-game perfect equilibrium for the overall game can only be considered after equilibria for all sub-games have been evaluated. Sub-game perfect equilibria exist if airlines choosing to play cannot improve profits and no additional airlines choose to enter the market. Furthermore, it is assumed that airlines making losses will choose not to play.

5. Numerical illustration

Fig. 2 depicts an example consisting of a four-node network with two airlines using airports 1 and 2 and airports 3 and 4 as hubs, respectively. When airlines try to minimize distance by connecting hubs and spokes via the ILP in Section 3, direct competition on two routes results (1–3 and 2–4). The remaining routes compete using direct and indirect flights, with one or two hubs involved in an indirect flight (e.g. route 3–4 offered by airline 2 versus 3–1–2–4 offered by airline 1). Table 1 shows parameter values as computed in Adler and Berechman (2001). Demand is set at 1000 business and 1000 non-business travellers from each origin to each destination.

Each airline must solve the second stage of the game for 21 variables: six service levels, three aircraft sizes (one per non-directed) and 12 prices. When airline 1 begins the second-stage process, there is a monopolistic solution with airline 2 going bankrupt due to losses. When airline 2 begins the sequence, the opposite is true. The monopolist runs the complete network and captures the entire market with load factors of over 90% and aircraft sizes of 400 seats on flights between hubs and 300 seats otherwise. Other solutions using different starting positions were strictly worse for both airlines.

The game was re-played, this time tripling the demand, ceteris paribus. The result was a duopoly where the airline selected first had slightly higher profits. Both airlines ran complete HS networks with load factors lower than the previous game, around 80%. Prices were also lower as well as aircraft size, averaging 300 seats between hubs and 200 otherwise. The market was split evenly and
frequency rose, suggesting that travellers also benefit. This illustration may suggest that the air transport industry is a natural monopoly unless demand levels are sufficiently high.

6. Summary and conclusions

This paper suggests several operational research models that can be used together to solve the issue of which HS network in a deregulated, global market will ensure airline survival and profitability rates. These models can identify the most appropriate hubs for a new entrant or allocate additional hubs to existing carriers. Applications of interest include mega-carrier mergers and the subsequent allocation of appropriate hubs on various continents.

A two-stage non-cooperative n-player Nash game is proposed in the paper. In the first stage, all carriers select hubs and connections. An integer linear programming model is proposed to select the appropriate HS network, either by minimizing the great circle distance between hubs and spokes or the total number of passengers flying on more than one-leg journeys. The second stage of the game pits competing carrier networks against each other based on frequency, aircraft sizes and airfares. Decision variables are computed using a multinomial logit market share model implanted in a nonlinear mathematical program, which computes airline profits (or losses). An illustration shows the potential for monopoly power in low-demand networks and oligopoly potential given sufficient demand.

Acknowledgements

The Kreitman Foundation at Ben-Gurion University of the Negev provided generous post-doctoral funding for this research.

References

Adler, N., 1999. The choice of optimal multi-hub networks in a liberalized aviation market, Ph.D. dissertation, Faculty of Management, Tel Aviv University, Israel.


Ebery, J., Krishnamoorthy, M., Ernst, A., Boland, N., 2000. The capacitated multiple allocation hub location problem:

---

Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha ) (scale economy parameter in cost function)</td>
<td>1.2</td>
</tr>
<tr>
<td>( \beta ) (network economy parameter in cost function)</td>
<td>0.7</td>
</tr>
<tr>
<td>( \gamma_k \cdot \gamma_{in} ) (frequency elasticity of demand parameter in passenger utility)</td>
<td>0.73, 0.3</td>
</tr>
<tr>
<td>( \delta_k \cdot \delta_{in} ) (willingness-to-pay for direct flight parameter in passenger utility)</td>
<td>0.8, 0.17</td>
</tr>
<tr>
<td>( PC_t ) (passenger charge per transfer departure)</td>
<td>10</td>
</tr>
<tr>
<td>( PC_{int} ) (passenger charge per initial departure)</td>
<td>5</td>
</tr>
<tr>
<td>( LC ) (landing charge per MTOW)</td>
<td>10</td>
</tr>
<tr>
<td>( U ) (upper bound on plane size)</td>
<td>400</td>
</tr>
<tr>
<td>( L ) (lower bound on plane size)</td>
<td>150</td>
</tr>
<tr>
<td>( C_i ) (airport maximum runway capacity)</td>
<td>50</td>
</tr>
<tr>
<td>Demand per O–D per traveller type</td>
<td>1000</td>
</tr>
</tbody>
</table>

\( a \) Source: Adler and Berechman (2001).


