Benchmarking Airports from a Managerial Perspective

Nicole Adler[a,*], Vanessa Liebert[b] and Ekaterina Yazhemsky[c]

[a]Hebrew University of Jerusalem, Israel and Northwestern University, United States. e-mail: msnic@huji.ac.il.

[b]Jacobs University Bremen, Germany. e-mail: v.liebert@jacobs-university.de.

[c]Hebrew University of Jerusalem, Israel. e-mail: katy1@mscc.huji.ac.il.

Abstract

Benchmarking airports is currently popular both in the academic literature and in practice but has proved rather problematic due to the heterogeneity inherent in any reasonably sized dataset. Most studies either treat the airport production technology as a black box, or they separate the terminal and airside activities, assessing them individually. In this article we analyze airports as a single unit due to the direct complementarities, avoiding the artificial separation of inputs between the terminal and airside, and opening the black box through the use of network data envelopment analysis (DEA). To further improve the benchmarking process, we identify appropriate peers for 43 European airports over 10 years through a dynamic clustering mechanism according to pre-defined characteristics, and we restrict the integer linear program with respect to potential reductions in capital inputs. Compared to basic DEA models, the results of the network DEA structure provide more meaningful benchmarks with comparable peer units and target values that are achievable in the medium term. By identifying each airport's individual reference set, unique airport outliers influence the performance measurement less severely than occurs under basic DEA. In addition, the formulation is shown to be suitable in assessing different strategies for evaluating aeronautical and commercial activities, not only separately but in combination.

Keywords: Air transport, DEA, Business policy

1. Introduction

According to the Princeton dictionary, an airport is defined as "*an airfield equipped with control tower and hangars as well as accommodations for passengers and cargo*". Airports can be defined as an important basic infrastructure to a society in which aviation is one of the drivers of a modern economy. An alternative approach defines an airport as a private production system in which society maximizes social welfare by encouraging airport management to maximise profits, and at the same time, considering consumer surplus via some form of airport regulation if deemed necessary. Consequently, it is unclear whether airports should be considered as a not-for-profit, public good, as is the general approach in the United States, or as a private enterprise maximizing shareholder value. Since it would appear to be true that large regions of the world are gradually adopting the privatized form (Zhang and Zhang [1]) and that independent authorities running public airports in the United States do not behave differently to their private counterparts with respect to productivity (Oum et al. [2]), in this article we develop an airport benchmarking methodology from an airport manager's perspective in which we assume that the airport intends to maximize revenues or minimize costs.

Liebert and Niemeier [3] review airport benchmarking studies applied to a diverse range of activities using various methodologies. The most popular methods include price index total factor productivity (Hooper and Hensher [4]; Oum and Yu [5]; Vasigh and Gorjidooz [6], parametric stochastic frontier analysis (Pels et al. [7]; Oum et al. [8]) and non-parametric data envelopment analysis (DEA). DEA has been used to compare the performance of airports within national boundaries, U.S. (Gillen and Lall [9]; Sarkis [10]), U.K. (Parker [11]), Spain (Martín and Román [12]; Murillo-Melchor [13]), Australia (Abbott and Wu [14]), Taiwan (Yu [15]), Portugal (Barros and Sampaio [16]) as well as airports around the world (Adler and Berechman [17]; Lin and Hong [18]). It is rather difficult to draw general inferences since many of these articles arrive at directly opposing conclusions. For example, Murillo-Melchor [13] show that Spanish airports in their dataset suffer from decreasing returns to scale whereas Martín et al. [12] concluded increasing returns to scale for the same set of airports. Abbott and Wu [14] found most Australian airports enjoy increasing returns to scale, Pels et al. [7] argue that European airports operate under constant returns to scale in air traffic movements and increasing returns to scale on the terminal side and Lin and Hong [18] argue that most airports are not operating at an optimal scale. Graham and Holvad [20] and Abbott and Wu [14] argue that Australian airports are more efficient than their European counterparts, Lin and Hong [18] argue that the U.S. and European airports are more efficient than their Asian and Australian counterparts and Pels et al. [7] conclude that widespread European airport inefficiency is not specific to a country or region. Consequently, Morrison [21] has called for a balanced approach and dialogue between airport managers and researchers.

The majority of previous studies have treated airport technology as a single production process, avoiding the complexity inherent in airport systems. Gillen and Lall [9] and Pels et al. [7] were the first to argue that the airport could be analyzed as two separate decision-making processes, one serving airside activities and the other serving landside production. The approach developed in this research connects the two sides of the production function, while at the same time opening the black box via network DEA (Färe [22]). We argue that a single black box approach would be insufficient to capture the rich picture underlying this approximation, as demonstrated in Fig. 1. Since the liberalization of the aviation industry in Europe in the late eighties, airports have focused on both aeronautical and commercial landside activities. The network DEA approach recognises the fact that generalized and fixed costs connected to the two sets of activities can only be split in an artificial manner and that while aeronautical revenues draw from passengers, cargo and air traffic movements, the non-aeronautical revenue is more closely tied to passenger throughput. Although airports may have limited control over traffic volume, non-aeronautical revenues drawn from non-airport related activities, such as airport cities, are indeed within the purview of airport management. As argued in Oum et al. [23], the omission of outputs such as commercial services is likely to bias efficiency results as it underestimates the productivity of airports whose managers focus on generating additional revenue sources. Many airports attempt to increase revenues from non-aeronautical sources which are not directly related to aviation activities in order to cross-subsidize aviation charges in turn attracting more airlines and passengers to their airport (Zhang and Zhang [24]). Revenue source diversification that exploits demand complementarities across aeronautical and non-aeronautical services appears to improve airport productive efficiency (Oum et al. [2]). We would argue that it is more reasonable to analyze airports as a single unit because of the direct complementarities, thus avoid the need to separate inputs between the terminal and airside. In general, the airport technology may be defined as a network that consists of multi-production processes and stages as described in Fig. 1. Consequently, in this paper we develop a network DEA modelling approach in order to measure the relative cost and revenue performance of airports with respect to aeronautical and commercial activities simultaneously, whereby activities are connected via passengers as the common intermediate product.

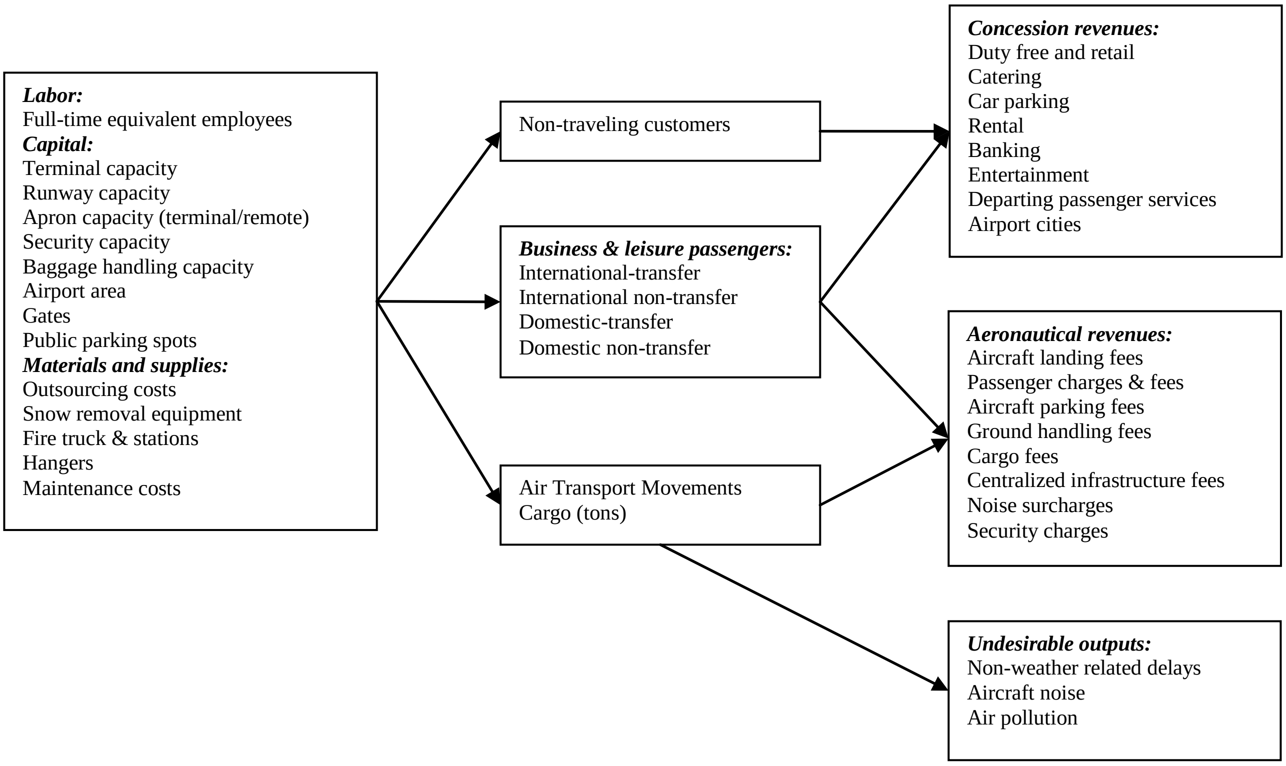


Fig. 1: Airport network technology

Another issue that appears in the airport benchmarking literature is the problem of comparability. A base assumption within the DEA context that has been questioned is the homogeneity of the decision-making unit under analysis and the appropriateness of this assumption with respect to airports (Morrison [21]). The aim of the formulation presented here is to broach the direct question of airport benchmarking in light of the reasonable level of heterogeneity found in a multiple airport study. This is necessary to generate sufficient data points for purposes of analysis. In order to ensure comparability, we apply a dynamic clustering approach (Golany and Thore [25]) using integer linear programming which forms reference sets based on similar mixes of inputs or outputs and intermediate products. Certain inputs may be beyond managerial control in the short to medium term yet affect airport performance (Adler and Berechman [17]). In general, capital is frequently treated as a non-discretionary variable over which airport management has little to no control (Banker and Morey [26]). In this research, capital has been defined in terms of declared runway and terminal capacity which are agreed upon within a multiple stakeholder setting and result in a number that accounts for the airport system configuration. For example, some airports consist of a reasonably large number of runways however, for reasons of weather and/or geographical layout, only a smaller portion may be in use in at a given time. Declared capacity takes this in to account as compared to simply counting the number of runways. On the terminal side, check-in counters, security and passport control, and gates, all together produce a throughput level per hour that is otherwise assumed to be linear within a standard DEA framework. We argue that the capacity of an airport, as a proxy for capital, may be adjusted to a certain extent in the medium term, hence the model restrictions permit terminal and runway declared capacity to change up to a pre-determined level. Pure capital investment is not an appropriate measure even within a specific country because the accounting processes differ, rendering the information incompatible. Finally, principal component analysis (PCA) combined with DEA (Adler and Golany [27]; Adler and Yazhemsky [28]) is applied in the input-oriented model in order to reduce the curse of dimensionality and the resulting bias, reducing the set of peer airports from 53% to 38% in the current application.

The aim of this research is to develop a comprehensive methodology tailored to airport benchmarking from a managerial perspective. A comparison with basic DEA results demonstrates that the additional restrictions in the network PCA-DEA dynamic clustering formulation lead to more reasonable peer comparisons, permitting an analysis of strategies which could potentially be adopted over the short and medium term for planning purposes. The model in this research allows airport managers to include their industry knowledge in the form of limitations on airport size, the operating conditions and the restricted variability of capacity, which is encapsulated in the dynamic clustering approach. Given that our data set does not include all airports in Western Europe, the numerical analysis presented is purely a case study highlighting the potential of this modelling framework for benchmarking airports. To further this aim, we scrutinize three specific airports, namely Vienna, Hanover and Lyon, in order to examine the usefulness of the framework developed. For example, the results of the under-utilized airport in Hanover indicate that in the medium-term the airport could reduce operations to two of their three existing runways, instead of closing two runways as obtained with basic DEA, or alternatively attempt to increase cargo throughput as occurred at their two medium-term benchmark airports located in Venice and Hamburg. The formulations developed are suitable for assessing appropriate strategies with respect to aeronautical and commercial activities not only separately but also in combination, assuming cross-subsidization is an acceptable policy. According to the combined network DEA dynamic cluster revenue maximization approach, Lyon airport has achieved a sustainable level of aeronautical revenues and ought to search for appropriate commercial revenue opportunities as opposed to the results from the basic DEA, which suggest a further increase in aeronautical revenues of 40%. The methodology provides an airport manager with the tools for both exploratory data analysis and inefficiency estimation, removing the need for additional tests of homogeneity. Furthermore, utilizing an hourly capacity measure as both a terminal and airside proxy of physical capital appears to be new in airport benchmarking studies. Compared to the standard quantity measures such as the number of runways or gates, this proxy provides an improved managerial measure of the airport infrastructure as a system and allows us to consider bottlenecks at an airport.

This article is organized as follows: Section 2 presents individual modelling formulations that have been combined in Section 3 in order to produce airport benchmarks based on a network PCA-DEA dynamic clustering approach. Section 4 provides a description of the public data available for analysis and Section 5 compares the results of the combined formulations to those of basic DEA models and benchmarks a select subset of airports in order to demonstrate the utility of the approach developed in this research. Finally, Section 6 concludes and presents recommendations for further research.

2. Methodology

DEA is a non-parametric method of frontier estimation that measures the relative efficiency of decision-making units utilizing multiple inputs and outputs. DEA accounts for multiple objectives simultaneously without attaching ex-ante weights to each indicator and compares each decision-making unit (DMU) to the efficient set of observations, with similar input and output ratios, and assumes neither a specific functional form for the production function, nor the inefficiency distribution. DEA was first published in Charnes et al. [29] under the assumption of constant returns-to-scale[1] and was extended by Banker et al. [30] to include variable returns-to-scale. This non-parametric approach solves a linear programming formulation per DMU and the weights assigned to each linear aggregation are the results of the corresponding linear program. The weights are chosen in order to show the specific DMU in as positive a light as possible, under the restriction that no other DMU is more than 100% efficient. Consequently, a Pareto frontier is attained, marked by specific DMUs on the boundary envelope of input-output variable space. Formulation (1) presents an input-oriented model assuming variable returns-to-scale.

1 Constant returns to scale means that the producers are able to linearly scale the inputs and outputs without increasing or decreasing efficiency.

$$
\begin{aligned}
& \underset{\lambda,\theta}{Min} \ \theta \\
s.t. & \sum_{n=1}^{N} X^n \lambda^n \leq \theta X^a \\
& \sum_{n=1}^{N} Y^n \lambda^n \geq Y^a \\
& \sum_{n=1}^{N} \lambda^n = 1 \\
& \lambda^n, \theta \geq 0
\end{aligned}
\tag{1}
$$

where superscript a is the index of DMU^a, the unit under investigation; X^a represents the input values of DMU^a; Y^a is the output values of DMU^a and λ^n the intensity variable. θ represents the relative efficiency score, where a value of 1 indicates efficiency and a value smaller than 1 indicates the amount by which the relevant inputs ought to be decreased in order for DMU^a to be deemed relatively efficient.

In Section 2.1 we discuss the dynamic clustering mechanism that ensures comparable benchmarks are chosen from a dataset given exogenous parameter values. In Section 2.2 we present the network DEA model first designed to disaggregate the process of decision-making within a unit. Subsequently, in Section 2.3 we discuss the combination of principal component analysis and data envelopment analysis, which reduces over-estimation bias and in Section 2.4 we present a multi-dimensional scaling approach that produces a graphical representation of the data. Finally, in Section 2.5, we discuss a non-parametric statistical procedure that measures efficiency variation across different groups within the dataset in order to estimate the potential impact of environmental variables on the relative Pareto efficient frontier.

2.1 Dynamic clustering

Basic DEA benchmarking may lead to inappropriate targets for improvement in a dataset in which there are substantial differences in size among the DMUs under analysis. Sarkis and Talluri [31] propose second-stage clustering to identify benchmarks for poor performers, after applying DEA, to determine the relative efficiencies of airports. This study applies a dynamic clustering approach first proposed by Golany and Thore [25] that restricts the selection of best practice DMUs according to predefined boundaries within the DEA framework in a single stage process. The boundaries of the cluster are defined in relative terms, limiting the efficient reference set[2] to those DMUs whose input-output values are within the distance defined by the proportions.

In Fig. 2 we demonstrate the impact of the cluster restrictions for a simplified model with two outputs and a single input. DMU^a is compared to the Pareto frontier (darker line defined by DMUs 3 to 6) in a standard DEA formulation, with DMUs 4 and 5 acting as benchmarks. In our proposed approach, each inefficient airport may refer to a set of benchmarks that do not lie directly on the Pareto frontier, rather within the dotted radius. If DMU^a lies far enough away from the Pareto frontier as shown in Fig. 2, all potential benchmarks will lie in the interior of the envelope, resulting in DMUs 1 and 2 acting as benchmarks for DMU^a. The assumptions of this approach lead to the conclusion that DMU$^{a''}$, the hypothetical observation lying on the interior frontier, represents a relevant target which is more accessible than DMU$^{a'}$ in the short to medium term. It should be noted that as DMU^a is not compared to the overall Pareto frontier, no inference can be made with regard to the economic efficiency of this unit as defined under basic DEA. Instead, the motivation of this model is to find appropriate benchmarks and short to medium term targets in order to improve airport performance. Dynamic clustering improves on the Sarkis and Talluri two-stage procedure since additional information, such as the importance of each target DMU, can be drawn from the one step procedure.

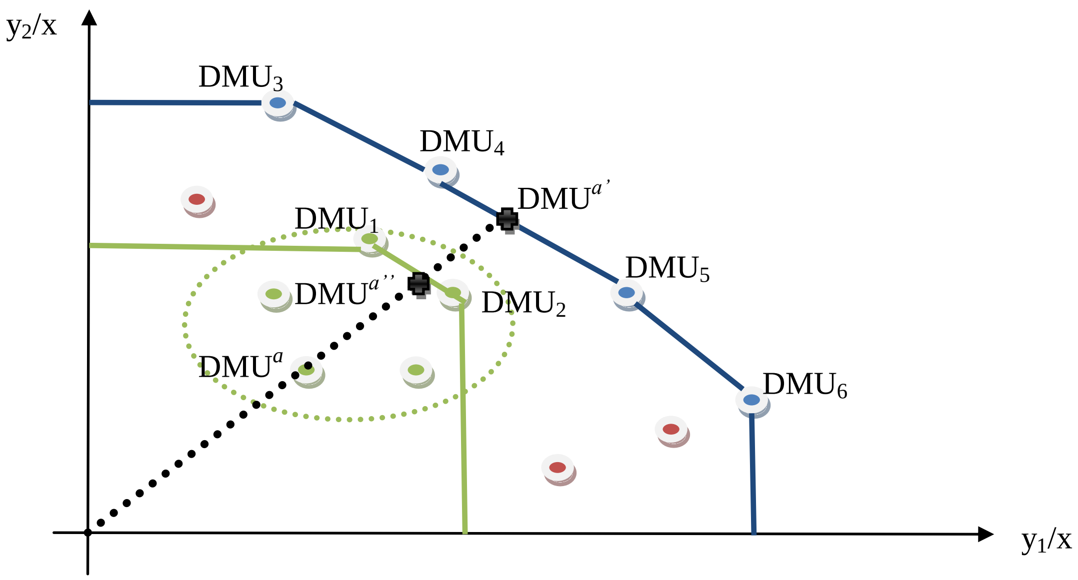


Fig. 2: Benchmark clustering

2 A reference set, or peer group, is defined by a subset of units "closest" to the unit under evaluation i.e. with similar mixes of inputs and outputs.

2.2 Network DEA

Network DEA models were first introduced by Färe [22] and Färe and Grosskopf ([32], [33]) and subsequently extended by Lewis and Sexton [34], Emrouznejad and Thanassoulis [35], Golany et al. [36], Chen [37], Kao [38] and Tone and Tsutsui [39]. Opening the black box permits an analysis of the optimal production structure of DMUs and their priorities, to determine both efficient subsystems and overall efficiency in order to allocate resources efficiently and determine appropriate targets. Castelli et al. [40] provide a classification of DEA models accounting for the internal structure of DMUs, depending on the assumptions of the modelling approach and then present mathematical formulations, extensions and applications. Fig. 3 demonstrates a network structure in which the outputs of some decision making sub-units (DMSU) become inputs for other sub-units. This framework has been widely applied in manufacturing production systems and supply chains. In transportation, network DEA has been applied by Yu and Lin [41] in order to simultaneously estimate passenger and freight technical efficiency, service effectiveness and technical effectiveness for 20 selected railways.

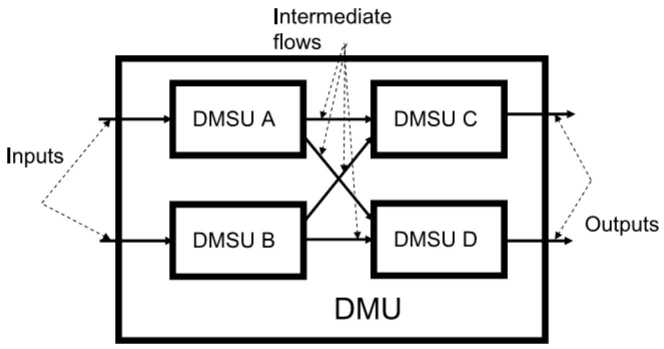


Fig. 3: Network Decision Making Unit & Sub-Units (Castelli et al. [40])

This research develops a network model that defines a multi-product airport in which capital, labour, materials and outsourcing of services produce traffic volume, in the form of aircraft movements, passenger and cargo. This throughput then generates revenues from aeronautical charges paid mostly by airlines, and from commercial terminal-side services serving passengers. The overall profits of this system are driven by services provided by outside parties including airlines and third party contractors as well as the airport processes themselves. Airport management retains reasonable control over labour, materials and levels of outsourcing but limited control over capital investments. In addition, management controls the variety and the pricing policies offered on the non-aeronautical side and partially controls aeronautical tariffs, dependent on the regulatory regime of the relevant country. Network DEA lends itself to a more accurate description of this process than standard performance analyses. In this research, network DEA describes the production process, demonstrating the sequential effects separating final and intermediate outputs including those under partial managerial control and those that are known to be non-discretionary. We have not defined multi-component parameters for shared inputs, as suggested in Beasley [42] and Cook et al. [43], because the joint resources (costs and airport capacities) could not be separately assigned to intermediate services (passengers, cargo and air traffic movements) and cannot be endogenized due to a causal (maybe non-linear) relationship between the overlapping intermediate outputs. Consequently, our benchmarking framework does not determine an aggregate efficiency score per airport as we are aware of airport management control restrictions. We concentrate separately on cost minimization given intermediate and final outputs, or on revenue maximization given the inputs and intermediate output levels required from the system.

2.3 Principal component analysis integrated with DEA

Dependent on the nature of the dataset, the results of the DEA model may not sufficiently distinguish between the efficient and inefficient DMUs due to an overestimation bias caused by the curse of dimensionality (Adler and Yazhemsky [28]). PCA-DEA is one of the methodologies that have been developed to reduce the number of inefficient DMUs incorrectly classified as efficient (Adler and Golany [27], [44]). The original variables are replaced with a smaller group of principal components (PCs), which explain the variance structure of a matrix of data through linear combinations of variables. The principal components are uncorrelated linear combinations ranked by their variances in descending order and those that explain little of the variance of the original data may be removed thus reducing the dimensions in the DEA linear program. In order to use principal components instead of the original data, the DEA model needs to be transformed to take into account the linear aggregation.

A rule-of-thumb computed in Adler and Yazhemsky [28] suggests that at least 76-80% of the information should be retained in the model in order to minimize the overestimation bias[3]. Clearly, if we use less than full information, we will lose some of the explanatory powers of the data but we will improve the discriminatory power of the model. It should be noted that as a result of the free sign in principal component analysis and the transformed constraints in the PCA-DEA model, the targets and benchmarks obtained could reflect a change in the current mix of input-output levels of the inefficient DMUs, along the lines of weight constrained DEA.

3 The rule-of-thumb defines the percentage of retained information required to balance the trade-off between the two incorrect definitions of (in) efficiency, namely efficient decision-making units defined as inefficient (under-estimation) and inefficient DMUs defined as efficient (over-estimation).

2.4 Visualizing multiple dimensions

Co-Plot, a variant of multi-dimensional scaling, aids both in exploring the raw data and in visualizing the results of DEA (Adler et al. [45]; Adler and Raveh [46]). Co-Plot positions each decision-making unit in a two-dimensional space in which the location of each observation is determined by all variables simultaneously according to a correlation analysis. The graphical display technique plots observations as points and variables as arrows, relative to the same arbitrary center-of-gravity. Observations are mapped such that similar DMUs are closely located on the plot, signifying that they belong to a group possessing comparable characteristics and behavior. A general rule-of-thumb states that the picture is statistically significant if the coefficient of alienation is less than 0.15 and the average of correlations is at least 0.75[4]. We apply Co-Plot to the set of variable ratios (each output divided by each input), in order to align the technique with DEA, such that Co-Plot graphically displays the DEA results in two dimensions. In general, the efficient DMUs appear in the outer circle of the plot signifying their relative achievements and we exogenously determine the color of the DMUs in order to clarify the results of the DEA.

2.5 Measuring variation across groups

In order to determine whether there are distinct differences between groups of airports, we apply the program evaluation procedure outlined in Brockett and Golany [47] and Sueyoshi and Aoki [48]. Four steps are required to implement the procedure. In the first step, the complete set of DMUs ($j=1,\ldots,n$) are split into k sub-groups and the model is run separately over each of the k groups. In the second step, for each of the k individual groups, the inefficient DMUs are moved to their hypothetical efficient level by projecting them onto the efficient frontier of their relevant group. In the third step, a pooled DEA is run with all n DMUs based on their adjusted variables. In the fourth step, a Kruskal-Wallis test is applied to determine if the k groups possess the same distribution of efficiency values within the pooled set. If the null hypothesis is correct, we expect to see most of the DMUs rated as efficient in step three. Note that in order to avoid inaccuracy in the nonparametric rank test, the number of observations in each of the k subgroups should be of similar size. If this is not the case, the size of the smallest subgroup is calculated and simple random sampling without replacement is applied to form subgroups of equally small sized samples. In order to test whether the findings are robust, Banker's F-test (Banker [49]) may be applied in the last stage of the procedure.

4 The coefficient of alienation is a single measure of goodness-of-fit for the configuration of n observations obtained from a smallest space analysis (Guttman [50]). The higher the correlation, the better the common direction and order of the projections of the n points along the arrow. The length of the arrow is proportional to the correlation.

3. Model formulations

In this section we describe three network PCA-DEA approaches with dynamic clustering that are then applied in Section 5. The application of network DEA to airports is new and to the best of our knowledge, we are aware of one working paper in the field, Lozano and Salmerón [51], in which capital utilization rather than managerial efficiency is analyzed based on network-DEA. Given the public data available for the study, Fig. 4 presents the airport network technology that we analyze based on a subset of variables described in Fig. 1. *X* represent inputs, *Y* outputs and *I* intermediate products. The number in brackets represents a node index in the network.

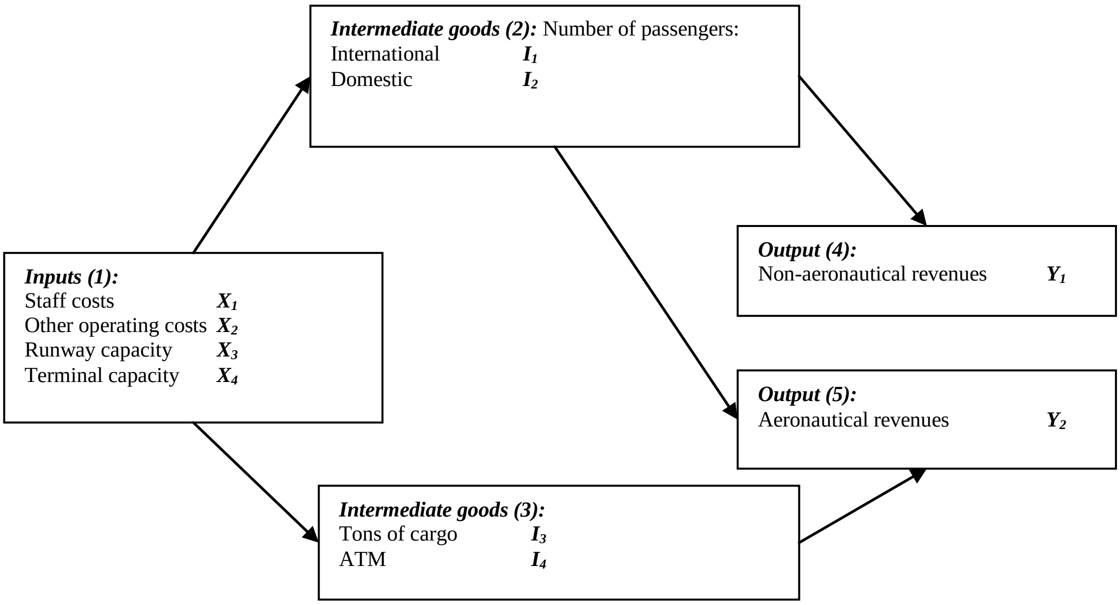


Fig. 4: Two-stage airport network technology

Model (2) assumes that airport management is interested in maximizing revenues, drawing from aeronautical activities and concessions given airport throughput on the terminal and airsides, which are in turn limited by the physical infrastructure and associated costs available to support the system. Drawing on discussions with airport managers and Pels et al. [7], we assume constant returns to scale with respect to revenues in that a doubling of the intermediate inputs, namely passengers, air traffic movements and cargo, should increase revenues at an equivalent rate. The network DEA formulation for the radial, output-oriented, constant returns to scale, mixed integer linear program applied in this research is presented in model (3.1), where superscript *a* is the index of DMU^a, the unit under investigation; X^a represents the input values of DMU^a; Y^a and I^a are the output and intermediate values of DMU^a respectively and the subscript of intensities for DMU^a, λ_{ij}^{n}, denotes the link leading from node *i* to node *j* in the network presented in Fig. 4. θ_1 and θ_2 represent the relative performance for the commercial and airside activities respectively, where a value of 1 indicates a position on the interior frontier and a value greater than 1 indicates by how much the relevant revenues ought to be increased in order for DMU^a to move towards the frontier. It should be noted that the first four rows of model (3.1) are not summed over *n*, in order to restrict envelopment intensities λ_{24}^{n} and λ_{235}^{n}, thus comparing airports that possess input levels that lie within a boundary of 10% to 300% of DMU^a inputs and between 20% and 200% of DMU^a intermediate outputs. Parameter values, α_l=0.1, α_u=3, β_l=0.2, β_u=2, were chosen such that a sufficiently rich set of airports exist in the cluster. Sensitivity analyses of our current dataset suggest that smaller bounds result in excessive limitations and pure self-comparisons over time whereas a wider set lead to unreasonable benchmarks whereby London Heathrow (LHR) and Tallinn (TLL), representing the largest and smallest airports in the dataset, are considered directly comparable.

In order to connect Fig. 4 and the clustering approach, λ_{12}^{n} and λ_{13}^{n} are binary variables and λ_{24}^{n} and λ_{235}^{n} are non-negative continuous variables. If λ_{12}^{n}=1 then DMU^n could be included in the peer group for DMU^a on the non-aeronautical side and if λ_{12}^{n}= λ_{13}^{n}=1 then DMU^n could be included in the peer group for DMU^a on the aeronautical side. λ_{12}^{n} consequently connects costs to the number of passengers produced and λ_{13}^{n} connects costs to cargo and air traffic movement production such that DMUs of similar size and cost structure represent potential benchmarks[5]. λ_{24}^{n} connects the number of passengers to non-aeronautical revenue derived and λ_{235}^{n} connects passengers, cargo and air traffic movements to aeronautical revenue. Since no trade-off between aeronautical and non-aeronautical activities is introduced in the model, benchmarks on each side of the airport activity are determined independently.

5 The effect of this approach is depicted in Figure 2 resulting in DMUs 1 and 2 acting as benchmarks for DMUa: $\lambda_{12}^{1,2}$ and $\lambda_{13}^{1,2}$ are binary variables equal to 1 since they lie within the boundaries of the first four equations in model (2) whereas $\lambda_{12}^{4,5}$ and $\lambda_{13}^{4,5}$ equal 0.

$$
\begin{aligned}
&\underset{\lambda,\theta}{Max}\ \ \theta_1 + \theta_2 \\
&s.t.\ \alpha_l X_j^a \lambda_{12}^n \le X_j^n \lambda_{12}^n \le \alpha_u X_j^a \lambda_{12}^n \quad \forall\, j = 1,2,3,4 \quad ,\ n = 1,\dots,N \\
&\quad \beta_l I_k^a \lambda_{12}^n \le I_k^n \lambda_{12}^n \le \beta_u I_k^a \lambda_{12}^n \quad \forall\, k = 1,2 \quad ,\ n = 1,\dots,N \\
&\quad \alpha_l X_j^a \lambda_{13}^n \le X_j^n \lambda_{13}^n \le \alpha_u X_j^a \lambda_{13}^n \quad \forall\, j = 1,2,3,4 \quad ,\ n = 1,\dots,N \\
&\quad \beta_l I_m^a \lambda_{13}^n \le I_m^n \lambda_{13}^n \le \beta_u I_m^a \lambda_{13}^n \quad \forall\, m = 3,4 \quad ,\ n = 1,\dots,N \\
&\quad \sum_{n=1}^{N} I_k^n \lambda_{24}^n \le I_k^a \quad \forall\, k = 1,2 \\
&\quad \sum_{n=1}^{N} Y_1^n \lambda_{24}^n \ge \theta_1 Y_1^a \\
&\quad \sum_{n=1}^{N} I_j^n \lambda_{235}^n \le I_j^a \quad \forall\, j = 1,2,3,4 \\
&\quad \sum_{n=1}^{N} Y_2^n \lambda_{235}^n \ge \theta_2 Y_2^a \\
&\quad \lambda_{24}^n \le \lambda_{12}^n, \quad \lambda_{235}^n \le \lambda_{12}^n, \quad \lambda_{235}^n \le \lambda_{13}^n \\
&\quad \lambda_{12}^n \in \{0,1\}\ \ \lambda_{13}^n \in \{0,1\} \quad binary \\
&\quad \theta_1, \theta_2, \lambda_{24}^n, \lambda_{235}^n \ge 0
\end{aligned} \tag{2}
$$

Alternatively, whilst it may be assumed that a private, unregulated airport pursues profit maximization, airports that are subject to economic regulation may behave as social welfare maximizers. Hence, maximizing aeronautical revenues may not be the target of airport management due to regulatory constraints. Furthermore, even profit maximizers may consider lower aeronautical charges as an opportunity to expand non-aeronautical activities and generate additional revenues by attracting airlines through lower airport charges. Consequently, the network DEA formulation for the radial, output-oriented, constant returns-to-scale, mixed integer linear program combining both aeronautical and concession activities is presented in (3). The goal is to maximize non-aeronautical revenue (Y_1) given international and domestic passengers, cargo and air traffic movements. Physical infrastructure (terminal and runway movements), costs (labour and materials) and intermediate outputs define the reference set for each DMU as in model (2). Aeronautical revenue (Y_2) is included in the analysis as a non-discretionary variable (Banker and Morey [26]). According to this model, benchmarks consist of airports achieving higher non-aeronautical revenues, given similar levels of aeronautical revenue whilst comparing airports of similar size and demand levels. In the following we will refer to (2) as the independent model where the clusters were independently defined and the performance of the aeronautical and non-aeronautical side estimated separately. Formulation (3) presents the combined model since a common set of benchmarks are considered but only non-aeronautical revenues are maximized.

$$
\begin{aligned}
& \underset{\lambda,\theta}{Max}\ \theta_1 \\
& s.t.\ \alpha_l X_j^a \lambda_{123}^n \le X_j^n \lambda_{123}^n \le \alpha_u X_j^a \lambda_{123}^n \quad \forall j = 1,2,3,4 \quad , \ n = 1,\ldots,N \\
& \quad \beta_l I_k^a \lambda_{123}^n \le I_k^n \lambda_{123}^n \le \beta_u I_k^a \lambda_{123}^n \quad \forall k = 1,2,3,4 \ \ , \ n = 1,\ldots,N \\
& \quad \sum_{n=1}^{N} Y_1^n \lambda_{235}^n \ge \theta_1 Y_1^a \\
& \quad \sum_{n=1}^{N} Y_2^n \lambda_{235}^n \ge Y_2^a \\
& \quad \sum_{n=1}^{N} I_j^n \lambda_{235}^n \le I_j^a \qquad \forall j = 1,2,3,4 \\
& \quad \lambda_{235}^n \le \lambda_{123}^n \\
& \quad \lambda_{123}^n \in \{0,1\} \qquad binary \\
& \quad \theta_1, \lambda_{235}^n \ge 0
\end{aligned}
\tag{3}
$$

The network PCA-DEA formulation for the radial, input-oriented, variable returns-to-scale, mixed integer linear program proposed in this research is presented in formulation (4). The cost minimization assumes variable returns to scale, since a doubling of output should not necessarily result in a doubling of staff, materials and outsourcing costs (Gillen and Lall [9]; Pels et al. [7]). As opposed to the output-oriented model, we have combined domestic and international passengers into one intermediate variable I_{pax} in order to reduce the number of variables. Furthermore, we have applied principal component analysis (PCA) to reduce the over-estimation bias and improve the level of discrimination in the results. The first principal component (PC_{cost}) combines staff costs and other operating costs, explaining 89% of the variance in the original data. PC_{cap} combines terminal and runways capacities, explaining 85% of the original information. Including all PCs would provide precisely the same solution as that achieved under the original DEA formulation.

Model (4) clusters airports according to revenue and traffic mix, whereby the total number of passengers is included in the commercial side and all intermediate activities are included in the aeronautical side. Parameter values were set at α_l=0.1, α_u=3, β_l=0.2, β_u=2. $\lambda_{24}^n, \lambda_{25}^n$ and λ_{35}^n are binary variables and λ_{12}^n and λ_{123}^n are non-negative continuous variables. S_{cost} and S_{cap} are slack variables and l_{cost} and l_{cap} are normalized eigenvectors based on costs and capacities respectively. θ_1 and θ_2 represent the restricted relative efficiency scores on the terminal side and airside respectively, whereby a score of 1 means that the airport lies on the (interior) frontier and less than 1 indicates the level of input retraction required in order to catch-up with the benchmarks identified. θ_1=1 indicates a cost minimization approach with respect to the non-aeronautical activities of the airport and θ_2=1 indicates cost minimization with respect to all activities of the airport (passengers, cargo, air traffic movements) whereby the source of revenues draws from both the non-aeronautical and the aeronautical sides. To restrict the variability of physical infrastructure, we assume that terminal and runway capacities may be adjusted up to 30% in the medium term (δ=0.7).

$$
\begin{array}{ll}
\underset{\lambda,\theta}{Min}\ \theta_1+\theta_2 & \\
s.t.\ \alpha_l Y_1^a \lambda_{24}^n \le Y_1^n \lambda_{24}^n \le \alpha_u Y_1^a \lambda_{24}^n & \forall \quad n=1,\ldots,N \\
\quad \beta_l I_{PAX}^a \lambda_{24}^n \le I_{PAX}^n \lambda_{24}^n \le \beta_u I_{PAX}^a \lambda_{24}^n & \forall \quad n=1,\ldots,N \\
\quad \alpha_l Y_2^a \lambda_{25}^n \le Y_2^n \lambda_{25}^n \le \alpha_u Y_2^a \lambda_{25}^n & \forall \quad n=1,\ldots,N \\
\quad \beta_l I_{PAX}^a \lambda_{25}^n \le I_{PAX}^n \lambda_{25}^n \le \beta_u I_{PAX}^a \lambda_{25}^n & \forall \quad n=1,\ldots,N \\
\quad \alpha_l Y_2^a \lambda_{35}^n \le Y_2^n \lambda_{35}^n \le \alpha_u Y_2^a \lambda_{35}^n & \forall \quad n=1,\ldots,N \\
\quad \beta_l I_m^a \lambda_{35}^n \le I_m^n \lambda_{35}^n \le \beta_u I_m^a \lambda_{35}^n & \forall \quad m=3,4 \quad , \quad n=1,\ldots,N \\
\quad \sum_{n=1}^{N} PC_{\cos t}^n \lambda_{12}^n + \sum_{i=1}^{2} l_{\cos t1}^{i1} S_{\cos t1}^i = \theta_1 PC_{\cos t}^a & \\
\quad \sum_{n=1}^{N} PC_{cap}^n \lambda_{12}^n + \sum_{i=1}^{2} l_{cap1}^{i1} S_{cap1}^i = \delta PC_{cap}^a & \\
\quad \sum_{n=1}^{N} I_{PAX}^n \lambda_{12}^n \ge I_{PAX}^a & \\
\quad \sum_{n=1}^{N} PC_{\cos t}^n \lambda_{123}^n + \sum_{i=1}^{2} l_{\cos t2}^{i1} S_{\cos t2}^i = \theta_2 PC_{\cos t}^a & \\
\quad \sum_{n=1}^{N} PC_{cap}^n \lambda_{123}^n + \sum_{i=1}^{2} l_{cap2}^{i1} S_{cap2}^i = \delta PC_{cap}^a & \\
\quad \sum_{n=1}^{N} I_{PAX}^n \lambda_{123}^n \ge I_{PAX}^a & \\
\quad \sum_{n=1}^{N} I_q^n \lambda_{123}^n \ge I_q^a & \forall \quad q=3,4 \\
\quad \sum_{n=1}^{N} \lambda_{12}^n = 1, \qquad \sum_{n=1}^{N} \lambda_{123}^n = 1 & \\
\quad \lambda_{12}^n \le \lambda_{24}^n, \qquad \lambda_{123}^n \le \lambda_{24}^n, \qquad \lambda_{123}^n \le \lambda_{25}^n, \qquad \lambda_{123}^n \le \lambda_{35}^n & \\
\quad \lambda_{24}^n,\ \lambda_{25}^n,\ \lambda_{35}^n \in \{0,1\} \quad binary & \\
\quad \theta_1,\ \theta_2,\ \lambda_{12}^n,\ \lambda_{123}^n \ge 0 &
\end{array}
\tag{4}
$$

4. Dataset

In this section we describe the variables collected, additional environmental variables that may be required to adapt the model to ensure homogeneity of the production process and the complete set of observations together with an initial exploratory data analysis. The dataset consists of 43 European airports located in 13 different countries. In order to increase the likelihood of comparability, we focus the case study on European airports. In Europe passenger terminals are normally operated and maintained by the airport operator whereas in the U.S., airports frequently contract out such activities to airlines making cross-comparisons somewhat problematic. However, disaggregated data for major European airports such as Paris Charles de Gaulle, Barcelona or Madrid, all of which belong to airport groups or authorities, were not available as financial data is only reported for the entire group. Consequently, we have pooled the data to an unbalanced set of 294 observations covering the time period from 1998 to 2007 (the Appendix lists the set of airports under study, the specific timeframe for which the data was available and whether ground handling processes are undertaken in-house or outsourced). All airports offer domestic and international routes, however airports located in smaller countries such as the Netherlands generally have very few domestic destinations. The passenger volume varies considerably from less than a million passengers at Tallinn and Durham Tees Valley airports up to more than 60 million at London Heathrow, the largest European airport in terms of passenger throughput and number three in the world (Airports Council International [52]).

Ten variables were collected in total for purposes of analysis based on publicly available data. The variables are categorized into three groups; four inputs (*X*), four intermediate products (*I*) and two outputs (*Y*). Table 1 presents summary information and specifies the data sources. The operating inputs consist of staff costs and all other non-labour related operating costs, which include materials and outsourcing. Although a smaller airport than London Heathrow in terms of air traffic movements, Frankfurt's staff costs are highest due to the level of vertical integration whereby the airport operates most of the services by itself or through wholly-owned subsidiaries. As an example, the airport manages the ground handling operations which represent one of the most labour intensive activities at an airport, a process traditionally organized by airlines or independent third party providers at Heathrow. Consequently, Heathrow spends the most on other operating costs, reflecting the high levels of outsourcing undertaken.

Generally, as a proxy for capital, physical data such as the number of runways, gates, check-in counters and overall terminal size is collected. However, such data is often problematic because the number of runways does not include information on the configuration or the impact of weather on the number of runways open within a given timeframe. Furthermore, the terminal area in square metres is somewhat subjective since some airports report gross terminal area including sections of an airport that are not open to the public. If the dataset covers more than one country, the monetary measurement of physical capital also creates difficulties due to different national accounting standards and depreciation methods or periods across countries. For example, the airports of the British Airports Authority (BAA) depreciate their runways over 100 years whereas the airports operated by the Aéroports de Paris depreciate over a period of 10 to 20 years (Graham [53]). In this research, terminal capacity is defined in terms of passengers handled within an hour, thus combining the capacities of all terminal facilities including check-in counters, security controls, baggage delivery and retail area into one common capacity figure. The airside is defined by the declared runway capacity, specified as the number of departing and arrival movements specified per hour. Airport stakeholders negotiate this parameter biannually which is primarily used to avoid congestion at schedule facilitated airports and aid in the allocation of slots at coordinated airports (IATA [54]). The advantage of using declared capacity is that the parameters account for bottlenecks across the terminal and runway systems, providing two individual capacity measures. Amsterdam possesses the highest agreed terminal and runway capacities in our sample with 26,000 passengers and 110 movements per hour. Due to their geographical location near the coast, they require a special runway configuration to operate as a hub airport. The smallest airport with respect to runway capacity is Florence in the Tuscany region, with a maximum hourly rate of twelve movements. Due to its short, single runway system (1,688 m), the airport can handle aircraft up to the size of a Boeing 737 or an Airbus A319 (Aeroporto di Firenze [55]).

The annual traffic volume is represented by the number of passengers, commercial air traffic movements and tons of cargo (trucking is excluded). The passengers are divided according to domestic and international destinations. Unfortunately, we could not collect enough data to separate the passengers between intercontinental and European flights or account for transfer passengers, which would be preferable since these groups probably generate different revenue streams. Non-aviation revenues include revenues from retail activities and restaurants, concessions and income from rents and utilities. Aviation revenues are generated from (often regulated) landing and passenger charges, ground handling undertaken in-house and cargo activities. The largest non-aeronautical revenues were generated at Heathrow, whereas Frankfurt earned the highest aviation revenues. Commercial revenues equal 67% of total airport revenues on average in the dataset, clearly supporting the argument that non-aeronautical activities should not be ignored in a productivity analysis of airports from a managerial perspective, particularly when considering the possibility of cross-subsidization.

All financial data is deflated to the year 2000 and adjusted by the purchasing power parity according to the United States dollar in order to ensure comparability across countries. In addition, the data has been normalized by the standard deviation to limit the influence of outliers in the dataset.

Table 1: Variables in airport performance analysis

Variable	Description	Name	Average	Maximum	Minimum	Source
Staff costs	Wages and salaries, other staff costs	*X1*	81,704,359	1,080,756,267	5,962,213	Annual Reports
Other operating costs	Costs of materials, outsourcing and other	*X2*	103,364,471	725,987,196	5,010,381	Annual Reports
Declared runway capacity	Total movements per hour	*X3*	49	110	12	IATA [56], Airport and Coordinator Websites
Terminal capacity	Total passenger throughput per hour	*X4*	6,768	26,000	450	IATA [56], Airport Websites
International passengers	Annual passenger volume	*I1*	10,300,571	61,517,733	355,579	IATA [56], Airport Websites
Domestic passengers	Annual passenger volume	*I2*	2,433,287	9,932,208	48	IATA [56], Airport Websites
Cargo	Metric tons (trucking excluded)	*I3*	214,076	2,190,461	37	IATA [56], Airport Websites
Air transport movements	Total commercial movements	*I4*	152,133	492,569	16,000	IATA [56], Airport Websites
Non-aeronautical revenues	Revenues from concessions own retail and restaurants, rents, utilities and other	*Y1*	117,906,043	1,107,046,057	4,629,813	Annual Reports
Aeronautical revenues	Landing, passenger and aircraft parking charges; revenues from ground handling, cargo revenues and other	*Y2*	175,507,645	1,739,331,693	7,199,668	Annual Reports

5. Results

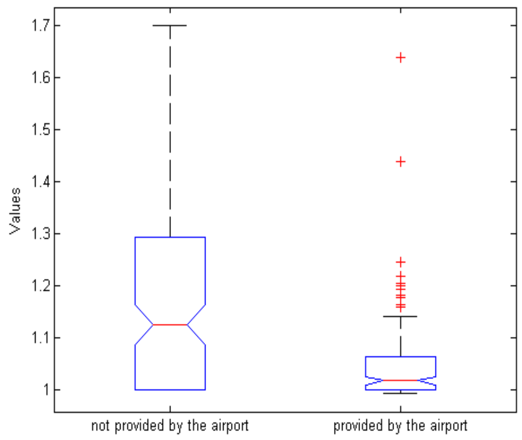
In the following section we identify the impact of vertical integration and subsequently include the information in the dynamic clusters. In section 5.2 we compare and contrast the results of a basic DEA model with the network PCA-DEA formulation. Section 5.3 discusses the benchmarking results for a subset of airports, specifically Vienna which acts as a benchmark in the current case study, Hanover as an example of an input-oriented inefficient case and finally Lyon, in order to compare and contrast the independent output model (formulation 2), which assesses the performance estimates of both revenue generating activities separately, and the combined output model (formulation 3) in which only commercial revenues are maximized.

5.1 Efficiency variation across groups

When estimating the relative efficiencies, it would appear that airports offering in-house ground handling services operate on a different production frontier to airports that outsource this activity. This is not immediately obvious since airports providing ground handling services in-house have higher labour costs whereas airports that outsource the activity have higher ‘other’ costs. Both models have higher revenues than airports that permit third parties to provide the service since no costs appear on the books and only minor concessional fees may be collected because the contracts themselves do not appear on the airport's accounting books. To evaluate the potential for different productivity levels, the non-parametric program evaluation procedure was applied to basic DEA which contains the last stage of formulations (2) and (4), combining both sources of revenues into one efficiency estimate. Based on Fig. 4, the output orientation assumes constant returns-to-scale and includes nodes {*2345*}, while the input orientation includes the inputs and outputs from nodes {*123*} and assumes variable returns-to-scale. In our sample, 21 airports offer ground handling and 22 outsourced or never offered this service which translates into 156 DMUs in the ground handling group and 138 DMUs otherwise (see Appendix). The results are clear and significant that airport operators providing ground handling appear to be revenue maximisers but were highly inefficient in cost minimization relative to airports from the non-ground handling group. Graph (a) in Fig. 5 shows the DEA efficiency scores on the vertical axis for the two groups from a revenue maximization perspective and (b) shows the DEA efficiency scores from the cost minimization perspective across the two groups. Airports with ground handling activities perform on average 10% better in maximizing their outputs as their aeronautical revenues per passenger are naturally higher whereas in the input-oriented model, airports that do not provide ground handling achieve on average 10% higher efficiencies since no costs are associated with this service.

(a) Output-orientation

Source	Sum of squares	Degrees of freedom	Mean Squares	Chi-sq	Prob > Chi-sq
Groups	243,606	1	243,606	33.7	6.4e-009
Error	1,873,769	292	6,417		
Total	2,117,376	293			



(b) Input-orientation

Source	Sum of squares	Degrees of freedom	Mean Squares	Chi-sq	Prob > Chi-sq
Groups	357,302	1	357,302	49.4	2.0e-012
Error	1,760,162	292	6,028		
Total	2,117,465	293			

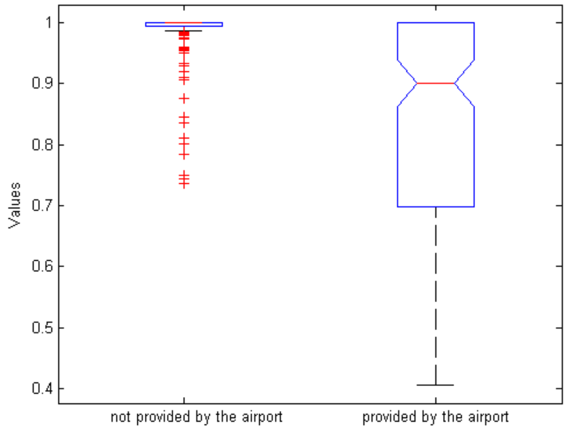


Fig. 5: Kruskal-Wallis ANOVA for outsourcing[6]

In order to test the robustness of our findings, the Banker F-test [49] is also applied both in the third stage of the program evaluation procedure and on basic DEA scores when two sub-groups of DMUs face the same frontier as suggested in Banker [49], assuming exponential[7] and half-normal[8] inefficiency distributions (Table 2).

Table 2: Banker F-test for outsourcing

(a) Output-orientation

Inefficiency distribution	Test applied on	Test Statistic	Prob>F
Exponential	Entire dataset	2.5726	5.62797E-16
	Program evaluation procedure	3.9766	4.69994E-31
Half-normal	Entire dataset	5.5646	9.70871E-24
	Program evaluation procedure	9.1030	3.84569E-36

(b) Input-orientation

Inefficiency distribution	Test applied on	Test Statistic	Prob>F
Exponential	Entire dataset	2.8054	1.33497E-18
	Program evaluation procedure	9.1584	1.27351E-70
Half-normal	Entire dataset	7.8721	2.68437E-32
	Program evaluation procedure	22.7297	1.3301E-123

The results also proved to be consistent for a basic DEA model in which air traffic movements, passengers, cargo and commercial income including ground-handling revenues were selected as output in order to adjust for the effect of outsourcing. Staff and other operating costs and runway and terminal capacities were defined as inputs. Having now considered both costs and revenues in the efficiency estimation, the radial variable returns-to-scale, input-oriented model still indicated significant efficiency differences across both groups based on the results of a Kruskal-Wallis test. In summary, both the non-parametric Kruskal-Wallis and parametric F-test reach the same significant result supporting a rejection of the null hypothesis (Fig. 5 and Table 2). After liberalization in 1996, airports that provided ground-handling were required to permit competitors' access. Munich and Frankfurt have claimed substantial losses in this segment on a regular basis (Dietz [57]; Hutter [58])[9]. However, the strong labour unions in Germany have prevented airport management from either cutting wages or outsourcing this service to third-party providers without guarantees that

6 The vertical axis of the boxplot represents the efficiency scores computed in the third step of the program evaluation procedure (a score of one implies relative efficiency).

7 Test statistic= $\dfrac{\sum_{j=1}^{N_1}\tilde{u}_{j1}/N_1}{\sum_{j=1}^{N_2}\tilde{u}_{j2}/N_2}$ and is distributed as F with ($2N_1$, $2N_2$) degrees of freedom, where $\tilde{u}_j = \theta_j - 1$ in the output oriented and $\tilde{u}_j = \dfrac{1}{\theta_j} - 1$ in the input oriented model and belong to the range [0, ∞).

8 Test Statistic= $\dfrac{\sum_{j=1}^{N_1}\tilde{u}_{j1}^2/N_1}{\sum_{j=1}^{N_2}\tilde{u}_{j2}^2/N_2}$ and is distributed as F with (N_1, N_2) degrees of freedom.

9 Most German airports are fully or at least major publicly owned and if ground handling is operated in-house by the parent company, the airport pays salaries based on public tariffs, which are on average 20% higher compared to private ground handling companies. Some German airports, such as the minor-private airport Hamburg, outsourced the ground handling segment to a 100% subsidiary in order to set flexible tariffs however this was not deemed acceptable by the public shareholders of Munich for example.

workers would continue under the same conditions. Thus, at least in the short-term, the degree of outsourcing can be regarded as a political factor that is beyond managerial control and ground handling is included in the network DEA formulation as an environmental variable which will further limit the potential benchmark set via clustering. Consequently, we analyse airports that offer ground handling services entirely separately from airports outsourcing this activity to airlines or independent providers, in turn limiting comparisons to airports operating similar strategies.

5.2 Comparison of basic and network DEA

In contrast to our formulations (2 to 4), basic DEA does not restrict potential benchmarks nor does it permit a limited deviation in one or more variables. In order to assist the comparison between basic and network DEA results, we have exogenously divided the dataset according to in-house or outsourced ground handling provision and applied DEA individually to each category. For the output orientation, the technology of Fig. 4 reduces to nodes {*235*} and {*24*} with respect to aeronautical and commercial activities respectively, while the input orientation case collapses to nodes {*12*} when assessing the non-aeronautical side and {*123*} with respect to both activities.

The basic DEA results generate consistently efficient airports that belong either to the set of smallest airports e.g. Bremen, Florence and Ljubljana, which provide ground handling in-house and Malta, Durham Tees Valley and Leeds/Bradford which outsource, or the largest airports in the sample such as Frankfurt. In neither output-oriented formulations do airports achieve 100% efficiency over the entire review period, although Salzburg, Ljubljana and Malta appear consistently close to the frontier. A notable exception is Cologne-Bonn, which remains cost efficient with respect to both activities but operates very inefficiently (between 44% and 77% over time) with respect to the commercial side. Cologne-Bonn is the European hub for the parcel service provider UPS, which rents office space and warehouses from the airport, suggesting a behaviour different to others in the sample (Cologne-Bonn Airport [59]).

Under basic DEA, all airports are compared against a single Pareto frontier and Salzburg represents an important benchmark for Vienna, Dusseldorf, Frankfurt, Hamburg and Munich. However, it is doubtful that the management of a primary or secondary hub airport would adopt the strategies of an airport that handles less than 2 million passengers per year with very low cargo throughput too. Durham Tees Valley, a small airport in East England with less than 700,000 passengers per year was defined as a benchmark for Lyon, Geneva, Oslo and the hub airport in Zurich, which would not occur in the formulations we present due to the dynamic clustering approach.

Airports in very small clusters are unique in character and in the extreme case tend to form their own reference set. In the current dataset, these mostly included the smaller and less congested airports such as Tallinn, Leeds/Bradford and Durham Tees Valley. These airports can be identified as outliers according to the Andersen and Petersen [60] super efficiency procedure. Such observations frequently influence the basic DEA Pareto frontier, for example Durham Tees Valley appears within the reference sets of Oslo and Zurich airports. In comparison, the results of the cost minimization formulation (3.3) categorizes Copenhagen and London Stansted as peer airports for Oslo and Zurich hence the modelling approach indicates benchmarks that are more homogeneous in character. Another unique example includes Dortmund, which acts a self benchmark in the cost minimization approach from 2003 to 2007, namely after their capacity expansion and severe reduction in cargo operations. Dortmund is the only airport that exhibits operational losses over the entire timeframe. The airport is partly owned by the local electricity distributor and losses are covered by their major shareholder (Dortmund Airport [61]). Dortmund is located in the Ruhr area (*Ruhrgebiet*) with a population of more than five million, representing the largest agglomeration in Germany. Airport competition includes Dusseldorf, Cologne-Bonn and Paderborn which are located in their catchment area (defined as 90 km around the airport) and intermodal competition includes high speed rail and the motorway, especially on domestic routes and traffic originating in Benelux. Hence despite their high capital investment, it may be necessary for the airport to further decrease their aviation charges in order to attract airlines and new destinations thereby generating additional commercial revenues.

The network DEA formulations provide the user with an exploratory data analysis that does not exist in the basic DEA results. The results of formulations (3.1) and (3.3) demonstrate that the average cluster size for each inefficient airport was reasonably small because the capacity of airports varies considerably across the sample and some airports suffer low utilization rates whereas other are highly congested. The operating costs at highly congested airports were large mostly due to employee costs hence airports with similar capacities did not necessarily belong to the same cluster. In general, large clusters indicate that various airports in the sample possess similar characteristics which in our dataset included Dusseldorf, Hamburg, Strasbourg, Venice and London Gatwick.

5.3 Benchmarking airports

In this sub-section, we describe the type of results and analysis that are achievable by collecting data and applying the formulations described in Section 3. We focus on relatively productive Vienna, relatively inefficient Hanover and finally Lyon, in order to describe the potential balance between the two revenue streams. Vienna is an example of an airport that has gradually improved on both the cost inputs and revenue outputs over time. Vienna appears in the reference set of Cologne-Bonn and

Dusseldorf in the input-oriented case. Between 1998 and 2007 Vienna's costs and revenues increased on average by similar proportions (99% and 94% respectively)[10], while traffic volume grew by 76% for total passengers, 54% for air traffic movements and 37% for cargo. The input-oriented case in Fig. 6 shows that from 1998 to 2003, Vienna lies close to the arrows that display the ratio of intermediate outputs to costs. In 2004, both staff costs and other operating costs substantially increased partly due to the introduction of a 100% hold baggage screening policy and the founding of a subsidiary for infrastructure maintenance (Vienna International Airport [62]). After 2004 greater emphasis has been allocated to the issue of runway utilization, viewed in Fig. 6 by the proximity of the later years to the capital asset related ratios. Vienna airport moves in a positive direction towards an improved utilization of the runways which increased from 48% to 66% between 2000 and 2007. Hence, despite substantial cost increases, the airport still managed to increase its relative performance as their costs per air traffic movement decreased over time. On the aeronautical output side, the airport management changed their regulated tariff structure by increasing passenger charges from an average price of 3.90€ in 1999 to 5.90€[11] in 2007 and decreasing overall landing charges by 3%, whilst reducing them for larger aircraft by up to 20%. In summation, Vienna airport increased passenger charges out of the total aviation revenues collected from 33% in 2001 to 46% in 2007 and decreased the share of landing fees from 44% to 28% over the same period (Vienna International Airport [62]). These policies appear to have aided Vienna to achieve revenue productivity.

10 Staff costs increased by 111% and other operating costs by 88%. Non-aeronautical revenues increased by 78% and aviation revenues by 109%.

11 These values are an average passenger price and were computed by dividing total passenger revenue charges by passenger throughput as obtained from the annual reports. They could therefore deviate from the passenger charges specified in the charges manual. Both prices were given in nominal values.

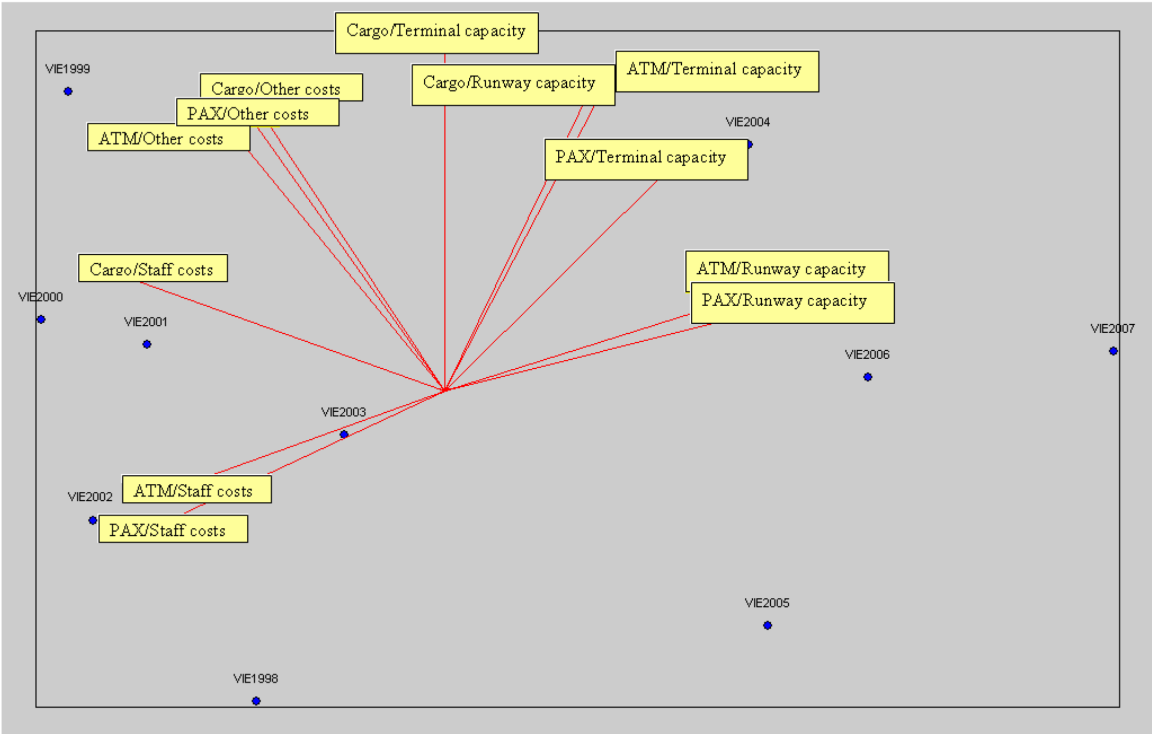


Fig. 6: Co-Plot graphic display of Vienna's input-oriented strategy[12]

Hanover airport is the ninth largest airport in Germany handling 5.6 million passengers in 2008 (ACI [52]). It is partly owned by the Fraport AG Company which owns and operates Germany's gateway hub in Frankfurt/Main. Whilst Hanover's non-aeronautical revenues from rents and utilities increased over the decade analyzed due to the development of a large airport city, the relative cost performance θ_1 consistently dropped from 72% in 1998 to 60% in 2006[13]. Over the same time period, passenger volume increased by 17%, air traffic movements by 8% and cargo dropped by 37% however staff costs and other operating costs increased by 80% and 64% respectively, as shown in Table 3. Fig. 7 displays the gradual decline in productivity (see the red arrow) and the change in benchmarks over time from Venice to Nuremberg (see white dots), the latter representing a relatively more expensive airport to operate[14]. From the technical perspective, Hanover shows potential to expand airport activities due to a declared runway capacity of 60 movements per hour. Capacity utilization at Hanover remained at a stable 23%, whereas Florence and Venice achieve 40% utilization and Hamburg slightly more than 50%. Nuremberg, Hanover's benchmark, achieves a

12 Coefficient of alienation is 0.06 and average of correlations is 0.89.

13 If δ=1 is assumed, terminal and runway capacities may decrease to a lower limit of zero. Hanover's cluster of 100 DMUs is stable over time, a common set of benchmarks exists between 1998 and 2006 (Florence, Hamburg and Venice) and a comparison of θ1 over time is possible. Over the same period θ2 is close to 1 due to overestimation bias caused by the relatively limited cluster size of 30-40 DMUs.

14 If δ=0.7 is assumed, terminal and runway capacities may be adjusted up to 30% in the medium term. As a result θ1=0.7 since 2001, although the dynamic clustering shows the change in productivity over time through changes in the set of benchmarks (see Table 3 and Fig. 6).

capacity utilization of less than 40% which is still higher than that of Hanover. Bremen and Dresden are the long term benchmarks according to basic DEA, which appear as black dots on the left edge in Fig. 7[15]. When comparing the results for Hanover in 2007 with respect to network PCA-DEA (θ_1=0.7) and basic DEA (θ_1=0.49) it becomes clear that the long term goal for Hanover, ceteris paribus, would be to close two of the three runways. The medium-term network DEA results suggest that it would be sufficient to close the equivalent of a single runway.

Table 3: Benchmarking Hanover airport

Airport	Declared Runway Capacity	Terminal Capacity	Staff Costs (US$)	Other Operating Costs (US$)	Domestic Passengers	International Passengers	Air Transport Movements	Cargo (tons)
DMUs under review								
HAJ1998	50	4,000	39,137,748	41,838,277	1,014,723	3,814,405	70,815	10,954
HAJ1999	50	4,000	44,100,404	40,270,806	1,080,384	4,017,528	76,914	7,724
HAJ2000	60	4,000	49,858,032	35,123,344	1,246,083	4,284,201	83,687	9,027
HAJ2001	60	4,000	49,344,453	39,876,433	1,067,834	4,089,724	75,368	6,712
HAJ2002	60	4,000	48,501,264	37,857,069	1,018,412	3,733,509	73,278	6,058
HAJ2003	60	4,000	51,602,584	46,057,783	1,010,975	4,033,895	74,960	6,338
HAJ2004	60	4,000	54,282,812	48,081,380	1,060,005	4,189,164	74,251	6,091
HAJ2005	60	4,000	61,893,620	58,038,723	1,137,940	4,499,445	76,585	6,551
HAJ2006	60	4,000	66,510,634	58,753,898	1,222,533	4,476,766	76,255	5,954
HAJ2007	60	4,000	70,453,727	68,607,888	1,215,036	4,429,546	76,263	6,912

	Changing benchmarks over time according to formulation 3.3 and dual values (λ_{12}) for δ=0.7					
DMUs under review	Florence 2000	Hamburg 1998/9	Venice 2004/5	Genoa 2000	Nuremberg 2001/2/3/7	Vienna 1999
HAJ1998	0.46	0.29	0.25			
HAJ1999	0.43	0.34	0.23			
HAJ2000	0.28	0.28	0.43			
HAJ2001	0.15	0.29	0.33	0.23		
HAJ2002	0.28	0.17	0.37		0.18	
HAJ2003	0.13	0.22	0.27		0.38	
HAJ2004		0.35	0.21	0.35	0.09	
HAJ2005	0.37		0.32		0.03	0.28
HAJ2006		0.19	0.14		0.55	0.13
HAJ2007			0.18		0.57	0.25

Hanover's management may find it rather difficult to improve capacity utilization due to the airport's highly competitive location. The airport faces direct competition from Hamburg and Bremen that are in close proximity, as well as Dortmund, Paderborn and Münster-Osnabrück, which primarily serve charter and low cost carriers and are less than two hours drive by car, as shown in Fig. 8. Potential competition includes regional airports located in Braunschweig-Wolfsburg, Kassel-Calden and Magdeburg-Cochstedt, none of which currently operate commercially although plans exist to offer commercial flights from Kassel-Calden in 2012 (Flughafen Kassel-Calden [63]). Additional intermodal competition includes the ICE high speed rail alternative and a highly connected motorway network. In conclusion, Hanover faces direct, potential and intermodal

15 According to basic DEA, θ_1 decreased slightly from 54% in 1998 to 50% in 2006 and the benchmarks include Bremen, Dresden, Hamburg and Florence over the entire period.

competition hence the airport needs to cut costs by as much as 40%, further develop non-airport related activities and attempt to attract cargo throughput. The latter strategy may increase runway utilization and seems reasonable given the high level of connectivity of the city and the lack of night flights restrictions due to their location.

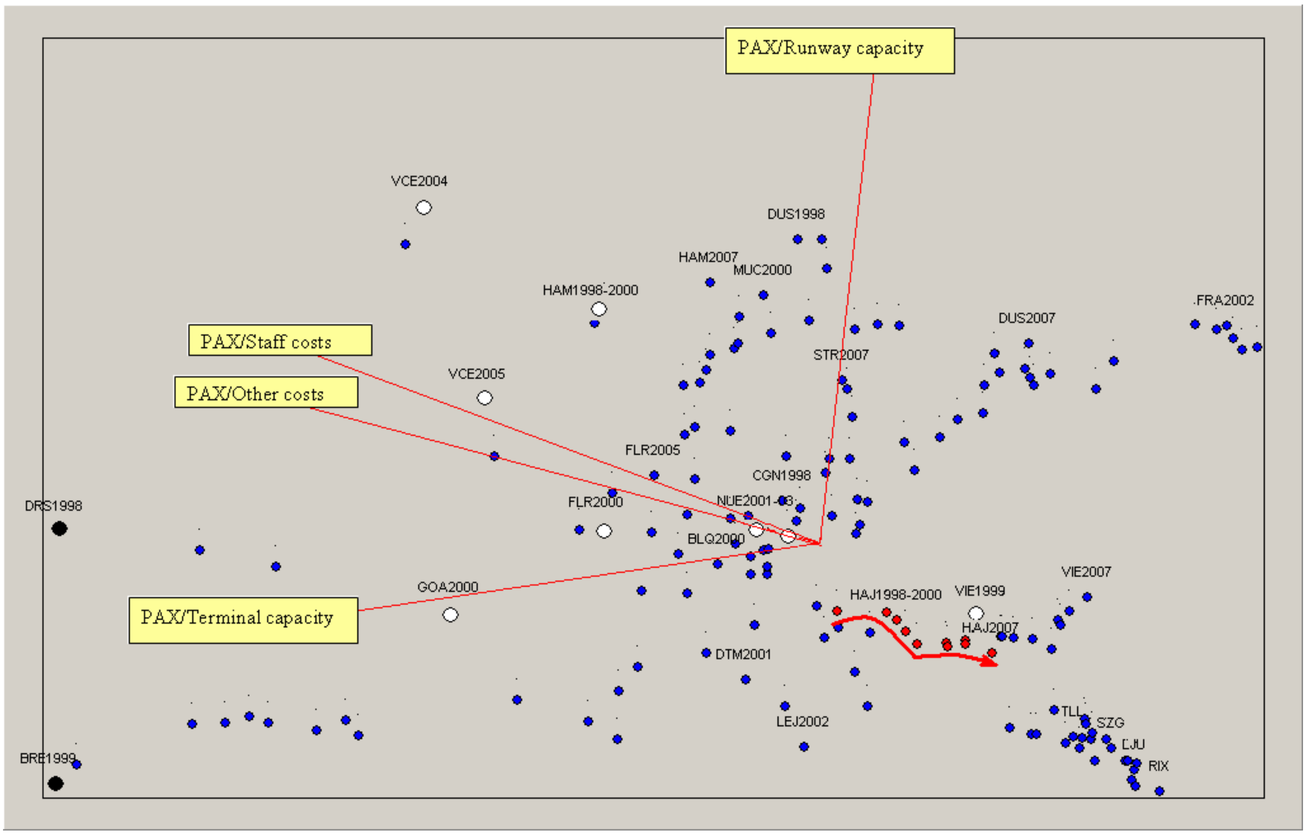


Fig. 7: Co-plot of input minimization results with emphasis on Hanover[16]

Fig. 8: Catchment area of Hanover airport (2 hour drive)[17]

16 Coefficient of alienation is 0.137 and average of correlations is 0.829.

Our final analysis presents the benchmarking results for Lyon Saint-Exupéry, which is the fourth largest airport in France, with a passenger throughput of 7.9 million in 2008 (ACI [52]). Similar to the majority of French regional airports, Lyon is fully publicly owned and operated by the regional Chamber of Commerce. Lyon airport became a major regional hub airport for the national carrier Air France at the end of the 1990s and today, Easyjet is their second largest customer (Lyon Aéroport [64]). In the independent revenue maximizing formulation (2), Lyon improved in aeronautical performance (θ_2) from a score of 2.09 to 1.02 between 1998 and 2005[18], benefiting from a change in the tariff structure similar to that of Vienna airport. The important peer airports include privatized BAA Glasgow and Basel-Mulhouse, both of which focus on low cost carrier traffic. Glasgow serves Easyjet and Scottish Loganair in competition with Ryanair at Prestwick, and Basel-Mulhouse serves Easyjet, which achieved a market share of almost 50% in 2007 (Flughafen Basel-Mulhouse [65]). With respect to non-aeronautical activities, Lyon increased its performance (θ_1) from 1.68 in 1998 to 1.41 in 2005, in part due to the large increase in car parking revenues, rents and utilities which contributed to a 67% increase in overall commercial revenues (see Table 4). Benchmarks on the non-aeronautical side include Basel-Mulhouse and Marseille, with the former generating more than 50% of their revenue from commercial sources, the majority of which are derived from retail sales, rents and utilities.

17 Source: adapted from ADV [66].

18 Lyon's benchmark clusters are stable over time according to formulation 3.1.

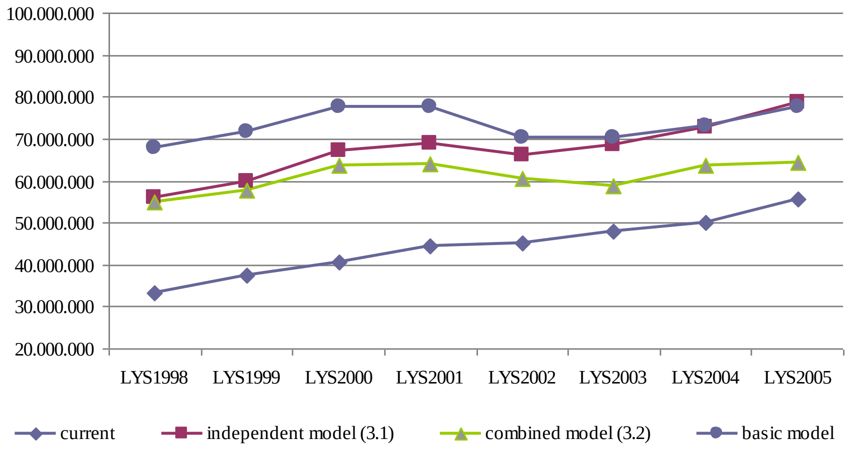
Table 4: Output benchmarks for Lyon airport

Airport		Domestic Passengers	International Passengers	Movements	Cargo	Aeronautical Revenues (US$)	Non-aeronautical Revenues (US$)
DMUs under review							
LYS1998		2.478.508	2.742.712	106.170	39.749	25.552.145	33.325.320
LYS1999		2.565.033	2.935.515	116.894	39.050	29.032.107	37.403.371
LYS2000		2.715.196	3.311.666	129.373	40.126	35.476.207	40.519.260
LYS2001		2.714.678	3.393.929	132.903	38.902	41.348.698	44.483.257
LYS2002		2.523.982	3.254.242	120.529	35.349	48.933.813	44.982.158
LYS2003		2.571.177	3.368.718	118.489	35.494	58.338.026	47.915.432
LYS2004		2.633.962	3.594.650	123.958	34.874	61.171.391	50.077.362
LYS2005		2.682.123	3.879.242	128.868	38.725	68.845.652	55.678.876
Output benchmarks for Lyon airport in 2005							
Benchmark	Intensity (λ)	Domestic Passengers	International Passengers	Movements	Cargo	Aeronautical Revenues (US$)	Non-aeronautical Revenues (US$)
Short-term benchmark for non-aeronautical activities (model 3.2, θ_1=1.15)							
MLH2004	0.98	651.102	1.893.772	57.915	34.227	30.882.599	38.867.805
GLA2000	0.29	3.568.259	3.453.741	92.000	10.000	69.790.936	38.013.125
GLA2005	0.16	4.604.022	4.237.878	97.610	9.461	76.125.667	63.356.734
NCE2006	0.06	4.332.382	5.615.653	164.617	13.940	100.059.952	83.755.849
Medium-term benchmark for non-aeronautical activities (model 3.1, θ_1=1.4)							
MLH2004	1	651.102	1.893.772	57.915	34.227	30.882.599	38.867.805
MLH2002	0.61	792.765	2.264.199	88.000	31.285	34.865.786	45.079.574
MRS1998	0.39	3.943.382	1.568.411	87.030	55.993	19.795.312	31.681.597
Medium-term benchmark for aeronautical activities (model 3.1, θ_2=1.03)							
MLH2004	0.96	651.102	1.893.772	57.915	34.227	30.882.599	38.867.805
GLA2000	0.52	3.568.259	3.453.741	92.000	10.000	69.790.936	38.013.125
NCE2006	0.05	4.332.382	5.615.653	164.617	13.940	100.059.952	83.755.849
Long-term benchmark for both activities (basic DEA, θ=1.4)							
LCY2002	0.4	417.551	1.187.449	53.000	1.000	38.509.245	12.353.716
MLH2002	0.35	792.765	2.264.199	88.000	31.285	34.865.786	45.079.574
OSL2007	0.19	9.477.511	9.566.489	223.000	97.000	177.975.321	245.588.135
ATH2007	0.08	5.953.814	10.571.571	205.295	119.000	453.224.152	137.319.560

Given that aeronautical revenue maximization is not necessarily an optimal policy, irrespective of ownership form, the combined formulation (3) defines aeronautical revenue as a non-discretionary output and maximizes commercial revenue alone. The results of this model suggest that Lyon's short-term commercial revenue target should be $63 million, an increase of 15%, given current aeronautical revenues. The medium-term target (formulation 2) suggests an increase of 40% in non-aeronautical revenues to increase performance and the longer term, standard DEA target requires the same increase of 40% both on the commercial and non-aeronautical side respectively (Fig. 9). In the combined model, Basel-Mulhouse appears as an important benchmark and Glasgow acts as a benchmark of increasing intensity over the years. As also shown in Fig. 10, Lyon airport is moving in the direction of Basel-Mulhouse and Glasgow hence is increasing in performance over time.

Marseille no longer appears as a benchmark in the results of the combined model since the airport generates substantially lower aeronautical revenues in comparison.

Non-aeronautical revenues



Aeronautical revenues

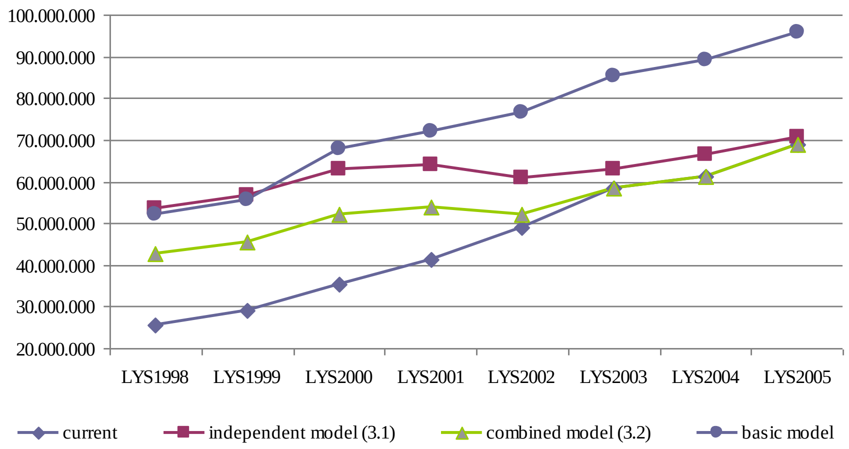


Fig. 9: Current and target output values for the independent and combined network model with respect to Lyon

In summary, Lyon's aeronautical revenues hit their short-term targets by 2003 and their medium-term benchmarks by 2005 (Fig. 9), suggesting that their landing and passenger charges are sufficient and should not be increased further. However, Lyon could still optimize commercial revenues in order to increase managerial productivity and better manage their joint revenue streams. The targets obtained from the basic DEA results would be very challenging in the short- or medium-term and should therefore be considered only as a long-term target if at all. As shown in the graphs of Fig. 9, several paths to the Pareto frontier can be defined for the airport, in which both aeronautical and non-aeronautical revenues can be expanded, either equi-proportionally, or with greater emphasis on the non-aeronautical revenues such that the airport remains profit maximizing.

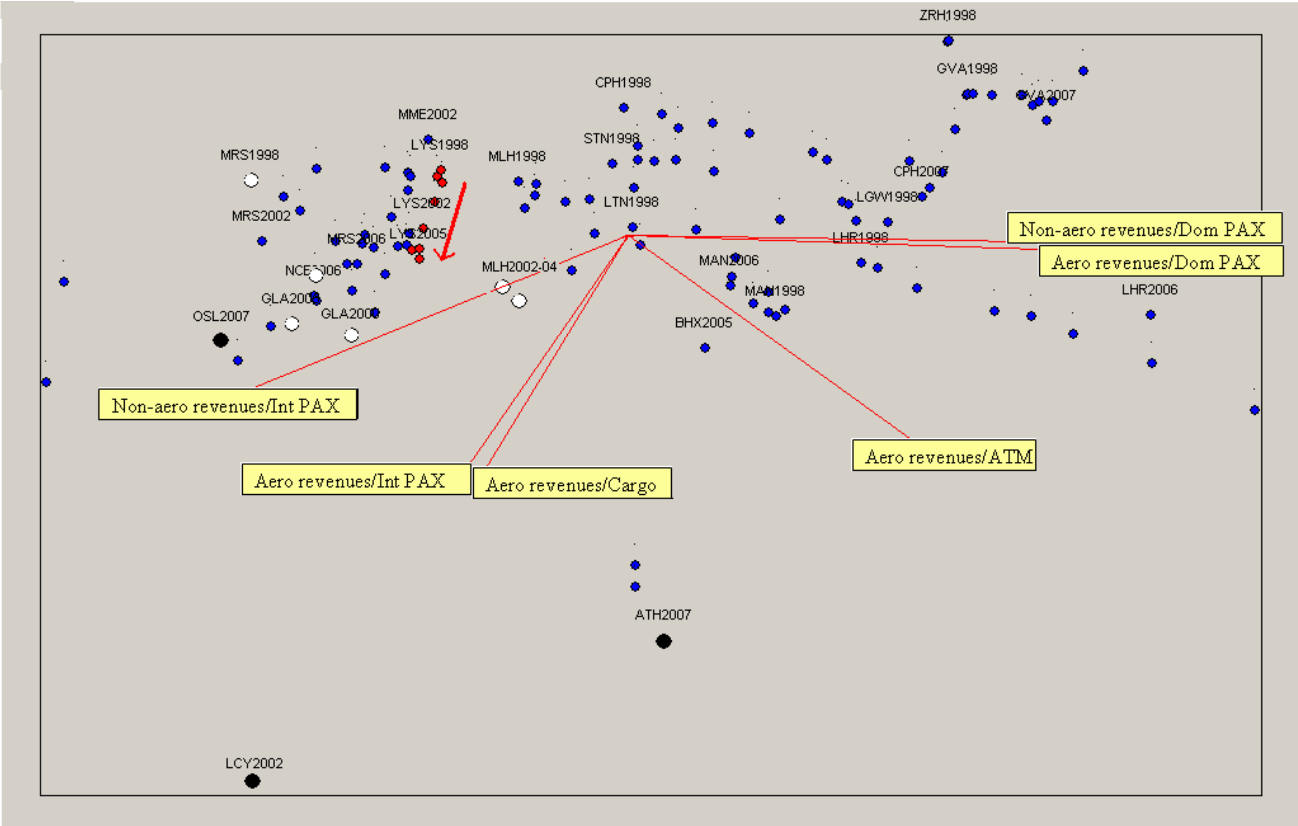


Fig. 10: Co-plot of output maximization results with emphasis on Lyon[19]

6. Conclusions and directions for future research

DEA has been repeatedly applied to the study of airport productivity however the basic DEA models treat the airport technology as a black box, which reduces the usefulness of the model for purposes of benchmarking. The focus of this paper has been to model the airport production process from a managerial perspective in order to provide a set of models that would aid benchmarking by applying a network DEA model. Usually network-DEA is applied to determine both the efficiency of sub processes and the overall efficiency (Golany et al. [36]) whereas in our research, network DEA helps decision-makers to describe the production process, demonstrating the sequential effects separating final and intermediate outputs including those under partial managerial control and those

19 Coefficient of alienation is 0.107 and average of correlations is 0.815.

that are known to be non-discretionary. Consequently, the approach connects aeronautical and commercial activities via intermediate products.

To improve the set of peer airports chosen, a dynamic clustering mechanism limits DEA's dual variables (benchmark intensities) was used, ensuring appropriate comparability within the dataset. The dynamic clustering approach proposed by Golany and Thore [25] restricts the selection of best practice DMUs according to predefined boundaries within the basic DEA framework. We extended this method by using integer linear programming which forms reference sets based on similar mixes of inputs or outputs and intermediate products. As a result each DMU optimizes only the last stage of the network, taken into account the information from previous stages. In addition, PCA-DEA is applied to reduce the number of variables when clusters are too small to avoid the curse of dimensionality. By identifying individual reference sets using dynamic clustering we provide individual benchmarks for inefficient DMUs, permitting identification of strategy changes over time and uniqueness with respect to economic regulation and airport infrastructure. The formulation was further adapted to ensure partial flexibility with respect to an expensive and complicated infrastructure system. Finally, the provision of ground handling was shown to severely affect efficiency estimates, leading to a separation in the comparison of those airports that undertake the process in-house compared to those that outsource.

Data proved to be the most difficult issue for this application. After defining salient variables (as in Fig. 1), we were then forced to reduce the model drastically in the light of data availability issues (as in Fig. 3). It would be extremely helpful were government organizations, such as the International Civil Aviation Organization, to standardize the data collection and publish data openly as this would enable fair and transparent comparisons. However, the results have shown that compared with the basic DEA approach, network DEA formulations provide more appropriate benchmarks which may enable airport managers to improve performance in the short and medium-term. In the case of Hanover, we show that in the short or medium term it is sufficient to close one of the three existing runways or expand their cargo operations to increase utilization, whereas basic DEA benchmarks require the airport to close the equivalent of two runways. In the case of Lyon we demonstrate that in the short-term the airport earns a sufficient level of aeronautical revenues and simply needs to focus on improvements on the commercial side. In comparison, the results of basic DEA require Lyon to increase aeronautical revenues by 40% in order to operate efficiently. In contrast to the results from basic DEA, we cannot conclude in our network DEA model that DMUs acting as benchmarks are necessarily efficient since they may deviate from the overall Pareto frontier if limited by the dynamic clustering restrictions. On the other hand, inefficient units may be consistently defined as relatively inefficient.

To be in a position to undertake benchmarking exercises requires the collection and publication of airport related data provided openly at the federal level since such information would be of public interest. Furthermore, an airport also produces undesirable outputs such as delays. Besides the capacity utilization which has been considered in our research, delay substantially affects airport and airline performance and should clearly be included in a benchmarking study. For improved managerial benchmarking, disaggregated data with regard to non-aeronautical activities would help to identify successful strategies on the commercial side. Other factors that are beyond managerial control include the competitive environment, ownership structure and economic regulation. These aspects influence managerial behaviour, and accounting for them may further improve comparability and permit the relevant authorities to analyze the impact of cost or incentive based regulation on managerial performance. In order to assess technological changes over the ten-year period, Malmquist DEA may be applied in future research. Furthermore, an airport is typical of an industry with lumpy investments, hence time lags between investment and changes in productivity should be considered, which would require adapting the formulations developed here in order to account for time.

Acknowledgements

We would like to thank the German Airport Performance (GAP) and German Aviation Benchmarking (GAB) research projects that are funded by the German Ministry of Education and Research for providing the data. Further, we thank Dr. Thomas Immelman of Hamburg Airport, Prof. Mike Tretheway of Intervistas, Prof. Gert Brunekreeft, Prof. Hans-Martin Niemeier, Prof. David Gillen, Prof. Peter Forsyth and the participants of the German Aviation Research Society (GARS) seminar on Airport Benchmarking in November 2009, the World Bank roundtable on Transport Productivity in February 2010 and the INFORMS Conference 2010 for fruitful discussions. Nicole Adler would also like to thank the Recanati Fund for partial funding of this research.

References

[1] Zhang A, Zhang Y (2003) Airport Charges and Capacity Expansion: Effects of Concessions and Privatization. *Journal of Urban Economics* 53:54-75

[2] Oum TH, Adler N, Yu C (2006) Privatization, corporatization, ownership forms and their effects on the performance of the world's major airports. *Journal of Air Transport Management* 12:109–121

[3] Liebert V, Neimeier H-M (2010) Benchmarking of Airports - A Critical Assessment. World Conference on Transport Research Proceedings

[4] Hooper PG, Hensher DA (1997) Measuring Total Factor Productivity of airports - an index number approach. *Transportation Research part E* 33:249-259

[5] Oum TH, Yu C (2004) Measuring airports' operating efficiency: a summary of the 2003 ATRS global airport benchmarking report. *Transportation Research part E* 40:515-532

[6] Vasigh B, Gorjidooz J (2006) Productivity Analysis of public and private airports: a causal investigation. Journal *of Air Transportation* 11:144-163

[7] Pels E, Nijkamp P, Rietveld P (2003) Inefficiencies and scale economies of European airport operations. *Transportation Research part E* 39:341–361

[8] Oum TH, Yan J, Yu C (2008) Ownership forms matter for airport efficiency: A stochastic frontier investigation of worldwide airports. *Journal of Urban Economics* 64:422-435

[9] Gillen D, Lall A (1997) Developing measures of airport productivity and performance: an application of data envelopment analysis. *Transportation Research part E* 33:261 – 273

[10] Sarkis J (2000) An analysis of the operational efficiency of major airports in the United States. *Journal of Operations Management* 18:335–351

[11] Parker D (1999) The performance of the BAA before and after privatization. *Journal of Transport Economics and Policy* 33:133–146

[12] Martín JC, Román C (2008) The relationship between size and efficiency: A benchmarking analysis of Spanish commercial airports. *Journal of Air Transport Management* 2: 183-197

[13] Murillo-Melchor C (1999) An analysis of technical efficiency and productive change in Spanish airports using the Malmquist Index. *International Journal of Transport Economics* 26:271–292

[14] Abbott M, Wu S (2002) Total factor productivity and efficiency of Australian airports. *Australian Economic Review* 35:244–260

[15] Yu MM (2004) Measuring physical efficiency of domestic airports in Taiwan with undesirable outputs and environmental factors. *Journal of Air Transport Management* 10:295–303

[16] Barros CP, Sampaio A (2004) Technical and Allocative Efficiency of Airports. *International Journal of Transport Economics* 31:355-377

[17] Adler N, Berechman J (2001) Measuring airport quality from the airlines viewpoint: An application of data envelopment analysis. *Transportation Policy* 8:171–181

[18] Lin LC, Hong, CH (2006) Operational performance evaluation of international major airports: an application of data envelopment analysis. *Journal of Air Transport Management* 12: 342–351

[19] Martín JC, Román C, Voltes-Dorta A (2009) A stochastic frontier analysis to estimate the relative efficiency of Spanish airports. *Journal of Productivity Analysis* 31:163-176

[20] Graham A, Holvad T (2000) Efficiency Measurement for Airports. Paper presented at the Trafik Dags PAA Aalborg Universitet 2000 Conference, Aalborg University

[21] Morrison WG (2009) Understanding the complexities and challenges of airport performance benchmarking. *Journal of Air Transport Management* 3:145–158

[22] Färe R (1991) Measuring Farrell efficiency for a firm with intermediate inputs. *Academia Economic Papers* 19:329–340

[23] Oum TH, Yu C, Fu X (2003) A comparative analysis of productivity performance of the world's major airports: summary report of the ATRS global airport benchmarking research report—2002. *Journal of Air Transport Management* 9:285–297

[24] Zhang A, Zhang Y (2010) Airport capacity and congestion pricing with both aeronautical and commercial operations. *Transportation Research part B* 44: 404-413

[25] Golany B, Thore S (1997) Restricted best practice selection in DEA: An overview with case study evaluating the socio-economic performance of nations. *Annals of Operations Research* 73:117-140

[26] Banker RD, Morey RC (1986) Efficiency Analysis for Exogenously Fixed Inputs and Outputs. *Operations Research* 34:513-521

[27] Adler N, Golany B (2001) Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe. *European Journal of Operational Research* 132:18-31

[28] Adler N, Yazhemsky E (2010) Improving discrimination in data envelopment analysis: PCA–DEA or variable reduction. *European Journal of Operational Research* 202:273–284
[29] Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research* 2:429–444
[30] Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management* Science 30:1078-1092
[31] Sarkis J, Talluri S (2004) Performance based clustering for benchmarking of US airports, *Transportation Research part A* 38:329-346
[32] Färe R, Grosskopf S (1996) Productivity and intermediate products: A frontier approach. *Economic Letters* 50: 65-70
[33] Färe R, Grosskopf S (2000) Network DEA. *Socio Economic Planning Sciences* 34:35–49
[34] Lewis HF, Sexton TR (2004) Network DEA: efficiency analysis of organizations with complex internal structure. *Computers and Operations Research* 31:1365-1410
[35] Emrouznejad A, Thanassoulis E (2005) A mathematical model for dynamic efficiency using data envelopment analysis. *Applied Mathematics and Computation* 160:363-378
[36] Golany B, Hackman ST, Passy U (2006) An efficiency measurement framework for multi-stage production systems. *Annals of Operations Research* 145:51–68
[37] Chen CM (2009) A network-DEA model with new efficiency measures to incorporate the dynamic effect in production networks. *European Journal of Operational Research* 194:687-699
[38] Kao C (2009) Efficiency decomposition in network data envelopment analysis: A relational model. *European Journal of Operational Research* 192:949-962
[39] Tone K, Tsutsui M (2009) Network DEA: A slacks-based measure approach. *European Journal of Operational Research* 197:243-252
[40] Castelli L, Pesenti R, Ukovich W (2010) A classification of DEA models when the internal structure of the Decision Making Units is considered. *Annals of Operations Research* 173:207-235.
[41] Yu MM, Lin ET (2008) Efficiency and effectiveness in railway performance using a multi-activity network DEA model. Omega-International Journal of Management Science 36:1005–1017
[42] Beasley J (1995) Determining teaching and research efficiencies. *Journal of the Operational Research Society* 46:441-452
[43] Cook WD, Hababou M, Tuenter HJH (2000) Multicomponent efficient measurement and shared inputs in data envelopment analysis: An application to sales and service performance in bank branches. *Journal of Productivity Analysis* 14:209-224
[44] Adler N, Golany B (2002) Including Principal Components Weights to improve discrimination in Data Envelopment Analysis. *Journal of the Operational Research Society* 53:985-991
[45] Adler N, Raveh A, Yazhemsky E (2007) DEA presented graphically using multi-dimensional scaling. In: Cook W, Zhu J (ed) Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis. Springer, New York, pp 171-187
[46] Adler N, Raveh A (2008) Presenting DEA graphically. *Omega-International Journal of Management Science* 36:715–729
[47] Brockett PL, Golany B (1996) Using rank statistics for determining programming efficiency differences in data envelopment analysis. *Management Science* 42:466-472
[48] Sueyoshi T, Aoki S (2001) A use of a nonparametric statistic for DEA frontier shift: the Kruskal and Wallis rank test. *Omega-International Journal of Management Science* 29:1-18
[49] Banker RD (1993) Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Science* 39: 1265-1273
[50] Guttman L (1968) A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika* 33:469–506
[51] Lozano S, Salmerón JL (2010) Network DEA models for air transport efficiency assessment. 14th Air Transport Research Society Proceedings
[52] ACI (2009) World Airport Traffic Report 2008. ACI World, Geneva
[53] Graham A (2005) Airport benchmarking: a review of the current situation. *Benchmarking International Journal* 12:99–111
[54] International Air Transport Association (2010) World Scheduling Guidelines, 19th Edition, http://www.iata.org/NR/Content Connector/ CS2000/SiteInterface/sites/whatwedo/scheduling/file/fdc/WSG-12thEd.pdf. Accessed 17 March 2010
[55] Aeroporto di Firenze (2010) Technical Data. http://www.aeroporto.firenze.it
[56] International Air Transport Association (2003) Airport Capacity/Demand Profiles - 2003 Edition. Air Transport Consultancy Services, Monetral-Geneva
[57] Dietz P (2009) Verluste auf dem Rollfeld. Frankfurter Rundschau 11 November 2009
[58] Hutter D (2009) 2000 Mitarbeitern droht die Kündigung. Süddeutsche Zeitung 29 May 2009

[59] Cologne Bonn Airport (2010) Press Office. http://www.airport-cgn.de/main.php?lang=2&id=140. Accessed 18 March 2010
[60] Andersen P, Petersen NC (1993) A Procedure for Ranking Efficient Units in Data Envelopment Analysis. *Management Science* 39: 1261-1264
[61] Dortmund Airport (2007) Geschäftsbericht 2007. Flughafen Dortmund GmbH
[62] Vienna International Airport (2001, 2004, 2007) Geschäftsbericht. Flughafen Wien AG
[63] Flughafen Kassel-Calden (2010) Ausbau. http://www.flughafenkassel.de/t3. Accessed 18 March 2010
[64] Lyon Aéroport (2008) Les indicateurs trafic de l'aéroport de Lyon-Saint Exupéry. Aéroport Lyon - Saint Exupéry
[65] Flughafen Basel-Mulhouse (2007) 2007-The Year-Key Figures. Flughafen Basel Mulhouse
[66] ADV (2010) Mitglieder/ Flughafenkarte. http://www.adv.aero/mitglieder.html. Accessed 19 March 2010
[67] SH&E (2002) Appendices: Study on the quality and efficiency of ground handling services at EU airports as a result of the implementation of Council Directive 96/67/EC. Report to European Commission, SH&E International Air Transport Consultancy, London

Appendix: Airport Dataset[20]

Code	Airport	Country	Time Period	Ground Handling
ABZ	Aberdeen	UK	1999	not provided
AMS	Amsterdam	Netherlands	1998-2007	not provided
ATH	Athens	Greece	2005-2007	not provided
BHX	Birmingham	UK	2005	not provided
BLQ	Bologna	Italy	2000-2005	provided
BRE	Bremen	Germany	1998-2007	provided
CGN	Cologne-Bonn	Germany	1998-2007	provided
CPH	Copenhagen	Denmark	1998-2007	not provided
DRS	Dresden	Germany	1998-2004	provided
DTM	Dortmund	Germany	2001-2007	provided
DUS	Dusseldorf	Germany	1998-2007	provided
FLR	Florence	Italy	2000-2005	provided
FRA	Frankfurt	Germany	2002-2007	provided
GLA	Glasgow	UK	1999-2006	not provided
GOA	Genoa	Italy	2000-2005	provided
GVA	Geneva	Switzerland	1998-2007	not provided
HAJ	Hanover	Germany	1998-2007	provided
HAM	Hamburg	Germany	1998-2007	provided
LBA	Leeds-Bradford	UK	1998-2006	not provided
LCY	London City	UK	2002	not provided
LEJ	Leipzig	Germany	1998-2002	provided
LGW	London Gatwick	UK	1998-2002	not provided
LHR	London Heathrow	UK	1998-2006	not provided
LJU	Ljubljana	Slovenia	2004-2007	provided
LTN	London Luton	UK	1998-2005	not provided
LYS	Lyon	France	1998-2005	not provided
MAN	Manchester	UK	1998-2006	not provided
MLA	Malta	Malta	2005-2006	not provided
MLH	Basel-Mulhouse	France	1998-2007	not provided
MME	Durham Tees Valley	UK	2002	not provided
MRS	Marseille	France	1998-2006	not provided
MUC	Munich	Germany	1998-2007	provided
NCE	Nice	France	1998-2006	not provided
NUE	Nuremberg	Germany	1998-2007	provided
OSL	Oslo	Norway	1999-2007	not provided
RIX	Riga	Latvia	2004-2006	provided
STN	London Stansted	UK	1998-2006	not provided
STR	Stuttgart	Germany	1998-2007	provided
SZG	Salzburg	Austria	2004-2007	provided
TLL	Tallinn	Estonia	2002-2007	provided
VCE	Venice	Italy	2000-2005	provided
VIE	Vienna	Austria	1998-2007	provided
ZRH	Zurich	Switzerland	1998-2007	not provided

20 Source: adapted from SH&E [67] and Airport Websites.