Comparison of Data Envelopment Analysis Methods Used in Airport Benchmarking

David Schaar and Lance Sherry, Ph.D.
Dept. of Systems Engineering and Operations Research
George Mason University
Fairfax, Virginia, USA

Abstract—Airport efficiency has been shown to contribute to the overall performance efficiency of the air transportation network. Airport performance efficiency is benchmarked annually and widely published. These benchmarks use several techniques, including several Data Envelopment Analysis (DEA) methods.

This paper examines the differences in results using three DEA methods (Cooper-Charnes-Rhodes (CCR), Banker-Charnes-Cooper (BCC), and Slacks-Based Measure of efficiency (SBM)) on data from 45 airports from 1996 to 2000. The results from the three DEA methods yielded wide variation in results. For example, the CCR analysis showed that efficiencies degraded from small to medium to large airports. The BCC analysis showed no significant difference in efficiency among the three classes of airports. The SBM analysis yielded degraded efficiency from large to medium to small airports. The implications of these results on the use of DEA in benchmarking and the need for guidelines for selection of DEA models and the interpretation of DEA results is discussed.

Keywords-Data Envelopment Analysis; CCR; BCC; SBM; airport efficiency

I. INTRODUCTION

Airport congestion is growing and traveler satisfaction is dropping [1], yet demand is expected to grow at major hub airports [2].

In this environment, the operational efficiency of airports is one of the important determinants of the system’s future success. An important component to achieving operational efficiency is to use performance measurement to understand which airports are performing well and which are underperforming and also to understand what the drivers are behind good and bad performance, all with the ultimate goal of improving performance in the areas that are deficient.

As we will show in this paper, much has been written in the academic literature about measuring airport efficiency but several opportunities for improvement exist in this work.

One such opportunity relates to the theoretical foundations for evaluating performance: A number of studies have been conducted where airports are ranked against each other. However, in these studies there has been limited focus on the selection of the evaluation models used to identify good and bad performers. Several choices exist in this area and a study of each model is necessary to understand which is the most appropriate in a given situation.

This paper will highlight this issue through the use of an example. We will re-run an earlier analysis of airport performance and highlight the impact of the choices in the areas listed above.

II. BACKGROUND

A. General Background

Performance benchmarking as a management tool is frequently cited as having been pioneered by Robert C. Camp at the Xerox Corporation in the late 1970’s [7],[8] where he used this tool to identify shortcomings of that company’s photocopier production and distribution. He used benchmarking to identify several key practices from the photocopier industry as well as from other industries, all serving to ultimately improve the company’s photocopier business. The performance measurement and benchmarking techniques have all evolved significantly since that time, but the ultimate goals for any benchmark remain the same: To identify performance gaps and to find practices that will help close that gap.

Benchmarking has since been applied both in industry and academia in numerous studies. Dattakumar and Jagadeesh [6] conducted an extensive literature review and found more than 350 publications on the topic as of June 2002.

Airport benchmarking studies began appearing in the literature in the early 1990’s with one of the first such studies being that of Tolofari, Ashford, and Caves [9] in which airports operated by the British Airport Authority were compared against each other. A series of benchmarking studies have since appeared with a variety of geographic foci and covering a diverse set of performance parameters. A few examples of these studies are listed in Table I.

TABLE I.
As is clear from Table I, besides being used in the analysis that will be examined here, Data Envelopment Analysis (DEA) is an analytical technique that is frequently used in benchmarking studies. DEA is a non-parametric methodology used to assess the efficiency of a Decision-Making Unit (DMU – e.g. an airport) in converting a set of inputs into outputs.

A number of different versions of the basic DEA model have been developed to address a series of potential shortcomings of the original DEA model. We will examine three variations of the DEA methodology and study the differences in assumptions of the three as well as compare the different outcomes of each. The purpose of this analysis is to understand whether the choice of methodology may have an impact on the outcome of an analysis.

### 2. CCR model

In all variations of the DEA models, the DMU(s) with the best inherent efficiency in converting inputs $X_1, X_2, \ldots, X_n$ into outputs $Y_1, Y_2, \ldots, Y_m$ is identified, and then all other DMUs are ranked relative to that most efficient DMU.

For DMU $0$, the basic DEA model (so-called CCR after Charnes, Cooper, and Rhodes [5]) is calculated as follows:

$$\max h_0 = \frac{\sum_{j=0}^{m} u_j y_{0j}}{\sum_{i=0}^{n} v_i x_{0i}}$$

subject to

$$\sum_{j=0}^{m} v_j x_{ij} \leq 1 \text{ for each unit } j$$

$$u_j, v_i \geq 0$$
The interpretation of $u_r$ and $v_i$ is that they are weights applied to outputs $y_{rj}$ and inputs $x_{ij}$ and the are chosen to maximize the efficiency score $h_0$ for DMU 0. The constraint forces the efficiency score to be no greater than 1 for any DMU. An efficiency frontier is calculated, enveloping all datapoints in a convex hull. The DMU(s) located on the frontier represent an efficiency level of 1.0, and those located inside the frontier are operating at a less than full efficiency level, i.e. less than 1.0.

The above fractional program is executed once for each participating DMU, resulting in the optimal weights being determined for each DMU.

Before solving the problem, the denominator in the objective function is removed and instead an additional constraint is added. Also, the original constraint is manipulated in order to convert the fractional program to a linear program. These two steps result in the following linear program:

$$\text{max } h_0 = \sum_{r} u_r y_{r0}$$

$$\sum_{r} u_r y_{rj} - \sum_{i} v_i x_{ij} \leq 0$$

subject to

$$\sum_{i} v_i x_{ij} = 1$$

$$u_r, v_i \geq 0$$

In simpler notation, this is written as:

$$\max(v,u) = uy_0$$

$$-vX + uY \leq 0$$

subject to

$$vX_0 = 1$$

$$v \geq 0, u \geq 0$$

Finally, before solving, the linear program is converted to its dual for computational efficiency reasons:

$$\min(h_0) = \sum_{r} u_r y_{r0}$$

$$\sum_{r} u_r y_{rj} - \sum_{i} v_i x_{ij} \leq 0$$

subject to

$$\sum_{i} v_i x_{ij} = 1$$

$$u_r, v_i \geq 0$$

With the addition of slack variables, the dual problem becomes:

$$\min(\theta, \lambda) = \theta$$

$$\theta x_0 - X\lambda = s^-$$

subject to

$$Y\lambda = y_0 + s^+$$

$$\lambda \geq 0, s^+ \geq 0, s^- \geq 0$$

The slack variables can be interpreted as the output shortfall and the input overconsumption compared to the efficient frontier.

C. BCC model

The CCR model is designed with the assumption of constant returns to scale. This means that there is no assumption that any positive or negative economies of scale exist. It is assumed is that a small airport should be able to operate as efficiently as a large one – that is, constant returns to scale. In order to address this, Banker, Charnes, and Cooper developed the BCC model [3].

The BCC model is closely related to the standard CCR model as is evident in the dual of the BCC model:

$$\min(\theta, \lambda) = \theta$$

$$\theta x_0 - X\lambda = s^-$$

subject to

$$Y\lambda = y_0 + s^+$$

$$e\lambda = 1$$

$$\lambda \geq 0, s^+ \geq 0, s^- \geq 0$$

The difference compared to the CCR model is the introduction of the convexity condition $e\lambda = 1$. This additional constraint gives the frontiers piecewise linear and concave characteristics.

D. SBM model

Finally, the second adjustment to the basic CCR model is the Slacks-Based Measure of efficiency (SBM), proposed by Tone [18]. The motivation for the development of this model is the observation that while both the CCR and the BCC models calculate efficiency scores, neither is able to take into account the resulting amount of slack for inputs and outputs. Consequently, the purpose of this model is to minimize the input and output slacks, resulting in this fractional program, which is converted to a linear program before solving:

$$\min(s^+, s^-) \rho =$$

$$1 - \frac{1}{m} \sum_{i=1}^{m} s_i^- / x_{i0}$$

$$1 + \frac{1}{s} \sum_{r=1}^{s} s_r^+ / y_{r0}$$
\[ x_0 - X\lambda = s^- \]
subject to \[ Y\lambda = y_0 + s^+ \]
\[ \lambda \geq 0, s^+ \geq 0, s^- \geq 0 \]  \hspace{1cm} (7)

III. METHOD OF ANALYSIS

A. Introduction

The analysis presented here is based on a re-run and extension of the analysis performed by Bazargan and Vasigh [11].

B. Inputs and Outputs

Input and output data was collected from a series of sources in order to assemble the same inputs and outputs as in the Bazargan and Vasigh study for the same airports and years (1996 – 2000). Table II lists the airports covered in the analysis and Table III lists the input and output variables and the source from which they were collected.

<table>
<thead>
<tr>
<th>TABLE II. AIRPORTS INCLUDED IN STUDY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
</tr>
<tr>
<td>ATL</td>
</tr>
<tr>
<td>DEN</td>
</tr>
<tr>
<td>DFW</td>
</tr>
<tr>
<td>DTW</td>
</tr>
<tr>
<td>EWR</td>
</tr>
<tr>
<td>IAH</td>
</tr>
<tr>
<td>JFK</td>
</tr>
<tr>
<td>LAS</td>
</tr>
<tr>
<td>LAX</td>
</tr>
<tr>
<td>MIA</td>
</tr>
<tr>
<td>MSP</td>
</tr>
<tr>
<td>ORD</td>
</tr>
<tr>
<td>PHX</td>
</tr>
<tr>
<td>SFO</td>
</tr>
<tr>
<td>STL</td>
</tr>
</tbody>
</table>

C. Model Selections

In the Bazargan and Vasigh study, the model selected was the CCR model. No discussion was provided regarding why this model was selected ahead of others; presumably it was selected on the premise of being the “default” DEA model.

One of the purposes of the present study was to examine the impact of different model selections. Hence, efficiency scores were calculated using three different, basic DEA models: CCR, BCC, and SBM. These are considered to be among the standard DEA models by Cooper, Seiford, and Tone [19]. These models were all executed in output-oriented mode, meaning that efficiency scores were calculated by analyzing which levels of outputs should be “possible” to generate if an airport were operating at full efficiency given the current levels of inputs. These scores were calculated using the “Learning Version” of the DEA-Solver software [4].

D. Model Configuration

One of the particular choices made in the Bazargan and Vasigh study, driven by practical considerations, was to create an artificial, “super-efficient” airport by assigning it the minimum input values present in the dataset and the maximum output values. The reason for this was that when the CCR model was initially executed with only the basic list of airports included, a very large portion of the airports were ranked as fully efficient (1.0). The cause of that outcome was the relatively small number of airports (45) as compared to the number of parameters considered (10). In general, this will frequently be a phenomenon in analysis having a low ratio of DMUs to parameters.

Bazarghan and Vasigh also describe the theory of adding a condition requiring all weights \( u \) and \( v \) to be \( \geq \varepsilon \) (an infinitesimal value) in order to avoid setting all but one input and one output’s weights to a non-zero value. However, it is not indicated if this was used in the study and if it was, there is no specification of the value used for this infinitesimal value is provided. Consequently, the model will in this case be run without this additional constraint.

IV. RESULTS AND DISCUSSION

A. Comparison of CCR Model Runs

The results of the CCR model runs are shown in Table IV, Table V, and Table VI. These values are in general lower than the values calculated in the Bazargan and Vasigh study and the reason for this can not be ascertained since the raw data of the previous study were not available. It can be postulated that the difference is due to differences in the inputs; in particular if the values assigned to the artificial, “super-efficient” airports were different in the two studies, since that would cause a general offset between the two.

What is noteworthy, however, is that the general trend from the Bazargan and Vasigh study is confirmed: Large airports are in general exhibiting lower efficiency scores than medium-sized airports, and medium-sized airports are
exhibiting lower scores than small airports. These results are displayed in Figure 1 and Figure 2. The Kruskal-Wallis test was performed to determine whether the differences in efficiency ranks among the groups were significant, and that was found to be the case. The results from the Kruskal-Wallis test are presented in Table VII.

### Table IV. Results of CCR Runs for Large Airports

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ATL</td>
<td>0.1910</td>
<td>0.1990</td>
<td>0.2000</td>
<td>0.2000</td>
<td>0.2000</td>
</tr>
<tr>
<td>DEN</td>
<td>0.1515</td>
<td>0.1667</td>
<td>0.1677</td>
<td>0.1435</td>
<td>0.1281</td>
</tr>
<tr>
<td>DFW</td>
<td>0.1216</td>
<td>0.1231</td>
<td>0.1380</td>
<td>0.1199</td>
<td>0.1251</td>
</tr>
<tr>
<td>DTW</td>
<td>0.1405</td>
<td>0.1310</td>
<td>0.1256</td>
<td>0.1400</td>
<td>0.1432</td>
</tr>
<tr>
<td>EWR</td>
<td>0.2335</td>
<td>0.2595</td>
<td>0.2614</td>
<td>0.2511</td>
<td>0.2884</td>
</tr>
<tr>
<td>IAH</td>
<td>0.2202</td>
<td>0.2141</td>
<td>0.2149</td>
<td>0.2145</td>
<td>0.2238</td>
</tr>
<tr>
<td>JFK</td>
<td>0.2500</td>
<td>0.2491</td>
<td>0.2485</td>
<td>0.2500</td>
<td>0.2500</td>
</tr>
<tr>
<td>LAS</td>
<td>0.2034</td>
<td>0.2088</td>
<td>0.1977</td>
<td>0.2005</td>
<td>0.1946</td>
</tr>
<tr>
<td>LAX</td>
<td>0.2953</td>
<td>0.2995</td>
<td>0.2950</td>
<td>0.2698</td>
<td>0.2719</td>
</tr>
<tr>
<td>MIA</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.2500</td>
</tr>
<tr>
<td>MSP</td>
<td>0.2064</td>
<td>0.2015</td>
<td>0.1932</td>
<td>0.2155</td>
<td>0.2192</td>
</tr>
<tr>
<td>ORD</td>
<td>0.1667</td>
<td>0.1667</td>
<td>0.1667</td>
<td>0.1667</td>
<td>0.1667</td>
</tr>
<tr>
<td>PHX</td>
<td>0.2723</td>
<td>0.2765</td>
<td>0.2587</td>
<td>0.2674</td>
<td>0.2607</td>
</tr>
<tr>
<td>SFO</td>
<td>0.1784</td>
<td>0.1962</td>
<td>0.1783</td>
<td>0.1946</td>
<td>0.1804</td>
</tr>
<tr>
<td>STL</td>
<td>0.1937</td>
<td>0.2083</td>
<td>0.2032</td>
<td>0.2117</td>
<td>0.2134</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>0.2050</td>
<td>0.2100</td>
<td>0.2065</td>
<td>0.2064</td>
<td>0.2077</td>
</tr>
</tbody>
</table>

### Table V. Results of CCR Runs for Medium Sized Airports

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BNA</td>
<td>0.2140</td>
<td>0.2140</td>
<td>0.2179</td>
<td>0.2143</td>
<td>0.2186</td>
</tr>
<tr>
<td>CLE</td>
<td>0.2054</td>
<td>0.2132</td>
<td>0.2142</td>
<td>0.2085</td>
<td>0.2183</td>
</tr>
<tr>
<td>DAL</td>
<td>0.5603</td>
<td>0.5087</td>
<td>0.5257</td>
<td>0.5319</td>
<td>0.5262</td>
</tr>
<tr>
<td>IND</td>
<td>0.2811</td>
<td>0.3043</td>
<td>0.3044</td>
<td>0.3026</td>
<td>0.3067</td>
</tr>
<tr>
<td>MCI</td>
<td>0.2674</td>
<td>0.2816</td>
<td>0.2812</td>
<td>0.2836</td>
<td>0.2843</td>
</tr>
<tr>
<td>MDW</td>
<td>0.2361</td>
<td>0.2298</td>
<td>0.2296</td>
<td>0.2260</td>
<td>0.2296</td>
</tr>
<tr>
<td>MEM</td>
<td>0.2206</td>
<td>0.2070</td>
<td>0.2065</td>
<td>0.2256</td>
<td>0.2289</td>
</tr>
<tr>
<td>MSY</td>
<td>0.2765</td>
<td>0.2799</td>
<td>0.2869</td>
<td>0.2845</td>
<td>0.2916</td>
</tr>
<tr>
<td>OAK</td>
<td>0.4802</td>
<td>0.4704</td>
<td>0.5455</td>
<td>0.5147</td>
<td>0.5455</td>
</tr>
<tr>
<td>PDX</td>
<td>0.2518</td>
<td>0.2764</td>
<td>0.2778</td>
<td>0.2855</td>
<td>0.2767</td>
</tr>
<tr>
<td>RDU</td>
<td>0.4480</td>
<td>0.4152</td>
<td>0.4027</td>
<td>0.3749</td>
<td>0.3807</td>
</tr>
<tr>
<td>SJC</td>
<td>0.3022</td>
<td>0.3156</td>
<td>0.3051</td>
<td>0.3057</td>
<td>0.2951</td>
</tr>
<tr>
<td>SJU</td>
<td>0.3374</td>
<td>0.3967</td>
<td>0.4215</td>
<td>0.4155</td>
<td>0.4378</td>
</tr>
<tr>
<td>SMF</td>
<td>0.4121</td>
<td>0.4252</td>
<td>0.4171</td>
<td>0.4205</td>
<td>0.4019</td>
</tr>
<tr>
<td>SNA</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9914</td>
<td>1.0000</td>
<td>0.9882</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>0.3662</td>
<td>0.3692</td>
<td>0.3752</td>
<td>0.3729</td>
<td>0.3753</td>
</tr>
</tbody>
</table>

### Table VI. Results of CCR Runs for Small Airports

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ALB</td>
<td>0.4772</td>
<td>0.5101</td>
<td>0.5047</td>
<td>0.4895</td>
<td>0.5099</td>
</tr>
<tr>
<td>BHM</td>
<td>0.5345</td>
<td>0.5400</td>
<td>0.5482</td>
<td>0.5461</td>
<td>0.5698</td>
</tr>
<tr>
<td>BOI</td>
<td>0.8089</td>
<td>0.5714</td>
<td>0.7842</td>
<td>0.5392</td>
<td>0.8086</td>
</tr>
<tr>
<td>COS</td>
<td>0.5191</td>
<td>0.5561</td>
<td>0.5672</td>
<td>0.5725</td>
<td>0.5625</td>
</tr>
<tr>
<td>DAY</td>
<td>0.3472</td>
<td>0.3708</td>
<td>0.3674</td>
<td>0.3620</td>
<td>0.3652</td>
</tr>
<tr>
<td>ELP</td>
<td>0.6856</td>
<td>0.6810</td>
<td>0.6863</td>
<td>0.6876</td>
<td>0.6757</td>
</tr>
<tr>
<td>GEG</td>
<td>0.4460</td>
<td>0.4662</td>
<td>0.4726</td>
<td>0.5016</td>
<td>0.4695</td>
</tr>
<tr>
<td>GSO</td>
<td>0.8082</td>
<td>0.8553</td>
<td>0.8350</td>
<td>0.8115</td>
<td>0.8459</td>
</tr>
<tr>
<td>GUM</td>
<td>0.5714</td>
<td>0.5714</td>
<td>0.5714</td>
<td>0.5714</td>
<td>0.5714</td>
</tr>
<tr>
<td>LIT</td>
<td>0.8031</td>
<td>0.8385</td>
<td>0.8493</td>
<td>0.8590</td>
<td>0.8710</td>
</tr>
<tr>
<td>OKC</td>
<td>0.5854</td>
<td>0.6006</td>
<td>0.6053</td>
<td>0.6155</td>
<td>0.5961</td>
</tr>
<tr>
<td>ORF</td>
<td>0.3994</td>
<td>0.4555</td>
<td>0.4217</td>
<td>0.4122</td>
<td>0.4306</td>
</tr>
<tr>
<td>RIC</td>
<td>0.4214</td>
<td>0.4470</td>
<td>0.4403</td>
<td>0.4326</td>
<td>0.4439</td>
</tr>
<tr>
<td>ROC</td>
<td>0.4341</td>
<td>0.4572</td>
<td>0.4570</td>
<td>0.4321</td>
<td>0.4374</td>
</tr>
<tr>
<td>TUL</td>
<td>0.5098</td>
<td>0.5332</td>
<td>0.5373</td>
<td>0.5349</td>
<td>0.5472</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>0.5568</td>
<td>0.5636</td>
<td>0.5765</td>
<td>0.5578</td>
<td>0.5803</td>
</tr>
</tbody>
</table>

Figure 1. Results of Bazargan and Vasigh study
Figure 2. Results of CCR re-run of Bazargan and Vasigh study

Figure 3. Results of BCC analysis

Figure 4. Results of SBM analysis

B. Comparison of CCR, BCC, and SBM Results

In applying the three DEA models (CCR, BCC, and SBM) to the same dataset, some striking results emerged: While the CCR analysis showed that small airports were more efficient than both medium sized and large airports, and that medium sized airports were more efficient that large airports, in the BCC model no significant difference among the three groups was observed, and in the SBM dataset the difference was in fact reversed. These findings are displayed in the following figures and tables, where the average efficiency scores are displayed and the results of the Kruskal-Wallis tests on the efficiency ranks.

### TABLE VII. Results of Kruskal-Wallis Test on CCR Ranks

<table>
<thead>
<tr>
<th>Year</th>
<th>Large airports</th>
<th>Medium airports</th>
<th>Small airports</th>
<th>Chi square</th>
<th>Asymptotic significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>36.13</td>
<td>22.40</td>
<td>10.47</td>
<td>28.691</td>
<td>0.000</td>
</tr>
<tr>
<td>1997</td>
<td>36.07</td>
<td>23.07</td>
<td>9.87</td>
<td>29.848</td>
<td>0.000</td>
</tr>
<tr>
<td>1998</td>
<td>36.33</td>
<td>22.60</td>
<td>10.07</td>
<td>30.200</td>
<td>0.000</td>
</tr>
<tr>
<td>1999</td>
<td>36.33</td>
<td>22.53</td>
<td>10.13</td>
<td>30.200</td>
<td>0.000</td>
</tr>
<tr>
<td>2000</td>
<td>36.27</td>
<td>22.60</td>
<td>10.13</td>
<td>29.716</td>
<td>0.000</td>
</tr>
</tbody>
</table>

### TABLE VIII. Results of Kruskal-Wallis Test on BCC Ranks

<table>
<thead>
<tr>
<th>Year</th>
<th>Large airports</th>
<th>Medium airports</th>
<th>Small airports</th>
<th>Chi square</th>
<th>Asymptotic significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>18.53</td>
<td>23.47</td>
<td>27.00</td>
<td>3.149</td>
<td>0.207</td>
</tr>
<tr>
<td>1997</td>
<td>22.33</td>
<td>25.67</td>
<td>21.00</td>
<td>1.006</td>
<td>0.605</td>
</tr>
<tr>
<td>1998</td>
<td>23.33</td>
<td>24.97</td>
<td>20.70</td>
<td>0.808</td>
<td>0.668</td>
</tr>
<tr>
<td>1999</td>
<td>23.40</td>
<td>23.37</td>
<td>22.23</td>
<td>0.077</td>
<td>0.962</td>
</tr>
<tr>
<td>2000</td>
<td>21.40</td>
<td>22.63</td>
<td>24.97</td>
<td>0.572</td>
<td>0.751</td>
</tr>
</tbody>
</table>
These results point very clearly to the fact that the choice of DEA model has a strong impact on the outcome of this study and that selecting the appropriate model is paramount. While the selection of the appropriate model is key, and previous studies have pointed this out, the literature does not provide many examples of or guidance for selecting the most appropriate model.

In comparing which inputs and outputs are assigned significance in the three different models, the following two observations can be made:

- The same weights were assigned to outputs in the BCC as in the CCR model, and these weights were all focused on a single output for each airport. However, input weights differed between the two models.
- In comparing the CCR and the SBM models, the same input variables are assigned non-zero weights, although these weights are different. On the output side, the SBM model assigns non-zero weights to a much broader set of variables than does the CCR model, which usually assigns non-zero values to only a single variable. It can be argued that this leads to the SBM model covering a more comprehensive and thereby accurate scope, rather than only considering a single output variable as in the CCR case.

Notably, running the BCC model, which considers variable returns to scale, resulted in either constant or decreasing returns to scale. While impossible to depict graphically in a multi-dimensional model, the concept of decreasing returns to scale is reflected in the shape of the convex hull spanning the DMU(s). In practical terms, this means that rather than observing any economies of scale in terms of efficiency, the model resulted in either no difference or actually a negative effect on airport efficiency as scale grew.

### C. Observations on the Weakness of Using Artificial, “Super-Efficient” DMUs

In running the CCR analysis using the artificial, “super-efficient” DMU, one serious problem that arises is the fact that for nearly every non-efficient DMU, the model selects just one single input and a single output to have a non-zero weight while all others are assigned a zero weight. That means that the resulting efficiency score is highly skewed and does not give an accurate representation of a DMU’s efficiency.

In the analysis, Bazargan and Vasigh make reference to the use of an infinitesimal value in a constraint to avoid setting any of the weights to 0, but make no reference as to whether this was used and if so, what values were used. As was observed in this analysis, using such a technique is important in order to achieve an accurate assessment of all DMUs across all input and output parameters rather than focusing the analysis on just a subset of parameters.

An alternative approach to handle the issue of setting weights to 0 is the so-called cross-efficiency score, as originally proposed by Doyle and Green [17]. This method calculates the optimal weights for each DMU in the typically manner for CCR, BCC, or SBM, but then applies each DMU’s weights to all other DMUs, and then computes an average score for each DMU. While this does aid in mitigating the effects of one DMU’s score being calculated on the basis of only a small set of inputs and outputs, it also can be argued to run counter to one of the fundamental principles of DEA; the assumption that a management tradeoff is made between performing well across the given inputs and outputs for an individual DMU, and therefore the model should permit each DMU to appear in the “best possible light” by selecting its optimal weightings.

### V. Conclusions and Future Work

The analysis in this paper has shown that depending on the DEA model chosen, radically different results may be obtained. Consequently, any study of airport efficiency needs to begin with a thorough examination of the models available and a motivation for why a particular model was selected. Without an upfront analysis of this kind, a study’s final results may be called into question.

In order to simplify such analysis, a thorough study of the implications of each of the most commonly used models is necessary. Such a study needs to examine the implications and interpretations of each model in an airport context since the characteristics of each model may take on different meanings depending on the application area.

Finally, a further, broader area worthy of additional analysis relates to the selection of inputs and outputs to airport efficiency benchmarks: Before proceeding to calculating the airports’ efficiency scores, any study needs to present a discussion of what the true goals of an airport are – for example maximizing passenger throughput, maximizing aircraft movements, minimizing delay, maximizing profits, minimizing costs, etc. Only after determination of the airports’ goals can the appropriate inputs and outputs be selected. Many of the analyses to-date have largely omitted this discussion and appear to have selected inputs and outputs merely on the availability of data.

### REFERENCES


### TABLE IX. RESULTS OF KRUSKAL-WALLIS TEST ON SBM RANKS

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean efficiency rank</th>
<th>Year</th>
<th>Mean efficiency rank</th>
<th>Year</th>
<th>Mean efficiency rank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Large airports</td>
<td>Medium airports</td>
<td>Small airports</td>
<td></td>
<td>Large airports</td>
</tr>
<tr>
<td>1996</td>
<td>9.67</td>
<td>23.60</td>
<td>35.73</td>
<td>1996</td>
<td>9.67</td>
</tr>
</tbody>
</table>

Chi-square 29.589 30.939 31.710 32.010 0.000 0.000 0.000 0.000 0.000

Asymptotic significance 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

### \( R \) \( K \) \( RUSKAL \) ALLIS TEST ON RANKS


