European Journal of Operational Research 132 (2001) 260–273

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/dsw

Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe

Nicole Adler [a,*], Boaz Golany [b]

[a] The Jerusalem School of Business Administration, Hebrew University, Mount Scopus, Jerusalem 91905, Israel

[b] Faculty of Industrial Engineering and Management, The Technion, Israel Institute of Technology, Haifa, Israel

Received 3 December 1998; accepted 28 February 2000

Abstract

US experience shows that deregulation of the airline industry leads to the formation of hub-and-spoke (HS) airline networks. Viewing potential HS networks as decision-making units, we use data envelopment analysis (DEA) to select the most efficient networks configurations from the many that are possible in the deregulated European Union airline market. To overcome the difficulties that DEA encounters when there is an excessive number of inputs or outputs, we employ principal component analysis (PCA) to aggregate certain, clustered data, whilst ensuring very similar results to those achieved under the original DEA model. The DEA–PCA formulation is then illustrated with real-world data gathered from the West European air transportation industry. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Data envelopment analysis; Principal component analysis; Air transportation networks

1. Background

The main focus of this paper is to evaluate which hub-and-spoke (HS) networks are desirable for an airline to choose, given the freedom to enter and exit markets, design networks and set prices and levels of service. We determine which network configuration, and consequently which airports, are likely to become the main gateways for both international and inter-regional routes using Western Europe as an example.

In 1978, the United States of America (USA) deregulated their air transportation industry. The European Union and the Far East are in the process of undertaking similar policies today. These new policies imply free market entry and unrestricted setting of airfare and level of service by airlines. The question then arises as to how this liberalisation will affect the airlines, airports and customers of the air transportation industry. This analysis can subsequently be applied to other markets that will undergo the process in the future, for example the Middle East and Eastern Europe.

* Corresponding author.

E-mail addresses: msnic@mscc.huji.ac.il (N. Adler), golany@ie.technion.ac.il (B. Golany).

PII: S0377-2217(00)00150-8

Furthermore, it is clear that the airline industry is presently undergoing international strategic alliances, which encourage regional (or continental) hubbing as part of a global airline network (see Oum and Yu, 1997). This research can be easily applied to a global network given the respective demand matrices.

Within the European Union (EU), national governments are increasingly required, as a result of EU law and financial necessities, to privatise home carriers and remove the subsidies that enabled the airlines to be loss-makers. It has been argued that these changes are likely to lead to the following results:

1. a decline in the number of air-lines operating;
2. a decline in aviation prices and an increase in service frequency;
3. employment of an easily recognisable HS network;

Bailey et al. (1985) wrote one of the first books on the subject of deregulating airlines, in which they argued that deregulation will cause major changes in the route system as a direct result of the pressures of open entry and exit. The competitive nature of the market and the newly acquired need for efficiency will require the carriers to utilise resources more effectively. An analysis of the US, on which the book is based, showed that carriers *seized the opportunity to redeploy their resources and have made major changes in their route networks. Almost all the carriers have emphasised connecting service by developing HS operations.* They argued that the cost structures of the carriers, the economies of aircraft size, plus the increase in load factors that result from a HS operation all led to an increase in "hubbing" after deregulation.

In the USA, many airlines went bankrupt or were merged with others in the aftermath of deregulation. The most famous case is probably Pan-Am, which ceased operations in December 1991 after failing to adapt to deregulation. Pan-Am began services again in 1996, though in a limited capacity within the USA only. Other examples include Eastern Airlines which was purchased by Texas Air in 1986, Midway which merged with Air Florida in 1985, Braniff and MarkAir which went bankrupt, and more. The surviving companies are utilising recognisable HS networks, with two or more hubs and very few direct links. Today, American Airlines openly declares that Chicago O'Hare, Dallas/Fort Worth and Miami International Airports provide the major hubs in its network. These three airports, which became hubs in the 1980s (i.e. after deregulation), are all integral elements of the HS network. Delta Airlines, a major American air transportation company, currently operates three major hubs in the USA, including Atlanta, Cincinnati and Salt Lake City. It also uses Frankfurt Airport in Germany to distribute traffic throughout Europe and beyond. Finally, it should be noted that various USA speciality and regional airlines have been established to soak up those travellers willing to purchase high-priced tickets in order to save time and additional legs, for example, Air South, AirTran, Casino Express, Frontier, etc.

Whilst such changes have already occurred in the USA, it is unclear how deregulation policies will affect the European market and to what extent. There is no guarantee that the effects of deregulation within Europe will necessarily be the same as those that were observed in the US. Button and Pitfield (1991) argue that for the following reasons, European deregulation will not be the same:

1. The European market is predominantly international, as opposed to the US. This will slow the effects of the changes, whereas in the US the laws were suddenly removed in a uniform manner.
2. The distances travelled within Europe are significantly less; approximately one-half the length of an average American trip (1450 km). Hubbing within Europe, which may lengthen the time required to travel, will consequently have a greater impact on the total time travelled. Thus different modes of travel will provide greater competition in Europe than in the States.
3. The US aviation industry is purely privatised, whereas the European aviation industry still has substantial public sector participation.
4. European policy-makers can use the knowledge that they have gained from the US changes to avoid some of the perceived pitfalls of

liberalisation. These include the problems of airline mergers and the market power that can be accessed through flight code sharing and domination of the computer reservation systems.

In this paper, we undertake a data envelopment analysis (DEA) approach towards identifying an appropriate network choice for an airline. DEA, since its introduction by Charnes et al. (1978), has become an increasingly popular technique for carrying out an examination of comparative efficiency when a simple monetary measure such as profit is not appropriate. The costs of the airport to the airlines are minor in comparison to total operating costs, since a mere 7% of total costs are paid directly to the airports in the form of landing and passenger related charges. Hence a reduction in airport charges is relatively inconsequential, particularly when considering that fuel and personnel charges generally amount for more than 70% of an airline's operating costs. Thus, in choosing a hub, an airport's quality levels are at least as interesting as their fees. Multiple inputs and outputs, e.g., number of aircraft movements, average delay time and average number of passengers per week, are weighted to provide a ratio of relative efficiency per potential network configuration. One unique element of this paper is the use of principal component analysis (PCA) in order to aggregate input data. An excessive number of inputs and/or outputs in a DEA model results in a large number of efficient units. It is consequently preferable to keep the ratio of the number of inputs and outputs to the number of decision-making units low and PCA can be used to aggregate inputs (or outputs) with minimum loss of information.

DEA is usually undertaken to compare decision-making units and to evaluate managerial strategies to improve the efficiency of those units that are not lying on the efficient frontier. In this paper we will be comparing networks simply to identify the most beneficial to an airline that is attempting to maximise profits and service quality whilst deterring new entrants. This is a relatively new area of DEA and little can be found in the literature. As an exception we point out the work of Desai and Storberck (1990) and Athanassopoulos and Storberck (1995) who used DEA to measure the spatial efficiency of siting decisions. For the purpose of the present research we will also need to "order" or rank the Pareto-efficient solutions, in order to chose one or two as potential networks for the deregulated airline. Here, we will employ Andersen and Petersen's procedure (1993) in order to rank the efficient decision-making units.

The paper outlines a micro-economic theory of airline behaviour in deregulated markets in Section 2. Section 3 describes the use of DEA and PCA in evaluating the network configurations with respect to levels of service and quality and develops a DEA–PCA model. Section 4 provides a specific solution to the question of Western European aviation. Section 5 includes a summary and conclusions. Appendix A specifies the abbreviated names of airports in the sample in Section 4.

2. Economic theory of an airline's behavior in deregulated markets

Under a deregulated aviation market regime, it has been shown that an airline, whose objectives are profit maximisation and market dominance, will choose a HS network over a fully connected (FC) configuration (see Berechman and Shy, 1996). Fig. 1 presents an example of a two-HS network whereby nodes 4 and 5 represent hubs.

The theory states that utilising network economies to advantage in order to prevent entry into the market, the airline will increase the frequency of flights between the hub(s) and spoke airports.

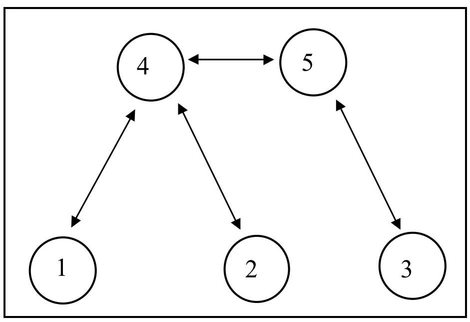

Fig. 1. 2-HS network configuration.

Consequently, the welfare of passengers flying directly to and from the hub increases whereas the welfare of passengers flying indirectly is not reduced (see for example Berechman and Shy, 1996; Morrison and Winston, 1986, 1995).

The economic model-based algorithm, given demand and parametric inputs, computes plane sizes and frequencies per network configuration such that the airline's profits are maximised. A network configuration includes the appointed hub(s) and specifies the connection of all other nodes with reference to the hub(s). Many network configurations can be used as input and the network that provides the airline with the maximum profit is considered the optimal network configuration. One of the by-products of choosing an optimal HS system for the airline is the impact this will have subsequently on the airport(s) chosen as hub(s). The liberalisation policies that will affect the airlines and encourage them to run a HS system will urge certain airports to increase their size in order to accommodate the additional traffic that will occur at the hubs chosen. The mathematical program enables us to evaluate the effect on the hub of a HS system. The landing and passenger-related charges paid by the airline become the revenue of the airport. The airport capacities are constraints in the model. Consequently the airport's tariffs and capacities will directly affect the solution outcome and can be used to develop a charges and capacity policy for airports, in addition to choosing the optimal HS network for the airlines. In the subsequent DEA, we can also analyse the effects of improving the level of service of a hub on the airline's choice of network configuration.

The mathematical model developed to evaluate the level of service, profits and load factors per network configuration can be found in detail in Adler and Berechman (2000). The information derived from the model will be used in the DEA. The model includes a nonlinear objective function and a set of nonlinear constraints, therefore any solution found may only be locally, rather than globally, optimal. Consequently, each network configuration is run from many different initial solutions in order to attempt to find the best solution.

In the West European application to be discussed in Section 4, we have selected four main airports to test as possible hubs within varied network configurations; London Heathrow, Schiphol Amsterdam, Charles De Gaulle Paris and Frankfurt. They were chosen because of their size and present demand patterns. Ten additional airports were chosen as possible secondary hubs, thus there were in total 40 hub combinations. The decision-making units (DMU) in this paper are potential network configurations, amongst which the airline must chose the most appropriate to its needs. In order to generate complete network combinations, two possible hub airport combinations are chosen and then all other airports in the sample are connected to either hub according to minimum distance. Subsequently, an additional 40 possible network combinations were computed with the same 2-hub combinations, but with a "balanced" network. In other words, eight airports were connected to each of the hubs according to minimum distance. Finally, each network combination needed to take into account the connections to America (via John F. Kennedy airport) and the Far East (via Narita, Tokyo). Four combinations were obtained in which each hub became the single international gateway, providing two combinations, and then each hub was connected to one of the international airports, providing a further two combinations. This resulted in a total of 320 network combinations or DMUs for the DEA.

Table 1 shows Schiphol as an international hub in four network configurations, with another four airports acting as intra-European hubs. The 20-node demand matrices for business and non-business travellers, which are endogenous to the model, are based on data drawn from aviation publications [1] and can be obtained from the authors upon request. The parameters of the model were computed econometrically (see Adler, 1999). The data is weekly and in dollars where relevant.

Table 1 shows that the Schiphol–Oslo network configuration is preferable to the other three

[1] ABC Guide and publications by the International Civil Aviation Authority and IATA.

Table 1
Summary table of four potential 2-HS networks

Secondary intra-European hub	Landing charges paid to airport	Passenger charges paid to airport	Operating costs	Station costs	Average load factor (%)	Potential profits to airline
Stockholm	17,877	7275	334,448	57,858	71	6,364,000
Barcelona	16,013	6606	336,225	52,032	77	6,413,000
Copenhagen	17,413	7901	336,959	58,231	73	−1,554,000
Oslo	8708	7665	554,865	37,663	74	30,517,000

configurations because of the substantially higher potential profits. Clearly, both sensitivity analyses and policy runs can then be evaluated in order to examine the effects of changes in airport charges and input parameters. The sensitivity analyses undertaken over the parameters of the model have shown the solutions to be stable.

The top three network combinations of the 320 evaluated in the mathematical model are given in Table 2.

The mathematical model has not accounted for the quality of the airports, with respect to average delay times, connecting capabilities, etc. which is indeed likely to affect the airline's decision as to which network configuration best suits themselves and their customers. Therefore, we will use the results of the mathematical model as part of the input of a DEA.

3. Data envelopment analysis

As has been shown in Section 2, airport charges are relatively inconsequential, therefore the airline may wish to include average delay times and congestion levels at airports, accessibility in terms of transportation to and from the airport and comfort of the terminals to the passengers in the decision-making process. These elements, amongst others, will affect the profitability of the airlines in the long run and the likelihood of passenger loyalty to the airline. Since the decision process requires a multi- criterion decision approach, the method chosen to account for all these elements was DEA.

Section 3.1 will describe the DEA methodology and the ranking technique chosen. An excess of inputs and outputs results in an excess of "efficient" decision-making units, therefore certain inputs were aggregated in order to facilitate the analysis, as described in Section 3.2. Section 3.3 develops a new DEA–PCA model formulation.

3.1. The DEA technique

DEA is a technique that measures the relative efficiency of decision-making units, in our case network configurations, with multiple inputs and outputs but with no obvious production function to aggregate the data in its entirety. Since it is unclear how to quantitatively combine the profits produced by the network with its requisite level of quality, the DEA technique can be utilised to evaluate the relative efficiency of each decision-making unit through the use of weighted averages. The efficiency is defined as the ratio of total weighted output to total weighted input as follows. By comparing n units with s outputs denoted by Y and r inputs denoted by X, the efficiency measure for unit a according to the variable-returns-to-

Table 2

Hub 1	Hub 2	Potential profit ($/week)	Balanced/unbalanced network	International hub
London Heathrow	Stockholm	118,387,880	Unbalanced	London Heathrow
London Heathrow	Oslo	115,501,000	Unbalanced	London Heathrow
London Heathrow	Manchester	112,483,312	Unbalanced	London Heathrow

scale model of Banker et al. (1984) (henceforth, BCC) can be defined as follows:

$$\underset{\lambda,s,\sigma}{\text{Min}} \quad T - e^{\mathrm{t}}s - e^{\mathrm{t}}\sigma, \tag{1}$$

$$\text{s.t.} \quad Y\lambda - s = Y^{a}, \tag{2}$$

$$-X\lambda - \sigma = -TX^{a}, \tag{3}$$

$$e^{\mathrm{t}}\lambda = 1, \tag{4}$$

$$\lambda, s, \sigma \geqslant 0, \tag{5}$$

where λ represents a vector of DMU weights chosen by the linear program, e^{t} a transposed vector of ones, σ and s vectors of input and output slacks, respectively, X^{a} and Y^{a} the input and output column vectors for DMU a, respectively, and T represents a constant.

The DEA results group the DMUs into two sets, those that are efficient and define the Pareto frontier and those that are inefficient. Golany and Yu (1995) developed a goal programming-discriminant function approach, which fine-tunes the seperation between efficient and inefficient units. However, we are attempting to choose the most appropriate network for an airline, thus we need to rank all the efficient networks. Several articles have been published in this field, including Cook and Kress (1990), Andersen and Petersen (1993), Golany and Roll (1993), Athanassopoulos and Thanassoulis (1995), Torgerson et al. (1996) and Friedman and Sinuany-Stern (1998). Andersen and Petersen (1993) achieve a full ranking by undertaking a DEA without assessing the DMU itself and then evaluating the extent to which the envelope frontier is extended when including the efficient DMU. However, the super-efficient methodology can give "specialised" DMUs an excessively high ranking. In our study, all networks connect the same 20 airports, thus specialisation is of less relevance and this methodology was used to rank the efficient network combinations.

We also considered the problem of defining a continuous Pareto frontier, since the networks are clearly discrete functions. The Free Disposal Hull model, see for example Tulkens (1993), solves the continuity problem, however it reduces the discriminatory powers of the solution, which in this case were problematic. Indeed, the poor initial results of the basic model were such that we needed to reduce the number of criteria in the DEA drastically. Consequently, we have combined PCA with DEA to reduce the inputs and outputs whilst minimising the loss of information.

3.2. Principal component analysis

PCA explains the variance structure of a matrix of data through linear combinations of variables, consequently reducing the data to a few principal components, which generally describe 80–90% of the variance in the data. If most of the population variance can be attributed to the first few components, then they can replace the original variables without much loss of information. As stated in Johnson and Wichern (1982), let the random vector $X = [X_1, X_2, \ldots, X_p]$ (in our case the original inputs chosen to be aggregated) have the covariance matrix V with eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p \geqslant 0$ and normalised eigenvectors $l_1, l_2, \ldots, l_p$. Consider the linear combinations, where the superscript t represents the transpose operator:

$$X_{\mathrm{PC}_i} = l_i^{\mathrm{t}}X = l_{1i}X_1 + l_{2i}X_2 + \cdots + l_{pi}X_p,$$
$$\mathrm{Var}(X_{\mathrm{PC}_i}) = l_i^{\mathrm{t}}Vl_i, \quad i = 1, 2, \ldots, p,$$
$$\mathrm{Cov}(X_{\mathrm{PC}_i},\ X_{\mathrm{PC}_k}) = l_i^{\mathrm{t}}Vl_k,$$
$$i = 1, 2, \ldots, p, \ \ k = 1, 2, \ldots, p.$$

The principal components are the uncorrelated linear combinations $X_{\mathrm{PC}_1}, X_{\mathrm{PC}_2}, \ldots, X_{\mathrm{PC}_p}$ ranked by their variances in descending order.

Generally inputs and outputs of a DEA need to be strictly positive (some papers have dealt with the issue of strictly zero data, see for example Thompson et al. (1993)). However, the results of a PCA can be negative. An affine transformation of data can be utilised with no change in the results when using the additive model (see Charnes et al., 1985) or without a change in the definition of efficient DMUs when using the BCC model (see Ali and Seiford, 1990). Indeed, Pastor (1996) proves that the BCC output-oriented model is input translation invariant and vice versa. Consequently all PC input data used subsequently in the DEA have been increased by the most negative value in the vector plus one when necessary, thus ensuring strictly positive data

$$\tilde{X}_{\mathrm{PC}_i} = X_{\mathrm{PC}} + b, \quad \text{where } b = \mathrm{Min}\{X_{\mathrm{PC}_i}\} + 1.$$

3.3. The DEA–PCA model formulation

In order to use PC scores instead of the original data, the DEA model needs to be transformed to take this into account. Indeed, if we were to use all the PCs we would not change the solution achieved under the original DEA formulation. When using both original inputs and outputs and some aggregated variables, separate $X = \{X_{\mathrm{o}}, X_L\}$ and $Y = \{Y_{\mathrm{o}}, Y_{L'}\}$, where $X_{\mathrm{o}}\,(Y_{\mathrm{o}})$ represents inputs (outputs) whose original values will be used in the subsequent DEA. $X_L\,(Y_{L'})$ represents inputs (outputs) whose values will be transformed through PCA. Let $L = \{l_{ij}\}$ be the matrix of the PCA linear coefficients of input data and let $L' = \{l'_{ij}\}$ be the matrix of the PCA linear coefficients of output data. Now, $X_{\mathrm{PC}} = LX_L$ and $Y_{\mathrm{PC}} = L'Y_{L'}$ are weighted sums of the corresponding original data, X_L and $Y_{L'}$. We can replace constraints (2) and (3) with the following:

$$\text{s.t.} \quad Y_{\mathrm{o}}\lambda - s_{\mathrm{o}} = Y_{\mathrm{o}}^a, \tag{2a}$$

$$Y_{\mathrm{PC}}\lambda - s_{\mathrm{PC}} = Y_{\mathrm{PC}}^a, \tag{2b}$$

$$-X_{\mathrm{o}}\lambda - \sigma_{\mathrm{o}} = -TX_{\mathrm{o}}^a, \tag{3a}$$

$$-X_{\mathrm{PC}}\lambda - \sigma_{\mathrm{PC}} = -TX_{\mathrm{PC}}^a. \tag{3b}$$

The slacks in the objective function are weighted to counter the transformation and ensure an equivalent problem/solution to that of the original linear program

$$\begin{aligned}\mathrm{Min}\ & T - e^{\mathrm{T}}s_{\mathrm{o}} - e^{\mathrm{T}}L'^{-1}(s_{\mathrm{PC}}^{+} - s_{\mathrm{PC}}^{-}) - e^{\mathrm{T}}\sigma_{\mathrm{o}} \\ & - e^{\mathrm{T}}L^{-1}(\sigma_{\mathrm{PC}}^{+} - \sigma_{\mathrm{PC}}^{-}),\end{aligned} \tag{1a}$$

whereby L^{-1} (L'^{-1}) represents the inverse matrix of input (output) weights attained in the PCA. In addition, two more constraints must be included in the formulation to maintain equivalence with Eqs. (1)–(5):

$$L'^{-1}(s_{\mathrm{PC}}^{+} - s_{\mathrm{PC}}^{-}) \geqslant 0, \tag{6}$$

$$L^{-1}(\sigma_{\mathrm{PC}}^{+} - \sigma_{\mathrm{PC}}^{-}) \geqslant 0. \tag{7}$$

Eqs. (6) and (7) relate to the "new" slacks relevant only to the PC data. These slacks do not need to be non-negative, hence the introduction of two slack variables. The new formulation is exactly equivalent to the original model, provided the PCs explain 100% of the variance in the original input and output matrices. Clearly, if we use less than the full information we will lose some of the explanatory powers of the full data and may not achieve the exact results of the original DEA. The removal of some constraints in the form of Eqs. (2a) or (3a) for PCs with small explanatory value will alter the feasible area. If these constraints were redundant in the solution, then the solution will not change.

We illustrate this point via a small example showing that PCs explaining 100% of the variance attain the same results as the original data. Furthermore, using slightly less information, whereby the PC values explain 80–90% of the variance, we do not change the solution dramatically. The example is drawn from the banking example in the WDEA package of Warwick University. The DEA compares 10 different banks with two inputs and four outputs. Using the additive model and original data, eight of the DMUs are efficient. A PCA was run on the four outputs with the results as shown in Table 3.

The first two PCs explain 88% of the variance of the data matrix and the first three explain 96.2%. The DEA–PCA model was run with all four PCs and achieved the exact results of the original model. The model was then run with two and three PCs and the results can be found in Table 4.

The results are very similar, even when full information was not utilised. Specifically, the number of efficient units remained exactly the same when removing one or two PC equations. The only difference can be found with respect to Bank 8, which was efficient in the full-information model but became inefficient in the other two formulations. It should be noted that Bank 8 was not in the efficient reference set of either Banks 1 or 4

Table 3
PCA results of the 10 banks example

Eigenanalysis of the correlation matrix				
Eigenvalue	2.242	1.277	0.328	0.152
Proportion	0.560	0.319	0.082	0.038
Cumulative	0.560	0.880	0.962	1.000
Variable	PC1	PC2	PC3	PC4
A	−0.549	0.430	−0.250	−0.672
B	0.020	0.863	0.221	0.454
C	−0.574	−0.21	0.791	0.040
D	−0.608	−0.162	−0.514	0.583

Table 4
DEA input and output data and results with adapted BCC model

Banks	Original data						Results of DEA		
	Input 1	Input 2	Output 1	Output 2	Output 3	Output 4	Original data and 4 PCs	3PCs	2PCs
1	170	70	45	6	11	5	90.96	90.96	90.96
2	155	85	53	11	9	7	100	100	100
3	183	92	48	23	4	2	100	100	100
4	143	62	28	7	3	1.8	83.32	83.32	83.82
5	202	88	60	17	5	3	100	100	100
6	117	49	35	12	4	1.7	100	100	100
7	143	44	27	8	3	1	100	100	100
8	155	61	33	17	6	2	100	93.37	93.97
9	139	53	42	8	7	3	100	100	100
10	183	63	52	12	15	4	100	100	100

when the original data were applied. Indeed, Bank 8 is a relative specialist in output 2 production.

4. Application of the DEA–PCA model to the Western European aviation industry

Section 4.1 discusses the inputs, outputs and data required to solve the real world, network choice decision with respect to Western Europe. Section 4.2 discusses the results of the base run of 320 potential networks and Section 4.3 discusses the robustness of the solution through the use of scenario analyses.

4.1. Inputs and outputs

The DMUs in this study are *potential* network configurations, amongst which the airline must choose the most appropriate to its needs. Refer to Section 2 for greater details and Golany and Roll (1989) for a discussion on how to choose appropriate DMUs and inputs and outputs in a DEA. The DEA is based on information attained from the mathematical model and alternative sources. The airlines revenue and costs are computed by the mathematical model, as are the number of passengers served, the frequency, the percentage of flights between the hubs and the average load factor. A major problem for the majority of DEA studies is a lack of relevant data, thus the use of a model to provide data is very advantageous. This is particularly true when considering future entities for which no data have been collated, for example new technological breakthroughs in the scientific or military fields. For greater details of the mathematical model, see Adler and Berechman (2000). Additional elements, that take into account the "attractiveness" of the network, will be based on both objective and subjective data. The inputs from the mathematical model may include:

- total landing related charges paid to airports in the network, based on landing, noise and environmental charges;
- total passenger charges paid to the airports for both transfer and non-transfer departure passengers carried;
- airline station costs, including ground staff salaries;
- airline operating costs, including fuel and crew salaries.

All of these costs are relevant to an airline's choice of network configuration and are different for each airport. The level of salaries and conditions of employment in each country are individual, as are the level of airport charges. These are considered input criteria since the airline wishes to minimise its costs but must generate the costs to produce flights.

Additional service quality inputs may include:

- Hub shopping facilities:
  - total number of shops;
  - number of duty-free shops;
  - number of tax-paid shops;
  - number of restaurants;
  - number of other concessions.
- Surface transportation systems from/to hubs:
  - local road links;
  - motorway links;
  - underground links;
  - rail links;
  - international train links;
  - high-speed train links;
  - city-centre bus links;
  - local public bus links;
  - long distance bus links;
  - direct sea connections;
  - distance to city centre.
- Comfort of hub airports:
  - overall passenger convenience;
  - signposting in airports;
  - ease of making connections;
  - passport inspections;
  - customs inspection;
  - speed of baggage delivery;
  - availability of baggage carts;
  - comfortable waiting areas/lounges;
  - special services for overseas travellers;
  - on-time departure.

All these elements may affect a passenger's choice of airports, thus it is also of relevance to an airline in choosing an optimal network configuration. In the BCC output-oriented DEA model used here, the inputs are given. The airport facilities are non-controllable variables to the network, since the airline cannot directly affect their existence.

The outputs from the mathematical model, which may be of interest in a DEA include:

- profit;
- revenue;
- total number of passengers carried through the system;
- average load factor.

These outputs are the result of an airlines choice of network and levels of decision variables. Additional outputs, which were not included in the mathematical model, but may help to reflect the service quality of a network include:

- average delay in minutes at each of the two hub airports;
- minimum connecting times at hub airports.

Indeed, the higher the frequency and the greater the demand channelled through the hub airports, the more problematic these output variables become. They clearly represent undesirable outputs of the model.

The data required for the DEA is based on several sources. The information used to compute demand matrices, various costs and econometrically computed parameters can be found in Adler (1999). The data describing the comfort of hub airports is based on the International Air Transport Association (IATA) Airport Monitor Survey for 1993, which canvassed 30,000 passengers at 34 Airports in Europe and North America. The data drawn from this survey cover the four main hub airports; Schiphol Amsterdam, Charles de Gaulle Paris, Frankfurt and London Heathrow. The remaining data was obtained from the Airports Council International (ACI) database, which covered the year 1995. The information provided was received from the airport operators in each of the relevant countries and includes all airports in the sample.

The DEA was undertaken over 320 network configurations using 11 inputs and outputs, as follows. The input service quality data is in the form of PCs rather than original data (see Table 5).

We have used the BCC output-oriented, DEA–PCA model (for justification, see Section 3.3), whereby output is maximised given inputs. The average delays and minimum connecting times are used in a reciprocal form since they are undesirable outputs of the model and ought to be minimised.

4.2. Base-run results

The results consisted of a large number of efficient DMUs, approximately two thirds of the 320 network combinations, consequently the ranking of the efficient units is very important. The summary results of the base-run can be found in Adler (1999) and the top six DMUs are given in Table 6.

It should be noted that some problems were encountered due to infeasibility (see for example Seiford and Zhu, 1999). Inputs and outputs were adjusted slightly until such problems were eliminated.

Clearly three of the four major hubs dominate the highest rankings, which suggests that the advantage achieved by LHR in the mathematical model (due to the heavily biased demand matrices) has been overcome when taking into account quality levels. In addition, the unbalanced, minimum distance networks are calculated as relatively preferable, as are single international hubs. The preferable international hub was in all cases chosen from the Group 1 hub set. The preferable secondary, intra-European hubs would appear to include Vienna, Stockholm and Munich, with the highest ranking achieved by an Amsterdam–Vienna combination. However, whilst service quality may help to guarantee passenger loyalty in the long run, in the short term the top three combinations generate substantially lower profits for the airline. It would perhaps be more useful to take an existing airlines major hub and use this analysis to evaluate which secondary hub would improve both airline profitability and service quality as examined in Section 4.3.

Table 5

Inputs	Outputs
Mathematical model	
	Profit
	Average load factor
Additional service data	
Transportation index (4 PCs)	1/Average delay in minutes for the major hub
Shopping index (2 PCs)	1/Minimum connection time
Hub comfort index (1 PC)	

4.3. Scenario analyses

We have attempted to evaluate the sensitivity of the results by changing certain parameters, such as the existence of a high-speed rail link or a reduction in station costs, including landing or passenger related charges. Reducing the average delay in minutes at the major hubs by 20% had no effect on the results, in that the six highest ranked network configurations did not change.

Station costs include ground staff salaries and airport charges, which represent approximately 21% of the total costs of European airlines. A scenario analysis was undertaken and summary results are presented in Table 7.

Frankfurt's station costs were reduced by 20% in the FRA–MUC network combination and the

Table 6

Hub 1	Hub 2	Profits ('000 $/week)	Radial rank	Balanced/unbalanced	International hub
AMS	VIE	33,462	139.98	Unbalanced	AMS
CDG	ARN	59,266	133.33	Unbalanced	CDG
FRA	MUC	86,755	128.47	Unbalanced	FRA
AMS	FBU	138,004	124.87	Unbalanced	AMS
FRA	ARN	127,690	122.42	Unbalanced	FRA
AMS	MAD	112,384	121.95	Unbalanced	AMS

Table 7
Effects of changes in station costs

Base run			Decrease Frankfurt station costs by 20% for FRA–MUC network			Decrease Charles de Gaulle station costs by 20% for CDG–ARN network		
AMS	VIE	139.98	*FRA*	*MUC*	*143.45*	AMS	VIE	139.98
CDG	ARN	133.33	AMS	VIE	139.98	*CDG*	*ARN*	*133.33*
FRA	MUC	128.47	CDG	ARN	133.33	FRA	MUC	128.47
AMS	FBU	124.87	AMS	FBU	124.87	AMS	FBU	124.87
FRA	ARN	122.42	FRA	ARN	122.42	FRA	ARN	122.42
AMS	MAD	121.95	AMS	MAD	121.95	AMS	MAD	121.95

profits increased accordingly. The network configurations ranking moved from third to first position. However, CDGs equivalent reduction in station costs did not enable the CDG–ARN network to move from second to first position. Consequently, the Parisian landing and passenger related charges would have to decrease by more than 20% in order to attain first ranked position. It should also be noted that Frankfurt airports charges are relatively high, which may explain its rank improvement.

We also evaluated whether the introduction of a high-speed rail link, which exists today at Schiphol and Charles de Gaulle, would improve the combined financial-quality index of a LHR combination. A new, surface transportation PCA was run, with the results shown in Table 8.

The DEA–PCA model, with the new transportation principal components, ranked the top eight networks as shown in Table 9.

Whilst the introduction of a high-speed rail link at LHR did not help the highest ranked LHR network combination change its relative position (LHR–ARN is ranked in eighth position in the basic run too), this analysis has affected Frankfurt's ranking. Frankfurt is now the only airport without the high-speed link and its ranks have suffered as a result, with FRA–MUC moving from third to fourth position and FRA–ARN moving from fifth to seventh position.

We have established that LHR-based networks, when accounting for both financial and quality issues, are weak contenders to become a major hub in liberalised Western Europe. Thus, we have run a DEA on the 80 different combinations that connect LHR with a second potential hub in Europe, in order to evaluate which two-hub combination is most preferable. Given the same input–output data specified in the base-run, all networks lie on the new production possibility frontier, however, a few were considered super-efficient (see Table 10).

All super-efficient networks were unbalanced and only one of the top six DMUs split the international routes, with Milan connected to NRT and LHR connected to JFK. It would appear preferable for an airline utilising LHR to choose a

Table 8

Eigenvalue	1.7982	1.5251	1.2166	0.9185	0.7779	0.5758	0.1880
Proportion	0.257	0.218	0.174	0.131	0.111	0.082	0.027
Cumulative	0.257	0.475	0.649	0.780	0.891	0.973	1.000
Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Motorway	0.029	−0.036	−0.871	0.188	−0.031	−0.080	−0.444
Train	−0.556	−0.328	−0.160	0.047	−0.026	−0.605	0.435
High-speed train	0.136	0.458	−0.119	−0.748	0.095	−0.433	−0.040
City-centre bus	0.472	0.172	0.242	0.511	−0.114	−0.633	−0.120
Long-distance bus	−0.502	0.270	0.215	−0.008	−0.665	−0.024	−0.431
Direct sea connection	0.436	−0.311	−0.168	−0.228	−0.722	0.060	0.328
Distance to city centre (km)	−0.078	0.694	−0.262	0.300	−0.113	0.189	0.552

Table 9

Hub 1	Hub 2	Profits ('000 $/week)	Radial rank	Balanced/unbalanced	International hub
AMS	VIE	33,462	139.98	Unbalanced	AMS
CDG	ARN	59,266	133.33	Unbalanced	CDG
CDG	CPH	118,217	129.87	Unbalanced	CDG
FRA	MUC	86,755	128.47	Unbalanced	FRA
AMS	FBU	138,004	124.87	Unbalanced	AMS
AMS	MAD	112,384	121.95	Unbalanced	AMS
FRA	ARN	127,690	118.59	Unbalanced	FRA
LHR	ARN	184,657	115.99	Unbalanced	LHR

Table 10

Hub 2	Balanced/unbalanced	International hub	Profits ('000 $/week)	Radial rank
FBU	Unbalanced	LHR	181,304	118.54
ARN	Unbalanced	LHR	184,657	115.99
MAD	Unbalanced	LHR	170,333	115.44
MAN	Unbalanced	LHR	178,456	113.69
LIN	Unbalanced	Split	162,447	104.90
BCN	Unbalanced	LHR	161,408	104.20

secondary intra-European hub amongst those that lie on the periphery of Western Europe, such as the Scandinavian cities of Oslo and Stockholm.

5. Summary and conclusions

In choosing an optimal hub-and-spoke network under deregulation, an airline is likely to be interested in encouraging passenger loyalty by providing a high quality service. However, a production function accounting for both monetary and quality objectives in the airline industry is a priori unclear in its construction. The non-parametric technique, data envelopment analysis, analyses the relative efficiency of various homogeneous alternatives with multiple inputs and outputs, without the need for a production-function specification. Consequently, we chose this approach to evaluate many, potential, two-HS networks, based on monetary and load factor variables and the more subjective criteria of airport comfort. The inputs of the model include surface access to and from the hub airports, shopping and general comfort indices. The output criteria include profitability, average load factors, average delays and minimum connecting times. Some of this data were attained from the results of a mathematical program in which 320 potential network combinations were evaluated.

Principal component analysis is concerned with explaining the variance–covariance structure through a few linear combinations of the original variables. Its general objectives are data reduction and interpretation. Frequently principal components serve as intermediate steps in a larger analysis and become inputs in a subsequent multiple regression or cluster analysis. In this research we have used principal components to produce a reduced number of inputs for a subsequent data envelopment analysis.

A new DEA–PCA formulation was developed such that the model provides the same results were it to use the complete original data. Clearly, by reducing the information fed into the DEA, which occurs when the PCs explain less than 100% of the original variance, the DEA–PCA results will be minimally different. Decision-making units considered efficient in the original model may become inefficient in the re-formulation. If this occurs, it is most likely to happen to specialist units. For this reason, principal component analysis can also be used prior to the DEA analysis to identify outliers, where necessary. Some may argue that the use of principal components as opposed to original data would cause a more opaque result in the subsequent analysis. It may be considered preferable to ask the decision-makers to choose those inputs or outputs they consider the most important. This process could be undertaken alongside the principle component analysis. Furthermore, reducing the data through principal component analysis may allow us to take account of more information than would otherwise be the case.

The variable returns-to-scale, output-oriented, DEA–PCA model was used to evaluate the West European aviation industry. Sets of original input data were replaced by principal components to reduce the total number of input criteria and minimise the loss of information, thus improving the discriminatory power of the data envelopment analysis model. The results of the analysis include two groups of networks, those that are considered relatively efficient and those that are not. We used the Andersen and Petersen super-efficiency ranking technique in order to rank the group of efficient network combinations.

The top six super-efficient hub network combinations include Charles de Gaulle, Frankfurt and Schiphol. This is the opposite result to the mathematical model based on financial data alone, in which London Heathrow hub combinations were consistently preferable. The results of the mathematical model have been used as output in the DEA–PCA model, therefore the additional quality criteria have affected the decision process. The two models produce similar results with respect to the high ranking of unbalanced, minimum distance, two-HS connections and the use of one international hub and one intra-European hub, spatially placed on the outer edges of Western Europe. However, the results specified here suggest that London Heathrow will need to improve its quality factors in order for it to be considered preferable when accounting for more than simple airline profit maximisation. The results proved stable under various scenario analyses including the addition of a high-speed rail link between Heathrow and London city.

The model developed in this research can be used by both airlines, considering future network developments such as relocating or adding hubs, and airports, looking to improve their chances in being chosen as future hubs. The analysis can identify areas in which an airport can help to improve its ranking, for example in delay times or charges. This is becoming an issue as airports, as well as airlines, are slowly being privatised. Clearly this model can be reapplied to Western Europe, taking account of all airports in the region, other continents, such as the Far East, and global networks that are becoming more important through the use of code-sharing and international computer reservation systems.

Acknowledgements

This paper is partially based on Chapters 8 and 9 of Nicole Adler's Ph.D. thesis. I would like to thank the American Express Fellowship Endowment Fund for Ph.D. Studies at Tel Aviv University for financing my research.

Appendix A. Airport abbreviations

The airports that are considered in the mathematical models and results include:

European airports	
AMS	Schiphol, Amsterdam
ARN	Stockholm
ATH	Athens
BCN	Barcelona
BRU	Brussels
CDG	Charles de Gaulle, Paris
CPH	Copenhagen
FBU	Oslo
FCO	Fumicino, Rome
FRA	Frankfurt
GVA	Geneva
LHR	Heathrow, London
LIN	Milan
MAD	Madrid
MAN	Manchester
MUC	Munich
VIE	Vienna
ZRH	Zurich
Non-European airports	
JFK	John F. Kennedy, New York
NRT	Narita, Tokyo

References

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

Adler, N., Berechman, J., 2000. Choice of an optimal network in a liberalized aviation market. Transportation Research A, forthcoming.

Ali, A.I., Seiford, L.M., 1990. Translation invariance in data envelopment analysis. Operations Research Letters 9, 403–405.

Andersen, P., Petersen, N.C., 1993. A procedure for ranking efficient units in data envelopment analysis. Management Science 39 (10), 1261–1294.

Athanassopoulos, A.D., Storberck, J.S., 1995. Nonparametric models for spatial efficiency. Journal of Productivity Analysis 6 (3), 225–245.

Athanassopoulos, A.D., Thanassoulis, E., 1995. Separating market efficiency from profitability and its implications for planning. Journal of the Operations Research Society 46 (1), 20–34.

Bailey, E.E., Graham, D.R., Kaplan, D.P., 1985. Deregulating the Airlines. MIT Press, Cambridge, MA.

Banker, R.D., Charnes, A., Cooper, W.W., 1984. Some models for estimating technical and scale efficiencies in data envelopment analysis. Management Science 30 (9), 1078–1092.

Berechman, J., Shy, O., 1996. Airline deregulation and the choice of networks. In: Nijkamp, P. (Ed.), Advances in Spatial Equilibrium. Springer, Berlin, pp.138–156.

Button, K., Pitfield, D., 1991. Transport Deregulation. Macmillan Educational, London.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision-making units. European Journal of Operational Research 2, 429–444.

Charnes, A., Cooper, W.W., Golany, B., Seiford, L., Stutz, J., 1985. Foundations of data envelopment analysis for Pareto–Koopmans efficient empirical production functions. Journal of Econometrics 30, 91–107.

Cook, W.D., Kress, M., 1990. A data envelopment model for aggregating preference rankings. Management Science 36 (11), 1302–1310.

Desai, A.D., Storberck, J.S., 1990. A data envelopment analysis for spatial efficiency. Computers, Environment and Urban Systems 14, 145–156.

Friedman, L., Sinuany-Stern, Z., 1998. Scaling units via the canonical correlation analysis in the DEA context. European Journal of Operations Research, forthcoming.

Golany, B., Roll, Y., 1989. An application procedure for DEA. Omega 17 (3), 237–250.

Golany, B., Roll, Y., 1993. Alternate methods for treating factor weights in data envelopment analysis. Omega 21 (1), 99–109.

Golany, B., Yu, G., 1995. A Goal programming-discriminant function approach to the estimation of an empirical production function based on DEA results. Journal of Productivity Analysis 6 (2), 171–186.

Johnson, R.A., Wichern, D.W., 1982. Applied Multivariate Analysis. Prentice-Hall, Englewood Cliffs, NJ.

Morrison, S., Winston, C., 1986. The Economic Effects of Airline Deregulation. The Brookings Institution, Washington, DC.

Morrison, S., Winston, C., 1995. The Evolution of the Airline Industry. The Brookings Institution, Washington, DC.

Oum, T.H., Yu, C., 1997. Winning Airlines: Productivity and Cost Competitiveness of the World's Major Airlines. Kluwer Academic Publishers, Dordrecht, MA.

Pastor, J., 1996. Translation invariance in data envelopment analysis: A generalization. Annals of Operations Research 66, 93–102.

Seiford, L.M., Zhu, J., 1999. Infeasibility of super-efficiency DEA models. INFOR 37 (2), 174–187.

Thompson, R.G., Dharmapala, P.S., Thrall, R.M., 1993. Importance for DEA of zeros in data, multipliers and solutions. The Journal of Productivity Analysis 4, 379–390.

Torgerson, A.M., Forsund, F.R., Kittelesen, S.A.C., 1996. Slack-adjusted efficiency measures and ranking of efficient units. The Journal of Productivity Analysis 7, 379–398.

Tulkens, H., 1993. On FDH efficiency analysis: some methodological issues and applications to retail banking courts and urban transit. The Journal of Productivity Analysis 4, 183–210.