Integrated Framework for Robustness Analysis Using Ratio-Based Efficiency Model with Application to Evaluation of Polish Airports

Miłosz Kadziński*[,a], Anna Labijak[a], Małgorzata Napieraj[a]

[a]*Institute of Computing Science, Poznań University of Technology, Poznań, Poland*

Abstract

We consider a problem of evaluating efficiency of Decision Making Units (DMUs) based on their deterministic performance on multiple consumed inputs and multiple produced outputs. We apply a ratio-based efficiency measure, and account for the Decision Maker's preference information representable with linear constraints involving input/output weights. We analyze the set of all feasible weights to answer various robustness concerns by deriving: (1) extreme efficiency scores and (2) extreme efficiency ranks for each DMU, (3) possible and necessary efficiency preference relations for pairs of DMUs, (4) efficiency distribution, (5) efficiency rank acceptability indices, and (6) pairwise efficiency outranking indices. The proposed hybrid approach combines and extends previous results from Ratio-based Efficiency Analysis and the SMAA-D method. The practical managerial implications are derived from the complementary character of accounted perspectives on DMUs' efficiencies. We present an innovative open-source software implementing an integrated framework for robustness analysis using a ratio-based efficiency model on the *diviz* platform. The proposed approach is applied to a real-world problem of evaluating efficiency of Polish airports. We consider four inputs related to the capacities of a terminal, runways, and an apron, and to the airport's catchment area, and two outputs concerning passenger traffic and number of aircraft movements. We present how the results can be affected by integrating the weight constraints and eliminating outlier DMUs.

Key words: Data Envelopment Analysis, Ratio-based Efficiency, Robustness Analysis, Stochastic Multicriteria Acceptability Analysis, Airport Efficiency, Software

1. Introduction

The framework of Data Envelopment Analysis (DEA) offers a variety of methods for evaluating the relative efficiency of Decision Making Units (DMUs) which consume multiple inputs and produce

*Corresponding author: Institute of Computing Science, Poznań University of Technology, Piotrowo 2, 60-965 Poznań, Poland. Tel. +48-61 665 3022.

Email addresses: milosz.kadzinski@cs.put.poznan.pl (Miłosz Kadziński), anna.labijak@student.put.poznan.pl (Anna Labijak), napieraj.malgorzata@gmail.com (Małgorzata Napieraj)

multiple outputs [18, 38, 39]. Conceptually, efficiency is the ratio between virtual output and virtual input, i.e., respectively, outputs or inputs aggregated using some weights assigned to these factors [14]. Typically, DEA methods have been used to classify the DMUs into efficient and inefficient ones. By definition, the former ones have an efficiency score equal to one, whereas for the latter ones this measure is less than one. For the inefficient DMUs, such scores convey information on how close to being efficient they are. Analysis of these measures may lead to formulating the corrective actions, revealing an excess use of some inputs or shortfalls in the production of outputs, as well as to indicating a reference set of some comparable DMUs.

1.1. Critical View on the Traditional Methods of Data Envelopment Analysis

Although DEA has proven its usefulness when applied to a variety of real-world problems (see, e.g., [18, 23, 40]), some criticism has been leveled against its discriminative power and the way the efficiency scores are computed. Firstly, the efficiency measures for each DMU are derived from the analysis of the input/output weights which are the most favorable to it. However, a weight vector for which a DMU attains its maximal efficiency is not unique [36]. Thus, choosing among them is arbitrary to a large extent. Secondly, the underlying Linear Programming (LP) techniques require some normalization of weights for each DMU individually. This implies that scaling affects the optimal weights and a meaningful comparison of these weights across different DMUs is difficult. Thirdly, the efficiency measures fail to reflect how the efficiencies of DMUs compare to each other for other feasible weight vectors [53]. In fact, only extremely small share of feasible weights is taken into account in the analysis, while others are neglected despite being equally desirable. Fourth, DEA measures efficiency relative to the efficient frontier. This requires some assumptions about possible returns to scale (e.g., constant or variable). These may be, however, difficult to formulate or justify. Further, we may sometimes prefer a DMU judged as inefficient, which is dominated only by some convex combination of other DMUs, but not by any existing DMU [36]. Moreover, an efficiency frontier and, thus, the efficiency scores, vastly depend on the DMUs under consideration [58, 74]. The outcomes of DEA may be very sensitive even to the inclusion or removal of a single DMU. In the same spirit, the outcomes of DEA can be interpreted only when the number of DMUs is large enough in comparison with the number of inputs and outputs. Finally, while DEA is useful for indicating which DMUs are efficient, it does not discriminate between them. In some real-world situations, the share of efficient DMUs may be very large, and we may wish to identify among them a small subset of the most distinguishing ones.

Several techniques have been proposed in the literature to address these drawbacks. In particular, preference information on the relative comparisons of inputs and/or outputs may be used to reduce the space of feasible weight vectors [50, 63], and, thus, the conclusiveness of efficiency scores. Further, the cross-efficiency methods exploit the space of feasible weights to derive for each DMU an average efficiency obtained from the analysis of weights for which other DMU's efficiency is maximal [22, 59]. Moreover, the super-efficiency discriminates the efficient DMUs by indicating for each of them how much more efficient it can be relative to the remaining ones [2, 75]. Although following the right direction, these approaches do not address all aforementioned concerns comprehensively. Doing so, requires incorporation of robustness analysis into the DEA framework.

1.2. Existing Approaches for Robustness Analysis in Data Envelopment Analysis

Robustness analysis accounts for the uncertainties which can be observed in the real-world decision problems [33]. A conclusion is considered to be robust if it is true for all or for the most plausible combinations of parameter values [52, 67]. As noted in [20], this type of analysis provides information that may allow the users to avoid answering questions they find too demanding. It may also guide them in revising or enriching the provided preference information, progressively constraining the space of admissible values for the parameters of employed model. In the context of DEA, robustness concern refers to the relative efficiencies of DMUs for all feasible input and output weights or their representative sample. Advances in this regard, that we build on in this paper, have been presented in [53] and [36].

On one hand, [53] consider the whole set of weights that are compatible with the preference information concerning input/output variables. The so-called Ratio-based Efficiency Analysis (REA) does not make any assumptions in terms of the production possibilities beyond the set of DMUs that are under comparison. To materialize the relations between the DMUs' efficiencies, the method exhibits three kinds of results derived from the analysis of the whole set of feasible weights: efficiency bounds exposing the greatest and the least relative efficiencies of a DMU compared to a subset of other DMUs, dominance relation indicating for a pair of DMUs if one of them dominates the other in pairwise efficiency comparison, and ranking intervals indicating the range of efficiency ranks that are attained by a DMU. All these results are derived from comparing DMUs' efficiencies pairwise rather than measuring their distance from an efficient frontier as in the traditional DEA models. As a result, these outcomes are interpretable even if the set of DMUs is relatively small, being at the same time less sensitive to the inclusion of DMUs whose input/output values are distant from the performances of other units.

On the other hand, [36] apply simulation to provide stochastic indices which characterize the possible outcomes of a decision problem. In Stochastic Multicriteria Acceptability Analysis for Data Envelopment Analysis (SMAA-D), it is possible to handle imprecision and uncertainty regarding the input/output weights and performances of DMUs. The method computes rank acceptability indices which measure the variety of model variables that grant each DMU any rank from the best to the worst. In particular, the best (most acceptable) DMUs are those with high acceptabilities for the first rank. When compared with the basic DEA models, the stochastic measures originally provided in SMAA-D have been found useful for making the efficient DMUs more comparable [36].

1.3. Aim of the Paper

The aim of this paper is fourth-fold. Firstly, from a methodological point of view, we extend the range of outcomes considered in REA and SMAA-D. With respect to the robustness analysis, we show how to determine the least efficiency measure for each DMU, i.e., what is the lower bound of the efficiency range when the whole set of DMUs (including the DMU under consideration) is analyzed. When considering stability of the efficiency comparison for pairs of DMUs, we propose to consider the necessary and possible efficiency preference relations instead of the dominance relation. The necessary relation needs to be confirmed by all feasible weight vectors, while the possible one has to be supported by at least one feasible weight vector. We show that taking into account these results is more beneficial than analyzing the dominance relation because of their interpretability and intuitive convergence with the growth of the preference information for input/output variables.

When it comes to SMAA-D, we significantly enrich the range of stochastic indices that can be derived from the representative sample of weight vectors so that they additionally capture the efficiency scores and pairwise efficiency relations. In particular, we analyze the extreme observed efficiencies, the distribution of efficiency measures, and pairwise efficiency outranking (winning) indices indicating the probability that one DMU has an efficiency at least as good (better) than the other. In this way, we provide both exact and stochastic outcomes reflecting three different perspectives on DMUs' efficiency: scores, pairwise preference relations, and attained ranks.

Secondly, we clearly demonstrate the benefits of considering together the outcomes of thus revised REA and SMAA-D. On one hand, with the necessary, possible, and extreme outcomes of the revisited REA, we can analyze what happens for all, some, the most and the least advantageous model parameters. However, the difference between extreme ranks and efficiencies may often be very large, and in practical decision analysis the information on the sole possibility of attaining a particular rank or an efficiency in a given subinterval may be insufficient. Similarly, REA leaves incomparable the pairs of units which are possibly preferred to each other. In this perspective, SMAA-D may enrich REA with answering questions on how probable are the possible efficiency preference relations and what is the distribution of ranks or efficiencies between the best and the worst ones. These results can be further exploited to indicate the expected rank (efficiency) for a given DMU, the ranks (efficiencies) which are attained most often, and the probability of being judged as efficient (obtaining the highest efficiency).

On the other hand, even though the stochastic indices can be estimated with high accuracy using Monte Carlo simulation, they are not exact. In particular, it may be unlikely to hit the weight vector corresponding to the extreme results. This, in turn, implies that such results would not be reflected in the distribution of ranks or efficiency scores. For the same reason, an estimated pairwise efficiency outranking index equal to one or zero does not, respectively, confirm the necessity or exclude the possibility of one DMU being preferred over another. Still, all these input/output weights whose indications are not reflected in the estimations of stochastic indices are feasible. Thus, it is desirable to confront the indices derived from Monte Carlo simulation with the possible, necessary, and extreme outcomes of exact robustness analysis conducted with LP techniques.

By combining REA and SMAA-D within an integrated framework incorporating robustness and stochastic analysis, we provide a DEA-type variant of hybrid methods that have been recently proposed in Multiple Criteria Decision Aiding (MCDA) [32, 33]. In this way, we tighten the interrelations between DEA and MCDA (for a comparison of these two methodological frameworks, see, e.g., [3, 13, 17, 28, 29, 55, 61]).

The third contribution of this paper consists in presenting an open-source software implementing the methods for robustness analysis using ratio-based data envelopment model. They are made available in the form of independent software components on the *diviz* platform [42]. These modules can be subsequently combined, using an intuitive user interface, to construct complex algorithmic workflows. From a technological point of view, they are implemented as web-services, which read the input formatted with respect to a well-defined XML-based standard. The basic components we provide deliver either exact results using GLPK solver or stochastic indices using a Hit-And-Run sampling procedure [62]. Apart from analyzing in this way three types of results concerning efficiency scores, pairwise efficiency preference relations, and efficiency ranks, we enrich the range of DEA-based tools that can be used within *diviz* by providing modules which derive, e.g., cross-efficiency or super-efficiency scores. All implemented components allow incorporating linear weight constraints.

Finally, we apply the presented methodological framework to the real-world problem of evaluating efficiency of Polish airports. We take into account four inputs and two outputs. The inputs are related to the capacities of a terminal, runways, and an apron, and to the airport's catchment area. The outputs concern passenger traffic and number of aircraft movements. By illustrating the use of DEA-based robustness analysis for this particular problem, we prove the usefulness of the proposed approach for studying the performances and measuring the efficiency of airports. This type of research has aroused great interest in the recent years (see, e.g., [7, 24–26, 48, 70, 72]). Nevertheless, the introduced framework should be perceived as more general one; its use is not limited to this particular domain.

The remainder of this paper is organized as follows. Section 2 presents the new hybrid approach for DEA, combining and extending the ideas from REA and SMAA-D. We present how to compute and interpret robust outcomes and stochastic indices. We also discuss the interdependencies between these two types of results as well as the evolution of robust results with incremental specification of weight constraints. Section 3 concerns an open-source software implementing the proposed integrated framework for robustness analysis in DEA. Section 4 is devoted to the real-world case study investigating efficiency of Polish airports. In Section 5, we focus on the practical considerations. Section 6 concludes the paper.

2. Integrated Framework for Robustness Analysis Using Ratio-based Efficiency Measure

2.1. Notation and Basic Concepts

The following notation is used in the paper:

- $\mathcal{D} = \{DMU_1, \ldots, DMU_K\}$ – the set of considered DMUs; thus, K is the number of compared DMUs ($K = |\mathcal{D}|$);

- x_m – m-th input, $m \in \{1, \dots, M\}$;
- y_n – n-th output, $n \in \{1, \dots, N\}$;
- x_{mo} – an amount of m-th input consumed by $DMU_o \in \mathcal{D}$;
- y_{no} – an amount of n-th output produced by $DMU_o \in \mathcal{D}$;
- $v = \{v_1, \dots, v_m\}$ – a vector of input weights;
- $u = \{u_1, \dots, u_m\}$ – a vector of output weights;
- $S_v = \{v = (v_1, \dots, v_M)^T \neq 0 | v \geq 0, A_v v \leq 0\}$ and $S_u = \{u = (u_1, \dots, u_N)^T \neq 0 | u \geq 0, A_u u \leq 0\}$ – a space of feasible input and output weights, respectively; A_v and A_u are matrices of coefficients derived from linear constraint on weights representing the user's (Decision Maker's) preference information.

To measure the efficiency of each $DMU_o \in \mathcal{D}$, we apply the ratio of virtual output for $u \in S_u$ and virtual input for $v \in S_v$, defined as follows:

$$E_o(v, u) = \frac{\sum\limits_{n=1}^{N} u_n y_{no}}{\sum\limits_{m=1}^{M} v_m x_{mo}}. \tag{1}$$

For all feasible weights, the virtual inputs and outputs need to be strictly positive. For conditions satisfying this assumption, see [53].

Referring to the set of feasible weight vectors $(v, u) \in (S_v, S_u)$, robustness of the efficiency analysis may concern three points of view: efficiency scores, pairwise efficiency preference relations, and efficiency ranks. In this section, we discuss in detail two complementary ways for conducting such analysis. On one hand, LP techniques are employed to determine in an exact way: extreme efficiencies and ranks for each DMU as well as verifying the truth of the necessary and possible efficiency preference relations. On the other hand, Monte Carlo simulation algorithms are used to compute stochastic indices based on a representative sample of feasible weight vectors. The latter approach is based on normalizing input and output weights so that the following constraint is respected:

$$\sum_{n=1}^{N} u_n = \sum_{m=1}^{M} v_m = 1.$$

This normalization makes the space of feasible weights bounded. Then, a Hit-And-Run method is used to efficiently sample weights from the convex space of feasible weights [62, 66]. For this purpose, some probability distribution with joint density function in the feasible weight space needs to be assumed. Such distribution constitutes a form of partial preference information provided by an analyst. In general, our approach can work with any arbitrarily provided distribution. However, in most decision situations, its specification would be rather challenging. Thus, following SMAA-D [36] and MCDA-based Stochastic Ordinal Regression (SOR) [32, 33], when other weight distribution is not exogenously given, we use a uniform one. In this way, each weight vector has equal chances ($= 1/vol(W)$, where $vol(W)$ is the volume of the feasible weight space) to be considered within a sample of weights. This assumption is also in line with the spirit of robustness analysis, where each individual feasible weight vector is equally authorized to make some outcome non-necessary or possible, or shift the extreme bounds.

For each sampled input/output weight vector, we compute efficiency scores for all DMUs, and then normalize them by the maximal obtained efficiency. In this way, the final efficiency measures are in the interval between zero and one as in the traditional DEA methods. Such results are analyzed to derive estimates of the shares of feasible weight vectors for which: a DMU attains an efficiency score in some pre-defined efficiency subinterval or a specific rank, and for which some DMU is preferred to another.

When it comes to weight restrictions, as noted in [49], typical examples of such constraints are absolute weight bounds (e.g., $2 \leq v_1 \leq 5$), bounds on virtual inputs or outputs (e.g., $5v_1 + v_2 \geq 1$ or $2u_1 + 3u_2 \leq 1$), and bounds on the ratio of two weights (e.g., $0.5 \leq u_1/u_2 \leq 2 \implies 0.5u_2 \leq u_1 \leq 2u_2$). All these forms are admitted within the proposed framework.

2.2. Efficiency Scores

In this section, we discuss the measures that are useful for analysis of efficiency scores attained by the DMUs across all feasible weight vectors. When compared to REA, we additionally discuss how to determine the lower bound of the efficiency range when the whole set of DMUs (including the DMU under consideration) is analyzed. When compared to SMAA-D, we propose to consider the efficiency acceptability interval indices which capture the distribution of efficiency scores attained by each DMU.

2.2.1. Extreme Efficiency Scores

For each $DMU_o \in \mathcal{D}$, the best E_o^* and the worst $E_{o,*}$ efficiencies attained in the set of feasible weight vectors (S_v, S_u) may be computed using LP. The following program needs to be solved to determine E_o^*:

$$\begin{aligned} \max \quad & E_o^* = \sum_{n=1}^{N} u_n y_{no} \\ \text{subject to:} \quad & \sum_{m=1}^{M} v_m x_{mo} = 1, \\ & \sum_{n=1}^{N} u_n y_{nk} \leq \sum_{m=1}^{M} v_m x_{mk}, \quad k = 1, \ldots, K, \\ & (v, u) \in (S_v, S_u). \end{aligned} \tag{2}$$

The idea underlying problem (2) consists in finding the most advantageous feasible weight vector $(v, u) \in (S_v, S_u)$ for DMU_o in terms of its efficiency score. Note that E_o^* is equivalent to the efficiency originally proposed in the CCR model [14]. Thus, if $E_o^* = 1$, DMU_o is efficient; otherwise, it is inefficient. The worst efficiency $E_{o,*}$ can be determined with the following LP:

$$
\begin{aligned}
\min \quad & E_{o,*} = \sum_{n=1}^{N} u_n y_{no} \\
\text{subject to:} \quad & \sum_{m=1}^{M} v_m x_{mo} = 1, \\
& \sum_{n=1}^{N} u_n y_{nk} \geq \sum_{m=1}^{M} v_m x_{mk} - C(1 - b_k), \quad k = 1, \ldots, K, \\
& \textstyle\sum_{k=1}^{K} b_k \geq 1, \\
& b_k \in \{0, 1\}, \qquad k = 1, \ldots, K, \\
& (v, u) \in (S_v, S_u).
\end{aligned}
\tag{3}
$$

In the above problem, we adapt a more general technique for dealing with inconsistency in LP which is called "The Big-M (or Big-C) method" or "Exact Big-M MIP Formulation" [15, 43]. To prevent undesired compensations, this technique assumes that the value assigned to constant C is great enough. In our context, it is sufficient if:

$$C > max_{DMU_o, DMU_k \in \mathcal{D}} \{max_{m=1,\ldots,M}\{x_{mk}/x_{mo}\} - min_{n=1,\ldots,N}\{y_{nk}/y_{no}\}\}.$$

For all values of C satisfying this condition, we are guaranteed to obtain the same results.

To find the least advantageous feasible weight vector $(v, u) \in (S_v, S_u)$ for DMU_o in terms of its efficiency score, we need to minimize its efficiency while ensuring that some $DMU_k \in \mathcal{D}$ is efficient. To guarantee that an efficiency score of some DMU is not less than one, we use binary variables b_k, $k = 1, \ldots, K$. If $b_k = 1$, then $C(1-1) = 0$ and $\sum_{n=1}^{N} u_n y_{nk} \geq \sum_{m=1}^{M} v_m x_{mk}$; thus, $E_k(v, u) \geq 1$. Since we require that $\sum_{k=1}^{K} b_k \geq 1$, this condition needs to be satisfied for at least one DMU_k, $k = 1, \ldots, K$. Otherwise, if $b_k = 0$, the use of C prevents constraint violation. The minimization of $E_{o,*}$ in the objective function implies that a solver will assign ones to the binary variables so that to implement the least advantageous scenario for DMU_o.

2.2.2. Efficiency Distribution

For each $DMU_o \in \mathcal{D}$, an efficiency acceptability interval index $EAII(DMU_o, b_i)$ is the share of feasible weight vectors $(v, u) \in (S_v, S_u)$ for which DMU_o attains an efficiency score in the interval $b_i \subset [0, 1]$ ($i = 1, \ldots, B$, where B is the number of subintervals (buckets)). Let us denote with $b_{i,*}$ and b_i^* the extreme values of the subinterval b_i. Thus, $b_i = (b_{i,*}, b_i^*]$ with the proviso that b_1 is also left-closed (i.e., $b_1 = [b_{1,*} = 0, b_1^*]$. The buckets are constructed in the following way:

$$\bigcup_{i=1}^{B} b_i = [0,1],\ b_i \cap b_j = \emptyset,\ i \neq j,\ \text{and } b_i^* - b_{i,*} = b_{i+1}^* - b_{i+1,*},\ \text{for } i = 1,\ldots,B-1.$$

While this is a default setting, in general, it is possible to construct buckets with different amplitudes so that $b_i^* - b_{i,*} \neq b_{i+1}^* - b_{i+1,*}$, for $i \in \{1,\ldots,B-1\}$.

In the following we consider estimations $EAIIs'$ of efficiency acceptability interval indices derived with Monte Carlo simulation. The same remark applies to pairwise efficiency outranking indices $PEOIs$ and efficiency rank acceptability indices $ERAIs$ defined in Sections 2.3.2 and 2.4.2, respectively.

Proposition 2.1. For each $DMU_o \in \mathcal{D}$, $\sum_{i=1}^{B} EAII'(DMU_o, b_i) = 1$.

To enrich the view on the efficiency scores obtained in the representative sample $(S_v, S_u)^S$ of weight vectors (S_v, S_u), we provide the following measures:

- the extreme efficiencies $E_o^{*'}$ and $E_{o,*}^{'}$ observed in $(S_v, S_u)^S \subset (S_v, S_u)$ for each $DMU_o \in \mathcal{D}$;
- an estimate of the expected efficiency $EE_o^{'} = \sum_{(v,u)\in(S_v,S_u)^S} E_o(v,u)/W$, where W is the number of weight vectors in $(S_v, S_u)^S$.

2.3. Pairwise Efficiency Preference Relations

In this section, we present the outcomes which materialize the outcomes of robustness analysis while referring to pairwise comparisons of DMUs. When compared to REA, we propose to consider a pair of efficiency preference relations instead of a single dominance relation. When compared to SMAA-D, we additionally analyze the pairwise efficiency outranking indices which indicate the probability that one DMU attains an efficiency at least as good as the other.

2.3.1. Possible and Necessary Efficiency Preference Relations

Applying all feasible weight vectors $(v, u) \in (S_v, S_u)$, we define two efficiency preference relations in the set of DMUs $\mathcal{D}$:

- Possible efficiency preference relation, $\succsim_E^P$, which is verified for a pair of DMUs $(DMU_o, DMU_k) \in \mathcal{D} \times \mathcal{D}$, in case $E_o(v,u) \geq E_k(v,u)$ holds for at least one $(v,u) \in (S_v, S_u)$;
- Necessary efficiency preference relation, $\succsim_E^N$, which is verified for a pair of DMUs $(DMU_o, DMU_k) \in \mathcal{D} \times \mathcal{D}$, in case $E_o(v,u) \geq E_k(v,u)$ holds for all $(v,u) \in (S_v, S_u)$.

The following LP needs to be considered to assess whether these relations hold:

$$\begin{aligned} \min/\max \quad & E_o = \sum_{n=1}^{N} u_n y_{no} \\ \text{subject to:} \quad & \sum_{m=1}^{M} v_m x_{mo} = 1, \\ & \sum_{n=1}^{N} u_n y_{nk} = \sum_{m=1}^{M} v_m x_{mk}, \\ & (v,u) \in (S_v, S_u). \end{aligned} \tag{4}$$

If $E_o^{max} = max\ E_o$ obtained in problem (4) is not less than one, there exists some $(v, u) \in (S_v, S_u)$ for which $E_o(v, u) \geq E_k(v, u)$, and, thus, $DMU_o \succsim_E^P DMU_k$. If $E_o^{min} = min\ E_o$ obtained in problem (4) is greater or equal to one, there is no feasible weight vector $(v, u) \in (S_v, S_u)$ for which $E_k(v, u) > E_o(v, u)$, and, thus, $DMU_o \succsim_E^N DMU_k$.

In [53], the robustness analysis for pairs of DMUs is materialized with the efficiency dominance relation $\succ_E$. It holds for (DMU_o, DMU_k) if DMU_o necessarily attains the efficiency not less than DMU_k, while attaining strictly greater efficiency for some feasible weight vector. Thus, $DMU_o \succ_E DMU_k$ iff $DMU_o \succsim_E^N DMU_k$ and $\neg(DMU_k \succsim_E^N DMU_o)$. We consider a separate consideration of $\succsim_E^N$ and $\succsim_E^P$ (rather than aggregating these two results into $\succ_E$) more beneficial for the three following reasons:

- in case $DMU_o \succ_E DMU_k$, we may indicate whether DMU_k is possibly weakly preferred to DMU_o or not (i.e., whether $E_o(v, u) > E_k(v, u)$ for all $(v, u) \in (S_v, S_u)$, or for some $(v', u') \in (S_v, S_u)$, $E_o(v', u') = E_k(v', u')$);
- in case $\neg(DMU_o \succ_E DMU_k)$ and $\neg(DMU_k \succ_E DMU_o)$, we may indicate if DMU_o and DMU_k are related by the necessary indifference or necessary incomparability; in the former case, for all $(v, u) \in (S_v, S_u)$, $E_o(v, u) = E_k(v, u)$; in the latter case, for some $(v', u') \in (S_v, S_u)$, $E_o(v', u') > E_k(v', u')$ and for some $(v'', u'') \in (S_v, S_u)$, $E_k(v'', u'') > E_o(v'', u'')$;
- the possible and necessary efficiency preference relations converge with the growth of weight constraints provided by the Decision Maker (DM) (see Appendix C), while the dominance relation does not [53].

2.3.2. Pairiwse Efficiency Outranking Indices

For a pair of DMUs, $(DMU_o, DMU_k) \in \mathcal{D} \times \mathcal{D}$, a pairwise efficiency outranking index $PEOI(DMU_o, DMU_k)$ is the share of feasible weight vectors for which DMU_o is not worse than DMU_k in terms of the efficiency score, i.e., $E_o(v, u) \geq E_k(v, u)$.

Proposition 2.2. For $DMU_o \in \mathcal{D}, PEOI(DMU_o, DMU_o) = 1$.

Proposition 2.3. For $DMU_o, DMU_k \in \mathcal{D}$, $1 \leq PEOI(DMU_o, DMU_k) + PEOI(DMU_k, DMU_o) \leq 2$.

The pairwise efficiency winning index $PEWI(DMU_o, DMU_k)$ is the share of feasible weight vectors for which $E_o(v, u)$ is strictly better than $E_k(v, u)$.

Proposition 2.4. For $DMU_o, DMU_k \in \mathcal{D}$, $PEWI(DMU_o, DMU_k) = 1 - PEOI(DMU_k, DMU_o)$.

In the following we consider estimations of the pairwise efficiency indices $PEOI'$ and $PEWI'$ which are computed with Monte Carlo simulation.

2.4. Efficiency Ranks

In this section, we discuss a set of results clearly indicating how the DMUs' efficiency ranks vary across the entire space of feasible weights. When compared to REA, to enhance understanding of the underlying logic, we discuss alternative formulations of linear programs for identifying the extreme ranks. When compared to SMAA-D, we propose to aggregate the rank acceptability indices into the estimates of expected efficiency rank for each DMU.

2.4.1. Extreme Efficiency Ranks

The rank of DMU_o relative to all DMUs in $\mathcal{D}$ is defined with the ranking function:

$$R_o(v,u) = 1 + \sum_{DMU_k \in \mathcal{D} \setminus \{DMU_o\}} h(o,k,(v,u)), \text{ where} \tag{5}$$

$$h(o,k,(v,u)) = \begin{cases} 1, & \text{if } E_k(v,u) > E_o(v,u) \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

To identify the best $R_o^* = \min_{(v,u)\in(S_v,S_u)} R_o(v,u)$ efficiency rank that $DMU_o \in \mathcal{D}$ can attain, the following Mixed-Integer Linear Programming (MILP) model needs to be considered [53]:

$$\begin{aligned} \min \quad & R_o^* = 1 + \sum_{k=1,k\neq o}^{K} b_k \\ \text{subject to:} \quad & \sum_{n=1}^{N} u_n y_{no} = \sum_{m=1}^{M} v_m x_{mo} = 1, \\ [*] \quad & \sum_{n=1}^{N} u_n y_{nk} \leq \sum_{m=1}^{M} v_m x_{mk} + C b_k \quad (k=1,\ldots,K, k\neq o), \\ & b_k \in \{0,1\} \qquad\qquad\qquad\qquad (k=1,\ldots,K, k\neq o), \\ & (v,u) \in (S_v, S_u), \end{aligned} \tag{7}$$

where C is a large positive constant. In the above problem, it is sufficient if:

$$C > max_{DMU_o,DMU_k\in\mathcal{D}}\{max_{n=1,\ldots,N}\{y_{nk}/y_{no}\} - min_{m=1,\ldots,M}\{x_{mk}/x_{mo}\}\}.$$

In problem (7), we identify the feasible weight vector $(v,u) \in (S_v,S_u)$ for which the number of DMUs with efficiency better than $E_o(v,u)$ is minimal. If $\sum_{n=1}^{N} u_n y_{nk}$ cannot be less or equal to $\sum_{m=1}^{M} v_m x_{mk}$ for some particular weight vector, a binary variable b_k corresponding to DMU_k, $k \neq o$, is instantiated with one. Then, being multiplied by a large positive constant C, $b_k = 1$ prevents violation of constraint [*] for the respective k. This scenario occurs only if $E_k(v,u) = \sum_{n=1}^{N} u_n y_{nk} / \sum_{m=1}^{M} v_m x_{mk} > 1$, (i.e., if $\sum_{n=1}^{N} u_n y_{nk} - \sum_{m=1}^{M} v_m x_{mk} > 0$) while $E_o(v,u) = 1$. Then, $E_k(v,u) > E_o(v,u)$, and each $b_k = 1$ identifies a unit ranked better than DMU_o. Otherwise, i.e., when $E_k(v,u) \leq E_o(v,u)$, b_k is instantiated with zero (then, $Cb_k = 0$). Since the objective function is minimized, the solver tries to assign as many zeros as possible to b_k, $k = 1,\ldots,K, k \neq o$, thus, minimizing the cardinality of the set of DMUs which are ranked better than DMU_o. As a result, the sum of binary variables b_k,

$k = 1, \ldots, K, k \neq o$, increased by one is equal to the best (highest) rank of DMU_o. For example, in case there are three units simultaneously ranked better than DMU_o, $R_o^* = 3 + 1 = 4$.

The worst efficiency rank of DMU_o, $R_{o,*} = \max_{(v,u) \in (S_v, S_u)} R_o(v, u)$, is obtained as the optimum of the following MILP problem:

$$\begin{aligned}
\max \quad & R_{o,*} = 1 + \sum_{k=1, k \neq o}^{K} b_k \\
\text{subject to:} \quad & \sum_{n=1}^{N} u_n y_{no} = \sum_{m=1}^{M} v_m x_{mo} = 1, \\
[*] \quad & \sum_{m=1}^{M} v_m x_{mk} \leq \sum_{n=1}^{N} u_n y_{nk} + C(1 - b_k) \quad (k = 1, \ldots, K, k \neq o), \\
& b_k \in \{0, 1\} \quad (k = 1, \ldots, K, k \neq o), \\
& (v, u) \in (S_v, S_u).
\end{aligned} \tag{8}$$

To prevent undesired compensations in the above problem, it is sufficient if:

$$C > max_{DMU_o, DMU_k \in \mathcal{D}} \{max_{m=1,\ldots,M} \{x_{mk}/x_{mo}\} - min_{n=1,\ldots,N} \{y_{nk}/y_{no}\}\}.$$

In problem (8), we identify the feasible weight vector $(v, u) \in (S_v, S_u)$ for which the number of DMUs with efficiency not worse than $E_o(v, u)$ is maximal. If $E_k(v, u) \geq E_o(v, u)$, a binary variable b_k is instantiated with one. Thus, the sum of binary variables b_k, $k = 1, \ldots, K, k \neq o$, is equal to the number of DMUs simultaneously ranked not lower than DMU_o. When increased by one, this number indicates the worst rank of DMU_o.

To enhance understanding of the underlying reasoning, in Appendix A we present alternative formulations of the above MILPs.

2.4.2. Efficiency Rank Acceptability Indices

For $DMU_o \in \mathcal{D}$ and rank $k = 1, \ldots, K$, the efficiency rank acceptability index $ERAI(DMU_o, k) \in [0, 1]$, is the share of feasible weight vectors that grant DMU_o rank k.

Proposition 2.5. For each $DMU_o \in \mathcal{D}$, $\sum_{k=1}^{K} ERAI(DMU_o, k) = 1$.

In what follows, we consider Monte Carlo estimations of the efficiency rank acceptability indices $ERAI'$. They can be used to compute an estimate of the expected rank for $DMU_o \in \mathcal{D}$:

$$ER_o^{'} = \sum_{k=1}^{K} k \cdot ERAI'(DMU_o, k).$$

In the Appendix, we provide additional relevant information concerning different types of discussed results. In Appendix B, we present the interdependencies between robust results and stochastic indices, thus, proving how they complement each other. In Appendix C, we elaborate on the evolution of results with incremental specification of weight constraints. Finally, in Appendix D, we discuss the impact of removing some DMUs from the considered set of units on the results.

3. Implementation on the Diviz Platform

3.1. Diviz

Diviz is an open-source software which allows to design, execute, and share complex workflows implementing procedures of decision analysis [42]. Even though it was originally designed for MCDA, its characteristics are general enough to account for methods of DEA. The software infrastructure consists of:

- a Java client for algorithmic workflow design and visual analysis of the outcomes,
- distant servers for executing the workflows, i.e., computing the results.

Decision analysis procedures as well as visualization or reporting tools are available in *diviz* via XMCDA web-services. They need to read inputs and write outputs formatted using the XMCDA standard. In this way, the web-services can interoperate and be combined into complex workflows.

3.2. Implemented Methods for Robustness Analysis Using Ratio-based Efficiency Measure

Methods for robustness analysis using ratio-based data envelopment model have been implemented and made available on *diviz* as a collection of individual components (modules). They can be subsequently used to construct complex algorithmic workflows. Each module requires three input files specifying, respectively, the list of DMUs, sets of inputs and outputs, and performance matrix. The linear weight constraints may be optionally provided in yet another input file. The modules implementing stochastic analysis need to be additionally provided with the number of weight vectors that should be sampled to compute the stochastic indices. The list of implemented modules is the following:

- *DEACCREfficiency* (computes E_o^* and $E_{o,*}$ for each $DMU_o \in \mathcal{D}$),
- *DEACCRPreferenceRelations* (verifies the truth of $\succsim_E^P$ and $\succsim_E^N$ for all pairs of DMUs),
- *DEACCRExtremeRanks* (computes R_o^* and $R_{o,*}$ for each $DMU_o \in \mathcal{D}$),
- *DEASMAACCREfficiencies* (computes $EAIIs'$, $E_o^{*'}$, $E_{o,*}^{'}$, and EE_o' for each for $DMU_o \in \mathcal{D}$; it requires specification of the number of efficiency subintervals (buckets) B and number of samples used in the Hit-And-Run algorithm),
- *DEASMAACCRPreferenceRelations* (computes $PEOIs'$ for all pairs of DMUs; it requires specification of the number of samples), and
- *DEASMAACCRRanks* (computes $ERAIs'$ for all DMUs and ranks; it requires specification of the number of samples).

To enrich the arsenal of methods that can be used to investigate efficiency of DMUs, we provide the following additional components: *CCRSuperEfficiency* (computes super-efficiency for each DMU [2]), *CCRCrossEfficiency* (computes cross-efficiency of each DMU either with an aggressive or benevolent approach [22, 59]), and *CCREfficiencyBounds* (computes four types of results: the minimal and maximal ratios of each DMU's efficiency and the best or the worst efficiency of any DMU [53]). Thanks to this, the practitioners can easily compare results of different methods, while teachers can present a wide spectrum of approaches to their students using the same data format and user interface. Moreover, all available DEA components in diviz are open-source, which enhances the addition of yet other methods by the researchers.

The structures of two exemplary modules, *DEACCREfficiency* and *DEASMAACCREfficiencies*, are presented in Figures 1 and 2. They exhibit the required inputs, provided outputs, possible parametrization, and computation procedures.

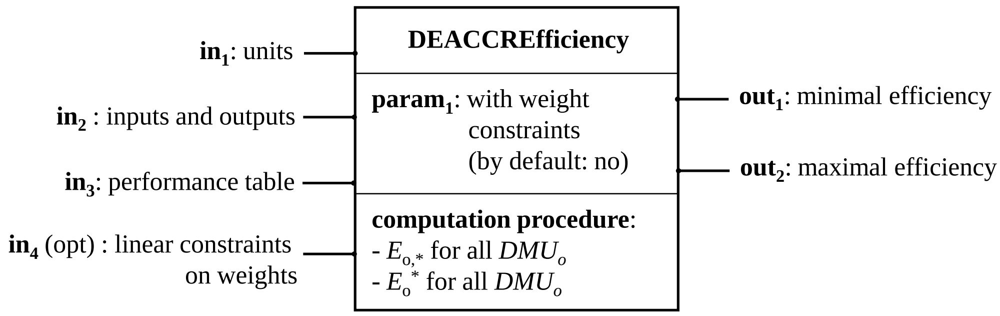

Figure 1: Structure of *diviz* module which computes the extreme efficiency scores for each DMU using LP.

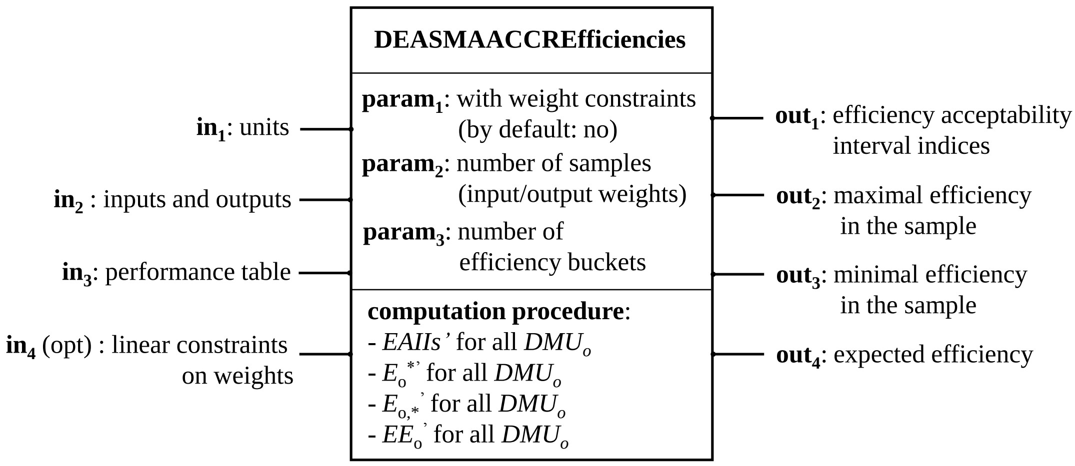

Figure 2: Structure of *diviz* module which computes the efficiency acceptability interval indices, observed extreme efficiency scores, and expected efficiency for each DMU using Monte Carlo simulation.

3.3. Workflow Design

The design of decision analysis workflows in *diviz* is performed via an intuitive graphical user interface. Each component is represented by a box which can be linked to data files or other computation modules. Thus, the design of the workflow does not require any programming skills, but rather understanding the role of each module [42]. To construct a workflow, the user chooses the modules (s)he is interested in from the list of available elements. Using a “drag-and-drop” function, (s)he adds them to the workspace along with the data files. Subsequently, the inputs and outputs of different components can be linked using connectors to define the structure of the workflow. In this way, the analysts may experiment with their own creations, suitably adjusting the arsenal of employed DEA methods to their own needs, while the researchers may design new software components that would built on the results delivered within our framework.

Once the design is finished, it is possible to execute the workflow. As already mentioned, the underlying calculations are performed on computing servers through the use of the XMCDA web-services. Thus, *diviz* requires connection to the Internet. From the point of view of practitioners, this allows to avoid performing heavy calculations on their local computers. The possibly multiple outcomes can be viewed either in *diviz* or in an external web-browser. The software maintains the history of all the past executions, which - in the context of efficiency analysis - is useful for studying the impact of additional weight constraints or removing the outlier DMUs on the results.

The *diviz* software enables to export any workflow as an archive (i.e., single file containing all necessary information including input data). This archive can be subsequently shared with other users, who can then import it (by loading the archive) into their software and execute it on the original data or continue the workflow's development. This is useful for the researchers for both dissemination and reproducibility of their results as well as for collaborative work on a particular case study.

Figure 3 presents the workflow for our case study concerning analysis of efficiency of Polish airports, whose results are discussed in Section 4. Each module delivers different results based on the input data concerning the DMUs (*DEA-unit.xml*), definition of inputs and outputs (*DEA-inOut.xml*), and underlying performances (*DEA-performanceTable.xml*). Note that, e.g., to draw the graph of necessary efficiency preference relation, the appropriate output of the *DEACCRPreferenceRelations* module is provided as the input for *plotAlternativesHasseDiagram* module.

4. Application to Efficiency Analysis of Polish Airports

4.1. Review of Airport Efficiency Applications

As noted in [70], continuous improvement of airports' competitiveness greatly affects economic development of countries. Over the last twenty years DEA has proven its usefulness for studying the performance and measuring the efficiency of airports. Such examination is important from several points of view [70]. Firstly, governments or private owners can verify that the resources available to the airport are used as effectively as possible. Secondly, airlines and passengers want to use efficient airports. Thirdly, managers can improve the competitiveness of the airports by following the best policy based on the competitors' performances.

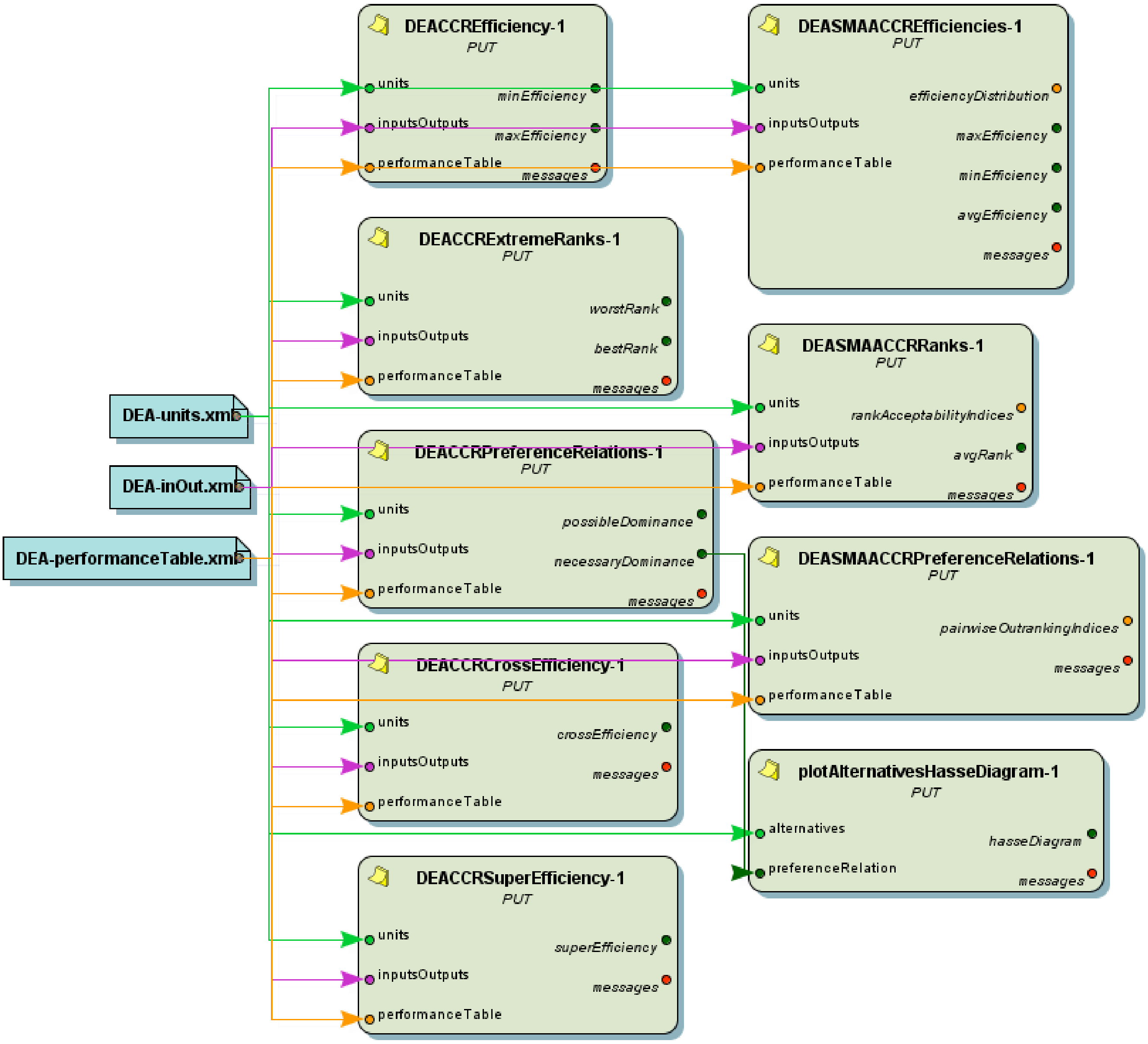

Figure 3: Algorithmic workflow for the efficiency analysis of Polish airports.

The literature concerning DEA application to measuring the efficiency and productivity of the airports can be viewed from a few perspectives:

1. Employed model:
   - CCR or BCC model for measuring airports' efficiency in a single year or season (e.g., [1, 26, 46–48, 56, 57, 71]);
   - DEA coupled with Malmquist productivity index to measure the airports' efficiency change over a few year period (e.g., [9, 25, 27, 44]);
   - DEA two-stage model, which first examines efficiency of the airports, and then uses a procedure to bootstrap DEA scores with a regression model for explanatory purpose (e.g., [7, 8]).

2. Type of considered inputs:
   - inputs related to the terminal services (e.g., number of check-in desks, gates, baggage collection belts, or parking spots, and terminal or baggage claim area) used, e.g., in [1, 8, 24–27, 47, 48, 56, 57, 71];
   - inputs related to the movement model (e.g., airport area, apron area, aircraft parking positions, numbers of runways and air routes connecting with other airports, runway length) used, e.g., in [26, 27, 48, 72];
   - monetary inputs (e.g., operational costs, labor costs, capital invested, capital stock, and airport charge) used, e.g., in [8, 44, 46, 56, 57, 71];
   - inputs related to the labor (e.g., number of employees) used, e.g., in [44, 46, 56, 57];
   - inputs related to the airport's localization (e.g., distance to the nearest city centre) used, e.g., in [1].
3. Type of considered outputs:
   - outputs related to the terminal services (e.g., number of passengers, cargo throughput, and mail tonnes) used, e.g., in [8, 25–27, 44, 46–48, 56, 71, 72]);
   - outputs related to the movement model (e.g., aircraft movement, commuter movements, and number of air carrier operations) used, e.g., in [8, 25–27, 47, 48, 71, 72];
   - monetary outputs (e.g., total revenue, operational revenue, sales to plane, sales to passengers, commercial revenue, handling revenue, and non-aeronautical fee) used, e.g., in [8, 56, 57].
4. Geographical scope:
   - single country (e.g., Argentina [8], Brazil [24], China [25], Italy [8], Japan [71], Spain [41, 44], Turkey [35], United Kingdom [9, 46], or United States [26, 56, 57]);
   - continental or intercontinental scope (e.g., Europe [1, 47, 48] or Asia-Pacific region [65, 70]).

4.2. Data Description

We analyze data concerning performances of 11 Polish airports. The geographical distribution of the airports is presented in Figure 4. Instead of the traditionally used basic inputs, such as the number of gates, aircraft parking positions, or runway length, we refer to more general and aggregated data on capacities of a terminal, runways, and an apron. These are derived from the report prepared by the world-wide leading consultancy companies [51] (see Table 1). As mentioned in [51], the values for $i_1 - i_3$ have been obtained directly from the airports. Additionally, we take into account a catchment area of each airport. The values for i_4 can be easily obtained from the Polish central statistical office. Detailed description of the four inputs is as follows:

Figure 4: Geographical distribution of Polish airports.

i_1: an annual capacity of a terminal defined as a passenger flow that an airport can accommodate without serious inconvenience (in million passengers per year); it takes into account limits on the traffic related to the terminal area, the numbers of gates and check-in counters, as well as severe congestion in access facilities;

i_2: a maximal throughput capacity defined as an average number of movements (arrivals and/or departures) that can be performed on the airport's runways (in number of movements per hour); it accounts for the configuration of runways, taxiways, waiting areas, and high speed exits, air traffic flow in the runway area (including an average runway occupancy time), and air traffic delays to the landing and takeoff moments;

i_3: a dynamic apron capacity defined as an average number of planes that can be served by the airport (in number of planes per hour); it is derived from the number and configuration of stands and ramps as well as an average stand occupancy time;

i_4: a catchment area of an airport defined as the number of inhabitants living within the range of 100 kilometers from the airport (in million inhabitants); it reflects the airport's potential for attracting the surrounding population.

When it comes to the outputs, we focus on the two primary indicators related to the terminal services and movement model, defined in the following way:

o_1: passengers traffic measured by the total number of passengers served by the airport (in million passengers per year);

o_2: number of aircraft movements (one total movement is a landing or takeoff of an aircraft) (in thousand movements per year).

The outputs are derived from the statistical data provided by the Civil Aviation Authority (CAA) in Poland [16] (see Table 1, columns $o_1 - o_2$). Let us emphasize that we have carefully selected the inputs and outputs so that they harmonize. Indeed, the inputs of each airport reflect its individually judged potential, whereas the outputs indicate the degree to which this potential is used in practice.

Table 1: Input and output performances for the problem of efficiency examination of Polish airports (all analyzed values concern 2009)

City	Short name	i_1	i_2	i_3	i_4	o_1	o_2
Warsaw	WAW	10.5	36	129.4	7.0	9.5	129.7
Cracow	KRK	3.1	19	31.6	7.9	2.9	31.3
Katowice	KAT	3.6	32	57.6	10.5	2.4	21.1
Wroclaw	WRO	1.5	12	18.0	3.0	1.5	18.8
Poznan	POZ	1.5	10	24.0	4.0	1.3	16.2
Lodz	LCJ	0.6	12	24.0	3.9	0.3	4.2
Gdansk	GDN	1.0	15	42.9	2.5	2.0	23.6
Szczecin	SZZ	0.7	10	25.7	1.9	0.3	4.2
Bydgoszcz	BZG	0.3	6	3.4	1.2	0.3	4.2
Rzeszow	RZE	0.6	6	11.3	2.7	0.3	3.5
Zielona Gora	IEG	0.1	10	63.4	3.0	0.005	0.61

4.3. Results

In this section, we discuss results derived from robustness and stochastic analysis of efficiency of Polish airports. As proven in the review presented in Section 4.1, such comprehensive analysis has never been conducted for any airport efficiency application. Moreover, while there exist some reports on measuring the efficiency of airports in many other countries, this aims to be the first comprehensive study for Poland.

First, we focus on the efficiency scores; then, we elaborate on the efficiency ranks; we conclude with the efficiency preference relations. All stochastic results presented in this section were derived from the analysis of 10000 input and output weights obtained with a Hit-And-Run algorithm. Then, we illustrate the impact of considering weight constraints and eliminating some outlier DMU.

For this purpose, we have constructed dedicated *diviz* workflows which are available online[1]:

- *DEAPolishAirports.dvz* for results presented in Sections 4.3.1, 4.3.2, and 4.3.3 without considering weight constraints;
- *DEAPolishAirportsWithConstraints.dvz* for results discussed in Section 4.3.4 when considering weight constraints;
- *DEAPolishAirportsWithoutOutlier.dvz* for results discussed in Section 4.3.5 when considering the set of airports without WAW.

[1] `http://www.cs.put.poznan.pl/mkadzinski/diviz/efficiency/` - when correcting the proofs, we will make the workflows available on the official website of diviz

These workflows can be used to reproduce the results discussed in this section. For this purpose: 1) download *diviz*[2], 2) launch it, 3) import the workflow ("Workflow - Import as new"), 4) run it on *diviz* ("Execution - Run"), and 5) view the results of interest by selecting a particular module's output. Moreover, they illustrate how to prepare the input data so that they can be later easily adapted to other problems.

4.3.1. Efficiency Distribution and Extreme Efficiencies

Table 2 (columns E_o^* and $E_{o,*}$) shows the best and the worst efficiency scores for each DMU, $DMU_o \in \mathcal{D}$. Five airports with $E_o^* = 1$ (WAW, KRK, WRO, GDN, and BZG) are deemed as efficient. Among the six inefficient airports with $E_o^* < 1$, POZ and IEG have, respectively, the least and the greatest gap that needs to be covered for reaching efficiency. Their maximal efficiency scores are equal to 0.799 and 0.258, respectively. The minimum efficiencies $E_{o,*}$ for all airports are less than 0.5. This means that for the least advantageous weight vector for each DMU, it is at least twice less efficient than another DMU. Interestingly, when taking into account the worst efficiency scores, POZ (judged inefficient) compares positively to KRK and BZG (judged efficient).

Table 2: Extreme efficiency scores (E_o^* and $E_{o,*}$), cross-efficiency (CE_o) and super-efficiency (SE_o) measures, extreme efficiencies observed in the sample ($E_o^{*'}$ and $E_{o,*}^{'}$), and estimate of the expected efficiency (EE_o') for each DMU_o

Short name	E_o^*	$E_{o,*}$	CE_o	SE_o	$E_o^{*'}$	$E_{o,*}^{'}$	EE_o'
WAW	1.000	0.452	0.773	2.277	1.000	0.560	0.944
KRK	1.000	0.213	0.689	1.123	1.000	0.257	0.664
KAT	0.591	0.108	0.362	0.591	0.519	0.131	0.281
WRO	1.000	0.338	0.731	1.039	0.991	0.387	0.702
POZ	0.799	0.218	0.551	0.799	0.732	0.258	0.533
LCJ	0.300	0.057	0.203	0.300	0.255	0.068	0.133
GDN	1.000	0.302	0.793	2.000	1.000	0.310	0.531
SZZ	0.271	0.089	0.193	0.271	0.265	0.092	0.145
BZG	1.000	0.184	0.849	1.745	1.000	0.196	0.726
RZE	0.409	0.069	0.275	0.409	0.359	0.085	0.221
IEG	0.258	0.001	0.016	0.258	0.051	0.001	0.010

The efficiency acceptability interval indices are provided in Table 3. We used 10 efficiency buckets with the same amplitude of 0.1. While for some airports the vast majority of attained efficiency scores is concentrated within a single bucket (e.g., for WAW in $(0.9, 1.0]$, or LCJ and SZZ in $(0.1, 0.2]$), for some other airports the distribution of scores is more balanced. In particular, for BZG the probability of attaining efficiency in seven different ranges between $(0.3, 0.4]$ and $(0.9, 1.0]$ is greater than 8%. Analogously, for WRO three out of ten different $EAIIs'$ are greater than 20%.

It is worthwhile analyzing the $EAIIs'$ along with the extreme efficiencies observed in the sample (see columns $E_o^{*'}$ and $E_{o,*}^{'}$ in Table 2). For most airports these differ from the true extreme efficiency scores computed with LP. In particular, for RZE, $E_{RZE}^* = 0.409 > E_{RZE}^{*'} = 0.359$ and $E_{RZE,*} = 0.069 < E_{RZE,*}^{'} = 0.085$, whereas for IEG, $E_{IEG}^* = 0.258 > E_{IEG}^{*'} = 0.051$. Such

[2] http://www.decision-deck.org/diviz/download.html

Table 3: Efficiency acceptability interval indices (in %)

	[0.0, 0.1]	(0.1, 0.2]	(0.2, 0.3]	(0.3, 0.4]	(0.4, 0.5]	(0.5, 0.6]	(0.6, 0.7]	(0.7, 0.8]	(0.8, 0.9]	(0.9, 1.0]
WAW	0.00	0.00	0.00	0.00	0.00	0.76	5.15	7.35	7.81	78.93
KRK	0.00	0.00	0.13	1.64	8.95	18.52	31.02	27.07	9.94	2.73
KAT	0.00	5.00	61.98	29.75	3.23	0.04	0.00	0.00	0.00	0.00
WRO	0.00	0.00	0.00	0.05	6.43	14.32	24.53	31.21	21.16	2.30
POZ	0.00	0.00	0.15	3.08	28.43	50.18	18.07	0.09	0.00	0.00
LCJ	4.39	95.02	0.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GDN	0.00	0.00	0.00	5.97	21.50	60.17	9.48	1.91	0.61	0.36
SZZ	0.94	97.68	1.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BZG	0.00	0.04	1.59	11.35	11.10	11.73	10.35	8.93	8.78	36.16
RZE	0.14	28.16	70.31	1.39	0.00	0.00	0.00	0.00	0.00	0.00
IEG	100.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

analysis allows to identify the ranges of scores which are attained only for marginal share of feasible weight vectors. In this perspective, the estimates of $EAIIs$ derived from Monte Carlo simulation may be equal to 0.0, while there exists some feasible input/output weight vector (not included in the sample) for which a DMU would attain efficiency contained in the underlying bucket (see, e.g., $EAII$(WAW,$(0.4, 0.5]$) or $EAII$(IEG,$(0.1, 0.2]$)).

Finally, the estimates of expected efficiency $EE_o^{'}$ (see column $EE_o^{'}$ in Table 2) may be used to rank the airports. In this case, WAW significantly outperforms other cities with $EE_{WAW}^{'} = 0.944$, and IEG is placed at the very bottom with $EE_{IEG}^{'} = 0.010$. When compared to cross-efficiencies (see column CE_o in Table 2), the advantage of using expected efficiencies consists in more in-depth exploitation of the space of feasible weights. In our case, $EE_o^{'}$ is derived from the analysis of 10000 uniformly distributed weight vectors, whereas for each $DMU_o \in \mathcal{D}$, CE_o is based only on 10 arbitrarily selected vectors for which the efficiency of some other DMU is maximal. This arbitrariness may lead to results which are surprising when compared with the indications of larger subset of feasible weights. For example, $CE_{WAW} = 0.773 < CE_{BZG} = 0.849$, while the outcomes of stochastic analysis are more favorable for WAW than BZG. In particular, $EAII'(WAW, (0.9, 1.0]) = 78.93 > EAII'(BZG, (0.9, 1.0]) = 36.16$, and $EE_{WAW}^{'} = 0.944 > EE_{BZG}^{'} = 0.726$. Additionally, when compared with super-efficiencies (see column SE_o in Table 2), EEs' enrich the conclusions that can be derived from the traditional efficiency analysis for all airports, including both the efficient and inefficient ones. Let us remind that for the inefficient DMUs, $E_o^* = SE_o$. Instead, EE_o' indicates an average performance of DMU_o, which is, in general, unique for each individual unit and different than E_o^*, $E_{o,*}$, CE_o, and SE_o. Thus, by indicating which units perform well subject to different preferences, EEs' make them more comparable.

4.3.2. Efficiency Rank Acceptability Indices and Extreme Ranks

Table 4 (columns R_o^* and $R_{o,*}$) shows the extreme ranks of each DMU, $DMU_o \in \mathcal{D}$, for all possible weight vectors. Obviously, the airports identified as efficient are potential top DMUs, i.e., the best rank for WAW, KRK, WRO, GDN, and BZG is one. BZG is more sensitive to the choice of a weight vector than other efficient DMUs, because its rank may drop to 8, while, e.g., WAW and

WRO are ranked fifth in the worst case. Among the inefficient airports, POZ is possibly ranked third, whereas no other inefficient airport is placed in top 5. KAT, LCJ, SZZ, RZE, and IEG are the least ranked DMUs though only the latter two are possibly ranked at the very bottom. The average width of the rank interval for the analyzed airports is four, and the least variation of the attained positions is observed for IEG, POZ, LCJ, and SZZ.

Table 4: Extreme ranks and efficiency rank acceptability indices (in %)

	R_o^*	$R_{o,*}$	ER'_o	1	2	3	4	5	6	7	8	9	10	11
WAW	1	5	1.3534	70.70	24.59	3.38	1.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KRK	1	6	3.5354	0.54	16.17	19.85	56.50	6.53	0.41	0.00	0.00	0.00	0.00	0.00
KAT	6	10	6.9947	0.00	0.00	0.00	0.00	0.00	0.58	99.37	0.05	0.00	0.00	0.00
WRO	1	5	2.7192	0.04	32.08	64.22	3.36	0.33	0.00	0.00	0.03	0.00	0.00	0.00
POZ	3	6	5.0994	0.00	0.00	0.13	9.89	69.89	20.09	0.00	0.00	0.00	0.00	0.00
LCJ	7	10	9.7795	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	22.05	77.95	0.00
GDN	1	6	5.0322	0.19	9.67	5.42	11.11	18.67	54.94	0.00	0.00	0.00	0.00	0.00
SZZ	7	10	9.1935	0.00	0.00	0.00	0.00	0.00	0.00	0.05	2.60	75.40	22.05	0.00
BZG	1	8	3.2662	28.56	17.49	7.00	17.81	4.58	23.98	0.58	0.00	0.00	0.00	0.00
RZE	7	11	8.0265	0.00	0.00	0.00	0.00	0.00	0.00	0.00	97.35	2.65	0.00	0.00
IEG	8	11	11.0000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.0

The rank acceptability indices are presented in Table 4 (columns $1-11$). Although WAW may be ranked in positions between 1 and 5, for over 70% of weight vectors it is ranked at the top (thus, attaining the best result among the efficient DMUs), and only for less than 5% it is ranked outside top two. Further, BZG is ranked first for over 28% weight vectors, while for the remaining efficient DMUs the probability of attaining the greatest efficiency is less than 1%. All airports but BZG attain a particular rank for the prevailing share of weight vectors. When it comes to the potentially efficient DMUs, e.g., WRO, KRK, and GDN are most often ranked third, fourth, and sixth, respectively. As for the inefficient airports, the indication of the most probable rank is even more clear. For the vast majority of weight vectors the ranks between 7 and 11 are attained by, respectively, KAT (7), RZE (8), SZZ (9), LCJ (10), and IEG (11). From another perspective, analysis of rank acceptability indices exhibits the range of ranks most often attained by the DMUs. For example, for over 99% (97%) of weight vectors, KRK (SZZ) is ranked between 2 and 5 (9 and 10), whereas, in general, its rank interval is $[1, 6]$ ($[7, 10]$). Finally, six airports have some rank acceptability indices equal to 0.0, even though analysis of the exact extreme results indicates that they may be possibly attained for at least one weight vector ((WAW, 5), (KAT, $9-10$), (LCJ, $7-8$), (BZG, 8), (RZE, $8, 10-11$), (IEG, $8-10$)). For each airport, ERAIs' can be aggregated into the estimates of an expected rank (see Table 4, column ER'_o). The airports with low ERs' (e.g., WAW, WRO, and KRK) are average good performers, while the units with high ERs' (e.g., RZE, SZZ, LCJ, and IEG) are on average far from being efficient, being ranked lower than the majority of airports.

4.3.3. Pairwise Efficiency Outranking Indices and Necessary/Possible Efficiency Preference Relations

The necessary and possible preference relations are provided in Table 5. Obviously, the truth of necessary efficiency relation implies the truth of a less demanding possible relation (for clarity, in Table 5 we list only these possible relations which are not necessary at the same time). There are 32 pairs of airports $(DMU_o, DMU_k) \in \mathcal{D} \times \mathcal{D}$, $o \neq k$, related by the necessary preference. For example, WAW and BZG are necessarily preferred to, respectively, six and four other airports.

Table 5: Necessary and possible efficiency preference relations

	Necessary preference		Additional possible preference
WAW	KAT, POZ, LCJ, SZZ, RZE, IEG	WAW	KRK, WRO, GDN, BZG
KRK	KAT, LCJ, SZZ, RZE, IEG	KRK	WAW, WRO, POZ, GDN, BZG
KAT	RZE	KAT	LCJ, SZZ, BZG, IEG
WRO	KAT, LCJ, SZZ, RZE, IEG	WRO	WAW, KRK, POZ, GDN, BZG
POZ	KAT, LCJ, SZZ, RZE, IEG	POZ	KRK, WRO, GDN, BZG
LCJ	IEG	LCJ	KAT, SZZ, RZE
GDN	KAT, LCJ, SZZ, RZE, IEG	GDN	WAW, KRK, WRO, POZ, BZG
SZZ		SZZ	KAT, LCJ, RZE, IEG
BZG	LCJ, SZZ, RZE, IEG	BZG	WAW, KRK, KAT, WRO, POZ, GDN
RZE		RZE	LCJ, SZZ, BZG, IEG
IEG		IEG	KAT, SZZ, RZE

The graph of necessary efficiency relation, subject to a transitive reduction, is illustrated in Figure 5. When analyzing this graph, the efficient airports are confirmed to be the best DMUs, because there is no other airport which is necessarily preferred to them. Note, however, than the inverse implication is not true, i.e., there may exist some inefficient DMU such that there is no other DMU necessarily preferred to it. Among the inefficient airports, POZ confirms its necessary superiority over the five remaining inefficient DMUs. Further, IEG, RZE, and SZZ should be viewed as the worst airports, because they are not necessarily preferred to any other airports.

The analysis of the diagram may be enriched with the view on the possible relations. For example, on one hand, LCJ is not possibly preferred to BZG, which means that the efficiency of BZG is always strictly greater than that of LCJ. On the other hand, although BZG is necessarily preferred to RZE, the latter is possibly preferred to the former. This means that there is at least one weight vector for which efficiencies of these two airports are equal.

Furthermore, it is interesting to analyze the graph of the necessary relation in the context of extreme ranks. Some of the observed interdependencies are straightforward. For example, while GDN (LCJ) is necessarily preferred to five (by six) other airports, its worst (best) rank is $11 - 5 = 6$ $(1 + 6 = 7)$. However, some other results are not that obvious. For example, POZ is necessarily preferred only by WAW, but its best rank is 3, whereas SZZ is not necessarily preferred to any other airport, but it is not ranked at the very bottom in the worst case.

Finally, let us note that the nodes which are not related by an arc in the diagram, indicate the airports which are incomparable in terms of $\succsim_E^N$ (e.g., (WAW,BZG), (POZ,WRO), (BZG,KAT), or

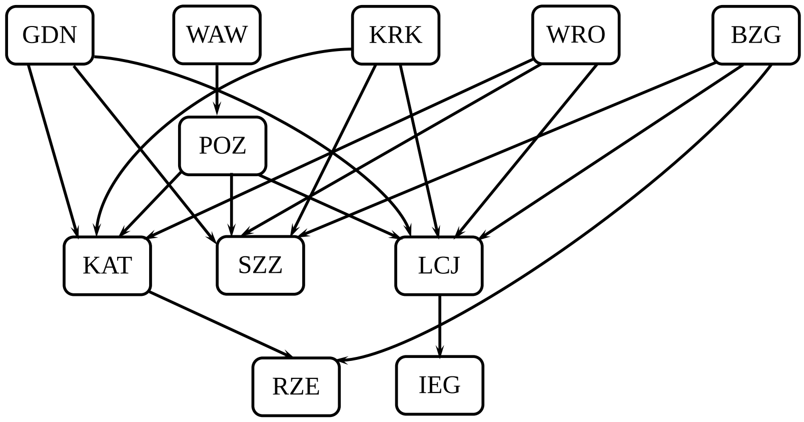

Figure 5: The necessary efficiency preference relation.

(LCJ,RZE)). For such pairs, one of the airports is possibly (for some weight vector) more efficient than the other, and vice versa. When considering the outcomes of the traditional robustness analysis, these pairs are left incomparable (no additional information is given). Instead of leaving the analyst only with information that the possible preference relations are observed for at least one weight vector, out approach provides estimates of the shares of weight vectors confirming these outcomes.

In Table 6, we present pairwise efficiency outranking indices for all DMUs. Obviously, for pairs of airports related by the necessary relation (e.g., (WAW,KAT) or (GDN,RZE)), the respective $PEOI'$ is 100%, while for pairs not related by the possible relation (e.g., (KAT,KRK) or (SZZ,WRO)), $PEOI'$ is 0%. When it comes to pairs related by the necessary incomparability, for some of them one airport is more efficient than the other for the vast majority of weight vectors. In particular, for (WAW,KRK) $PEOI'$(WAW,KRK) = 98.01% and $PEOI'$(KRK,WAW) = 1.99%, and for (WRO,POZ), $PEOI'$(WRO,POZ) = 99.58% and $PEOI'$(POZ,WRO) = 0.42%. As for the efficient airports, analysis of the pairwise efficiency outranking indices supports WAW in comparison with KRK, WRO, GDN, and BZG. For some pairs of airports, indicating the more advantageous one on the basis of $PEOIs$ is not possible. For example, for (WRO,BZG), $PEOI'$(WRO,BZG) = 53.25% and $PEOI'$(BZG,WRO) = 46.75%. Finally, although $PEOI'$ for (RZE,BZG) is equal to 0%, the possible efficiency relation for this pair holds, whereas even though $PEOI'$ for (SZZ,IEG) is equal to 100%, the necessary preference relation for this pair does not hold.

As justified in Section 2.1, when conducting Monte Carlo simulation, we assumed a uniform weight distribution for the space of feasible weights. Let us emphasize that with other exogenously given distribution the values of efficiency acceptability indices could be different. This is partially illustrated in Section 4.3.4, when a value of a density function assigned to some weight vectors is zeroed, because they are excluded from the feasible space by the provided weight constraints.

4.3.4. Incremental Specification of Weight Constraints

For illustrative purpose, in this section, we assume that the following set of linear weights constraints has been provided by the DM:

Table 6: Pairwise efficiency outranking indices (in %)

	WAW	KRK	KAT	WRO	POZ	LCJ	GDN	SZZ	BZG	RZE	IEG
WAW	100.0	98.01	100.0	95.41	100.0	100.0	99.84	100.0	71.37	100.0	100.0
KRK	1.99	100.0	100.0	18.35	99.56	100.0	83.45	100.0	43.53	100.0	100.0
KAT	0.00	0.00	100.0	0.00	0.00	100.0	0.00	99.97	0.56	100.0	100.0
WRO	4.59	81.65	100.0	100.0	99.58	100.0	89.13	100.0	53.25	100.0	100.0
POZ	0.00	0.44	100.0	0.42	100.0	100.0	62.51	100.0	26.86	100.0	100.0
LCJ	0.00	0.00	0.00	0.00	0.00	100.0	0.00	22.29	0.00	0.00	100.0
GDN	0.16	16.55	100.0	10.87	37.49	100.0	100.0	100.0	31.53	100.0	100.0
SZZ	0.00	0.00	0.03	0.00	0.00	77.71	0.00	100.0	0.00	2.82	100.0
BZG	28.63	56.47	99.44	46.75	73.14	100.0	68.47	100.0	100.0	100.0	100.0
RZE	0.00	0.00	0.00	0.00	0.00	100.00	0.00	97.18	0.00	100.0	100.0
IEG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.0

- input weights: $v_1 \geq 3v_3$, $v_1 \geq 5v_3$, $v_2 \geq 2v_3$, and $v_2 \geq 5v_4$;
- output weights: $u_1 \geq 5u_2$.

In Table 7, we provide extreme efficiency scores and ranks in two iterations, i.e., when considering the weight space without (1) and with (2) the above specified constraints. These illustrate that the ranges of efficiencies and ranks become more precise when preference information is taken into accounted. In particular, KRK, WRO, and BZG become not efficient. Their best efficiency score is less than one ($E^*_{o,2} < 1$ and $E^*_{o,1} = 1$) and they are ranked second in the best case ($R^*_{o,2} = 2$ and $R^*_{o,1} = 1$). This implies that only WAW and GDN remain efficient. Constraining the weight space is neither advantageous for IEG. Its best efficiency score drops from 0.258 to 0.188, while the best rank decreases from 8 to 11. As a result, IEG is ranked at the very bottom for all feasible weight vectors. Furthermore, with limited weight space, GDN attains the best lowest efficiency score ($E_{GDN,*,2} = 0.455 > E_{WAW,*,2} = 0.452$), while for KRK and POZ the increase of the worst efficiency is greater than 0.2. Even though their lowest scores are much better now, their least ranks remain unchanged. On the contrary, RZE (KAT) is now ranked 9 (8) for the least advantageous weight vector, while it was ranked 11 (10) without weight constraints.

Table 7: Extreme efficiency scores and ranks without (1) and with (2) weight constraints

Short name	$E^*_{o,1}$	$E_{o,*,1}$	$E^*_{o,2}$	$E_{o,*,2}$	$R^*_{o,1}$	$R_{o,*,1}$	$R^*_{o,2}$	$R_{o,*,2}$
WAW	1.000	0.452	1.000	0.452	1	5	1	5
KRK	1.000	0.213	0.962	0.439	1	6	2	6
KAT	0.591	0.108	0.554	0.210	6	10	6	8
WRO	1.000	0.338	0.922	0.445	1	5	2	5
POZ	0.799	0.218	0.779	0.433	3	6	3	6
LCJ	0.300	0.057	0.282	0.094	7	10	7	10
GDN	1.000	0.302	1.000	0.455	1	6	1	6
SZZ	0.271	0.089	0.260	0.113	7	10	9	10
BZG	1.000	0.184	0.954	0.189	1	8	2	8
RZE	0.409	0.069	0.383	0.169	7	11	7	9
IEG	0.258	0.001	0.188	0.001	8	11	11	11

In Figure 6, we depict the graph of the necessary efficiency preference relation derived from the analysis of constrained weight space. This graph is enriched when compared with the one presented in Figure 5. Precisely, there are five pairs for which the necessary relation has become true: (KAT,RZE), (KAT,IEG), (SZZ,IEG), (RZE,SZZ), and (RZE,IEG). Interestingly, even though KRK, WRO, and BZG are not efficient and ranked second in the best case, there is no other airport that would be necessarily preferred to them. This confirms the benefits of joint consideration of the three outcome perspectives: scores, ranks, and preference relations.

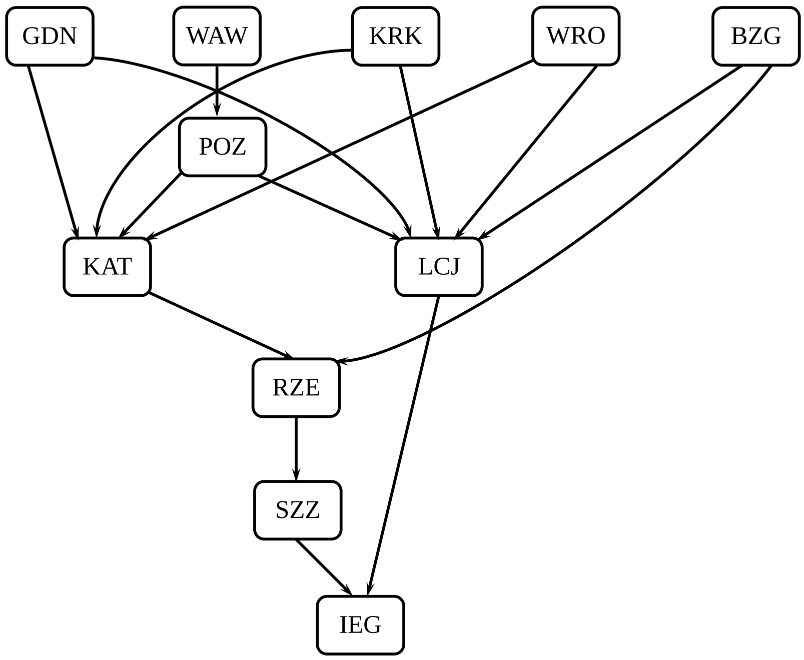

Figure 6: The necessary preference relation when accounting for the weight constraints.

To illustrate the effect of incorporating weight constraints on the acceptability indices, in Table 8 we present $EAIIs'$, $ERAIs'$, and $PEOIs'$ for BZG without and with weight constraints. The most evident effect of integrating these constraints into the efficiency model is that for the vast majority of feasible weight vector (about 97%) BZG attains efficiency scores in the range $(0.2, 0.6]$ and ranks $6-7$, while previously it attained the best efficiency scores ($(0.9, 1.0]$) and ranks on the podium $(1-3)$ for, respectively, over 35% and 50%. Moreover, BZG is now far less advantageous when compared with WAW, KRK, WRO, POZ, and GDN.

4.3.5. Elimination of Outlier DMUs

For illustrative purpose, in this section, we investigate the impact of removing some outlier airports on the obtained results. We refer to the backward approach presented in [6], and eliminate the units with super-efficiency greater than 2.0. Subsequently, we compare the original results from Sections 4.3.1-4.3.3 with the ones obtained while neglecting WAW[3]. Note that WAW is the largest

[3]In diviz, elimination of some DMU from the analysis can be conducted easily by setting a unit-specific attribute "*active*" to "*false*".

Table 8: Efficiency acceptability interval indices (in %), rank efficiency acceptability indices (in %), and pairwise efficiency outranking indices (in %) for BZG without (1) and with (2) weight constraints

	$[0.0,0.1]$	$(0.1,0.2]$	$(0.2,0.3]$	$(0.3,0.4]$	$(0.4,0.5]$	$(0.5,0.6]$	$(0.6,0.7]$	$(0.7,0.8]$	$(0.8,0.9]$	$(0.9,1.0]$
$EAII_1^{'}$	0.00	0.04	1.59	11.35	11.10	11.73	10.35	8.93	8.78	36.16
$EAII_2^{'}$	0.00	0.00	16.04	40.99	26.75	12.75	2.97	0.40	0.10	0.00
	1	2	3	4	5	6	7	8	9	$10-11$
$ERAI_1^{'}$	28.56	17.49	7.00	17.81	4.58	23.98	0.58	0.00	0.00	0.00
$ERAI_2^{'}$	0.00	0.01	0.07	0.40	1.22	84.45	13.85	0.00	0.00	0.00
	WAW	KRK	KAT	WRO	POZ	LCJ	GDN	SZZ	RZE	IEG
$PEOI_1^{'}$	28.63	56.47	99.44	46.75	73.14	100.0	68.47	100.0	100.0	100.0
$PEOI_2^{'}$	0.02	0.20	85.65	0.06	1.25	100.0	0.06	100.0	100.0	100.0

and busiest airport in Poland, which also proved to be the best in terms of robustness analysis in our results.

In Table 9, we provide the extreme efficiency scores and ranks as well as the estimates of expected efficiencies and efficiency ranks for the set of airports with ($\mathcal{D}_1$) and without ($\mathcal{D}_2$) considering WAW. For all airports, the extreme efficiency scores are not worse when WAW is neglected. Precisely, for POZ and SZZ the best efficiencies have been improved. In fact, POZ is the greatest beneficiary of removing WAW from the analysis ($E_{POZ}^{*,\mathcal{D}_2} = 0.989 > E_{POZ}^{*,\mathcal{D}_1} = 0.799$). Furthermore, the worst efficiencies have been improved significantly for all airports (e.g., $E_{WRO}^{*,\mathcal{D}_2} = 0.5 > E_{WRO}^{*,\mathcal{D}_1} = 0.338$) but GDN, SZZ, and IEG. When it comes to the extreme ranks, all airports have improved their worst ranks by one (e.g., $R_{BZG,*}^{\mathcal{D}_2} = 7 < R_{BZG,*}^{\mathcal{D}_1} = 8$). The same holds for the non-efficient airports in terms of their best ranks (e.g., $R_{POZ}^{*,\mathcal{D}_2} = 2 < R_{POZ}^{*,\mathcal{D}_1} = 3$). This confirms the robustness of the ranking intervals, which change at most by one when a single unit is removed or introduced.

As far as the expected efficiency scores and ranks are concerned, the estimates of these measures obtained from the Monte Carlo simulation after removing WAW indicate a clear improvement (i.e., greater expected efficiency and less expected rank) for all airports (e.g., $EE_{KRK}^{'\mathcal{D}_2} = 0.854 > EE_{KRK}^{'\mathcal{D}_1} = 0.664$ and $ER_{KRK}^{'\mathcal{D}_2} = 2.5769 < ER_{KRK}^{'\mathcal{D}_1} = 3.5354$).

In Figure 7, we depict the graph of the necessary efficiency preference relation and pairwise efficiency outranking indices derived from the analysis neglecting WAW. For all pairs of airports, the truth or falsity of $\succsim_E^N$ as well as the values of $PEOIs$ are the same as in Figure 5 and Table 6, respectively[4]. This confirms that the removal of some outlier DMUs does not influence the pairwise one-on-one results for the remaining units.

Finally, let us remind that the convergence of results with the removal/introduction of some unit cannot be predicted in case of efficiency acceptability interval indices and efficiency rank acceptability indices. To illustrate this phenomenon, in Table 10, we present the $EAIIs'$ and

[4]In general, the estimates of the pairwise efficiency outranking indices may differ slightly from one Monte Carlo simulation to another because there is no guarantee that the sets of feasible weight vectors sampled in these simulations are the same.

Table 9: Extreme efficiency scores and ranks, estimates of expected efficiencies and efficiency ranks with ($\mathcal{D}_1$) and without ($\mathcal{D}_2$) considering WAW

Short name	$E_o^{*,\mathcal{D}_1}$	$E_{o,*}^{\mathcal{D}_1}$	$E_o^{*,\mathcal{D}_2}$	$E_{o,*}^{\mathcal{D}_2}$	$EE_o^{'\mathcal{D}_1}$	$EE_o^{'\mathcal{D}_2}$	$R_o^{*,\mathcal{D}_1}$	$R_{o,*}^{\mathcal{D}_1}$	$R_o^{*,\mathcal{D}_2}$	$R_{o,*}^{\mathcal{D}_2}$	$ER_o^{'\mathcal{D}_1}$	$ER_o^{'\mathcal{D}_2}$
WAW	1.000	0.452	-	-	0.944	-	1	5	-	-	1.3534	-
KRK	1.000	0.213	1.000	0.420	0.664	0.854	1	6	1	5	3.5354	2.5769
KAT	0.591	0.108	0.591	0.220	0.281	0.362	6	10	5	9	6.9947	5.9953
WRO	1.000	0.338	1.000	0.500	0.702	0.901	1	5	1	4	2.7192	1.7638
POZ	0.799	0.218	0.989	0.370	0.533	0.699	3	6	2	5	5.0994	4.1006
LCJ	0.300	0.057	0.300	0.095	0.133	0.174	7	10	6	9	9.7795	8.7795
GDN	1.000	0.302	1.000	0.302	0.531	0.707	1	6	1	5	5.0322	4.0201
SZZ	0.271	0.089	0.274	0.089	0.145	0.192	7	10	6	9	9.1935	8.1943
BZG	1.000	0.184	1.000	0.312	0.726	0.891	1	8	1	7	3.2662	2.5440
RZE	0.409	0.069	0.409	0.137	0.221	0.286	7	11	6	10	8.0265	7.0305
IEG	0.258	0.001	0.258	0.001	0.010	0.014	8	11	7	10	11.000	10.000

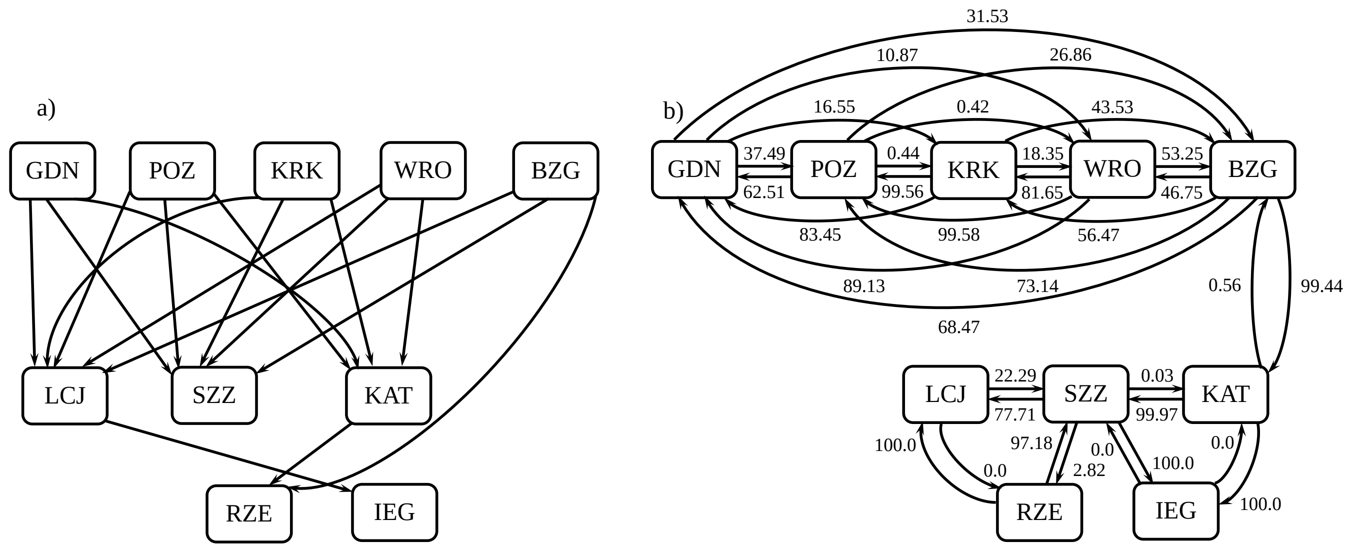

Figure 7: The necessary preference relation (a) and pairwise efficiency outranking indices (b) without considering WAW.

$ERAIs'$ for BZG obtained with and without considering WAW. The share of compatible weight vectors for which BZG attains the most advantageous efficiency scores ($(0.9, 1.0]$) and rank (1) has now increased, but for the remaining stochastic results one cannot observe any regularities except the ones already captured with the expected efficiency scores and ranks.

5. Practical Consideration

5.1. In what contexts is the proposed framework relevant?

The proposed integrated framework for robustness analysis should be used when facing at least one of the following characteristics:

1. The management wishes to investigate the performance of DMUs for all feasible input/output weights and/or their significant share. It is desirable, because the feasible weight vectors reflect relevant preference information and represent the full spectrum of priorities that can

Table 10: Efficiency acceptability interval indices (in %) and rank efficiency acceptability indices (in %) for BZG with ($\mathcal{D}_1$) and without ($\mathcal{D}_2$) considering WAW

	[0.0, 0.1]	(0.1, 0.2]	(0.2, 0.3]	(0.3, 0.4]	(0.4, 0.5]	(0.5, 0.6]	(0.6, 0.7]	(0.7, 0.8]	(0.8, 0.9]	(0.9, 1.0]
$EAII'^{\mathcal{D}_1}$	0.00	0.04	1.59	11.35	11.10	11.73	10.35	8.93	8.78	36.16
$EAII'^{\mathcal{D}_2}$	0.00	0.00	0.00	0.11	0.54	2.58	9.08	13.90	14.53	59.26

	1	2	3	4	5	6	7	8	9	10
$ERAI'^{\mathcal{D}_1}$	28.56	17.49	7.00	17.81	4.58	23.98	0.58	0.00	0.00	0.00
$ERAI'^{\mathcal{D}_2}$	46.38	7.25	17.23	4.41	24.19	0.54	0.00	0.00	0.00	0.00

be assigned to different inputs and outputs. This allows to judge the goodness of DMUs as overall performers in a more reliable way that in the traditional DEA approaches, which take into account only extremely small share of feasible weights.

2. The analyst does not want or know how to formulate, or is not able to justify the assumptions about possible returns to scale. In the proposed framework, the production possibilities are defined only by the considered DMUs, and the results are derived from pairwise comparisons among the existing units rather than measuring efficiency relative to the efficient frontier. This makes the results more reliable and less sensitive to changing the set of DMUs.
3. The number of compared DMUs is relatively small. Indeed, our framework can be used with any number of DMUs, even if it is not large enough compared to the overall number of inputs and outputs as required for the interpretability of traditional DEA outcomes (e.g., maximal efficiency scores, cross-efficiencies, or super-efficiencies). In our case study, the number of analyzed airports (11) is not significantly greater than the number of inputs and outputs (6). Nonetheless, the proposed approach can still provide interpretable and valuable results derived from the one-on-one comparisons of airports.
4. The management is interested in an in-depth analysis that would concern at least one of the following perspectives on DMUs' efficiency: scores, ranks, or preference relation. Firstly, the scores determine how much worse is a given DMU than the most efficient unit. Secondly, the ranks indicate how many DMUs are better/worse than a given DMU in terms of their efficiency ratio. Thirdly, the preference relation offers a unique one-on-one perspective for the efficiency analysis instead of one-against-all viewpoint being more typical for DEA.
These three prespectives are complementary, and the results offered by one of them, in general, cannot be derived from the analysis of another. Let us provide some examples supporting this claim:

    - if there is no unit that would be necessarily preferred to a given DMU, it may still be not efficient, thus, attaining an efficiency score less than one in the best case (see, e.g., the case of POZ in Section 4.3.5 which is not necessarily preferred by any other airport, but attains $E^{*,\mathcal{D}_2}_{POZ} = 0.989 < 1$);
    - if the intersection of the ranges of possible efficiency ranks for a pair of DMUs is non-empty, one of them may be still necessarily preferred to another (see, e.g., the case of KAT

and RZE in Section 4.3.5 with KAT $\succsim_E^{N,\mathcal{D}_2}$ RZE and $[R^{*,\mathcal{D}_2}_{KAT}, R^{\mathcal{D}_2}_{*,KAT}] \cap [R^{*,\mathcal{D}_2}_{RZE}, R^{\mathcal{D}_2}_{*,RZE}] = [6,9]$); furthermore, the pairwise outranking index may indicate that the vast majority of feasible weights ranks higher a DMU with less advantageous ranking interval (see, e.g., the case of SZZ and RZE in Sections 4.3.1-4.3.3, where $R^*_{SZZ} \leq R^*_{RZE}$ and $R_{*,SZZ} \leq R_{*,RZE}$, but $PEOI'$(RZE,SZZ) = 97.18%);

- if the least efficiency indicates a significantly worse performance of a given DMU when compared to the most efficient unit, it can be still better that the vast majority of other DMUs as proven by its worst possible rank (see, e.g., the case of GDN in Section 4.3.5 with $E^{\mathcal{D}_2}_{GDN,*} = 0.302$, thus, being over 3 times less efficient than the most efficient airport, while still being ranked better than 5 out of 9 other airports ($R^{\mathcal{D}_2}_{GDN,*} = 5$)).

Nonetheless, in practical efficiency analysis, one can use only a small subset of results that may be delivered within the proposed framework.

5.2. How to interpret different robust results and which managerial concerns they address?

Traditionally, DEA has been used for indicating which DMUs are efficient and inefficient, thus discriminating only between these two groups. In some real-world situations, the shares of efficient or inefficient DMUs may be very large, and the management may wish to identify a small subset of the most distinguishing ones among them. In their work, Tsou and Huang [64] discuss several ranking methods that have been proposed to improve the discrimination power of DEA. Our framework derives from the fact that each feasible weight vector provides a basis for the performance comparison, thus, offering greater discrimination among DMUs. The robust results which synthesize the outcomes obtained for different weight vectors can be used for answering the following relevant questions (we will provide the exemplary answers to these questions while referring to the results of our case study presented in Sections 4.3.1-4.3.3):

1. *Which efficient DMUs perform well compared to other DMUs?* For the efficient DMUs with $E^*_o = 1$ and $R^*_o = 1$, one should consider how frequently they attain the best ranks and efficiency scores and how bad they can be at worst (i.e., $E_{*,o}$ and $R_{*,o}$). This allows to distinguish the overall good performers exhibiting more universal good practices to follow from the more niche DMUs which are efficient only under very specific conditions, while being far from efficiency for the vast majority of feasible weights. In our study, the best examples of the former group are WAW and BZG, while GDN is the most representative for the latter group. Following these conclusions, WAW and BZG are advised to be used for benchmarking. Such discrimination between the efficient DMUs may also stimulate data mining to generate hypotheses about the drivers of strong or weak efficiency.
2. *Which inefficient DMUs do not perform significantly worse compared to other DMUs?* For the inefficient DMUs with $E^*_o < 1$ and $R^*_o < 1$, one should analyze how good they can be at best (i.e., E^*_o and R^*_o) as well as how often they attain their worst ranks and efficiency scores. This allows to discriminate between the inefficient DMUs which have the greatest potential

for becoming efficient and these for which attaining efficiency would be most challenging. The management may decide to implement the corrective actions for the former group in the first order, while the latter group seems to be crucial in terms of reducing the performance gap between the best and worst performers. In our study, the most advantageous inefficient airport is POZ, while KAT, RZE, SZZ, LCJ, and IEG require considerable improvement in their efficiency ratios, which is confirmed by a large spectrum of priorities that can be assigned to the inputs and outputs.

3. *Which DMUs are the average good or bad performers?* Irrespective of their efficiency status, all DMUs can be ranked from the best to the worst based on their average efficiency scores (EE_o) or ranks (ER_o). These results compare the DMUs using a large number of weight vectors, clearly exhibiting which units perform good for different priorities that can be assigned to inputs and outputs. In some situations, the expected efficiency scores or ranks can prove that inefficient DMUs are on average better than some efficient ones, thus, indicating the need for possible corrective actions also in the context of the efficient units. In our study, this is the case of GDN, since the comparison of expected efficiencies of POZ (deemed inefficient) and GDN (judged efficient) indicates that $EE'_{POZ} > EE'_{GDN}$. Although in general we built on the rankings and scores that DMUs can attain for the entire set of feasible weight vectors, these two measures can be used alike the existing DEA ranking methods for assigning a single efficiency score or position to each DMU.
4. *For which DMUs the relative efficiency scores and ranks vary much in the set of feasible weights?* For this purpose, one should analyze the difference between the extreme efficiency scores and ranks as well as the distribution of these measures across all feasible weights. High dispersion of scores and ranks indicates that the priorities of DMUs differ significantly. In some decision contexts, this should prompt investigation as to whether the guidelines for standard practice can be used as a tool to reduce variance in management. In our study, the best example of an airport for which such investigation should be conducted is BZG, which apart from being efficient in the best case, attains efficiency scores lower than 0.5 and ranks in the bottom half for about 25% of feasible weight vectors.
5. *How DMUs perform in one-on-one comparisons?* Traditionally, DEA referred to the efficiency scores and/or ranks. Although these two perspectives are deepened in the proposed framework, the necessary/possible efficiency preference relations and pairwise efficiency outranking indices offer a yet different one-on-one perspective, which is not influenced by the remaining DMUs. Indeed, the analyst may be sometimes more interested in the peer comparison. This is particularly useful if (s)he knows some units better. Then, they can be used as fixed benchmarks for the remaining DMUs. In our study, an expert interested, e.g., in the performance of SZZ and knowing POZ quite well, would get to know that POZ - despite being inefficient overall - is more efficient than SZZ for all possible priorities assigned to the inputs and outputs.

   The necessary preference relation may be very useful also in terms of formulating the corrective actions for the inefficient units. For such units, the efficient ones being necessarily preferred to them represent their hypothetical comparison units (HCUs). Differences in inputs and outputs between DMU and thus identified HCUs clearly indicate the productivity gaps and improvement potential. Moreover, when analyzing the graph of necessary efficiency preference relation, one can think of applying the step-wise benchmarking based on the specification of short-, medium-, and long-term targets. This requires identification of the paths that originate in the node representing some inefficient DMU and finish in one of the nodes corresponding to the efficient unit having no predecessors. In case there are multiple such paths, one may compare the underlying strategies to be potentially adopted. In our study, since GDN $\succsim_E^N$ KAT $\succsim_E^N$ RZE and BZG $\succsim_E^N$ RZE, the exemplary recommendation for RZE may be either to follow the example of BZG, or to focus first on reaching the efficiency level of KAT and only then following the practice of GDN.

The robust results can be also applied in other contexts which are important from the managerial perspective:

1. *Specification of performance targets* [31, 53]. In the traditional DEA methods, one investigated only the improvement that needs to be made to become efficient (i.e., to be ranked first or to attain the greatest efficiency score for some feasible weight vector). On the contrary, when referring to the robust results, the management may formulate more detailed and diverse questions. In the context of our study, they may concern, e.g., the improvement of performances that warrants that WRO is ranked at worst third for all feasible weights (while currently $R_{WRO,*} = 5 > 3$), or that BZG is necessarily preferred to KAT (while currently *not*(BZG $\succsim_E^N$ KAT) and $PEOI'$(BZG,KAT) = 99.44%), or that the efficiency of WAW is worse at most twice than that of the most efficient unit (while currently $E_{WAW,*} = 0.452 < 0.5$). The answers to these questions can be obtained with LP [53], directly indicating to the management how the DMU's performances should be bettered to attain the desired target.
2. *Identification of outlier DMUs.* The high values for the first rank efficiency acceptability indices and/or efficiency acceptability interval indices for the best scores can be used for detection of the outlier DMUs, similarly as super-efficiencies greater than a pre-defined threshold in a backward approach discussed in [6]. An obvious example of such an outlier in our study is WAW for which $EAII'(\text{WAW}, (0.9, 1.0]) = 78.93\%$ and $ERAI'(\text{WAW}, 1) = 70.70\%$.
3. *Adding discrimination among the DMUs* by introducing the restrictions on the relative values among different outputs and inputs which represent relevant managerial constraints. This is enhanced by the desirable evolution of the robust results with an incremental specification of weight constraints as discussed in Appendix C.

6. Conclusions

We have proposed an integrated framework for robustness analysis using a data envelopment model. While referring to a ratio-based measure, we considered three different viewpoints on the efficiency of Decision Making Units in the set of feasible input/output weights. Precisely, we evaluated the units' performance in terms of attained efficiency scores, pairwise preference relations, and ranks. On one hand, we assessed the extreme (in case of scores and ranks), necessary and possible (in case preference relations) performance of units using Linear Programming techniques. On the other hand, we used Monte Carlo simulation to enrich these exact outcomes with stochastic indices. The latter provide estimates of probability of attaining some result as well as some aggregated measures (e.g., expected efficiency score or efficiency rank) derived from a large set of feasible input/output weights. Apart from the complementary characteristics of the considered results, the discussed algorithms compare positively to the traditional techniques of efficiency analysis in terms of requiring less arbitrary assumptions, being less sensitive to a set of considered units, and offering greater discriminative power.

All these benefits have been illustrated on the problem of assessing efficiency of Polish airports. We took into account four inputs (i.e., capacities of a terminal, runways, and an apron, and a catchment area) and two outputs (i.e., passenger traffic and number of aircraft movements) related to the terminal services and movement model. Nevertheless, the scope of problems in which answering similar questions may be of interest to the analyst is very broad. Indeed, our approach can be used in a variety of efficiency analysis problems concerning, e.g., agricultural farms [4], banks [5, 34, 37, 68], container ports and terminals [19, 69], courts [54], local governments [21], shipping companies [45], urban rail firms [30], or transportation networks [73].

To support the applicability of our results in other decision contexts, we implemented an open-source software distributed as a part of the *diviz* platform. Apart from providing the modules for both robustness and stochastic analysis, we accounted for the well-known procedures of data envelopment analysis such as super-efficiency or cross-efficiency.
We envisage the following future developments:

- accounting for the hierarchical structure of inputs and outputs [60];
- admitting imprecise performance values;
- extension of the range of considered efficiency preference relations derived from robustness analysis, and studying their properties in terms of transitivity, completeness, reflexivity, continuity, and non-triviality;
- adapting the proposed framework to other data envelopment models such as additive DEA [28, 29];
- application to different decision problems in transport, medicine, environmental management, and education.

Acknowledgments

The authors thank Sébastien Bigaret and Patrick Meyer from Telecom Bretagne for helping us to make the software available on the *diviz* platform. The work of Miłosz Kadziński and Anna Labijak was supported by the Polish Ministry of Science and Higher Education under the Iuventus Plus program (grant no. IP2015 029674 – 0296/IP5/2016/74).

References

[1] Adler, N., Berechman, J., 2001. Measuring airport quality from the airlines viewpoint: An application of Data Envelopment Analysis. Transport Policy 8, 171–181.

[2] Andersen, P., Petersen, N., 1993. A procedure for ranking efficient units in DEA. Management Science 39, 1261–1264.

[3] Athanassopoulos, A., Podinovski, V., 1997. Dominance and potential optimality multiple criteria models decision analysis with imprecise information. Journal of the Operational Research Society 48, 142–150.

[4] Atici, K., Podinovski, V., 2015. Using data envelopment analysis for the assessment of technical efficiency of units with different specialisations: An application to agriculture. Omega 54, 72 – 83.

[5] Avkiran, N., 2015. An illustration of dynamic network DEA in commercial banking including robustness tests. Omega 55, 141–150.

[6] Banker, R., Chang, H., 2006. The super-efficiency procedure for outlier identification, not for ranking efficient units. European Journal of Operational Research 175 (2), 1311 – 1320.

[7] Barros, C., 2008. Technical change and productivity growth in airports: A case study. Transportation Research, Part A 42, 818–832.

[8] Barros, C., Dieke, P., 2008. Measuring the economic efficiency of airports: A Simar-Wilson methodology analysis. Transportation Research Part E 44, 1039–1051.

[9] Barros, C., Weber, W., 2009. Productivity growth and biased technological change in UK airports. Transportation Research Part E 45, 642–653.

[10] Bellini, T., 2010. Detecting atypical observations in financial data: the forward search for elliptical copulas. Advances in Data Analysis and Classification 4 (4), 287–299.

[11] Bellini, T., 2012. Forward search outlier detection in data envelopment analysis. European Journal of Operational Research 216 (1), 200 – 207.

[12] Bellini, T., 2015. The forward search interactive outlier detection in cointegrated var analysis. Advances in Data Analysis and Classification, 1–23.

[13] Bouyssou, D., 1999. Using DEA as a tool for MCDM. Journal of the Operational Research Society 50, 974–978.

[14] Charnes, A., Cooper, W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2 (6), 429 – 444.

[15] Chinneck, J., 2008. Feasibility and Infeasibility in Optimization: Algorithms and Computational Methods. Springer.

[16] Civil Aviation Authority, 2009. Analysis of the Polish airline market (in Polish), Warsaw, Poland.

[17] Cook, W., Tone, K., Zhu, K., 2014. Data envelopment analysis: Prior to choosing a model. Omega 44, 1 – 4.

[18] Cooper, W., Seiford, L., Zhu, J., 2011. Handbook on Data Envelopment Analysis. International Series in Operations Research & Management Science. Springer, New York.

[19] Cullinane, K., Wang, T.-F., Song, D.-W., Ji, P., 2006. The technical efficiency of container ports: Comparing data envelopment analysis and stochastic frontier analysis. Transportation Research Part A: Policy and Practice 40 (4), 354 – 374.

[20] Dias, L., Clímaco, J., 2002. Aiding Decisions with Multiple Criteria - Essays in Honor of Bernard Roy. In: Bouyssou, D., Jacquet-Lagrèze, E., Perny, P., Słowiński, R., Vanderpooten, D., Vincke, P. (Eds.), New Advances in Multiple Criteria Decision Analysis. Kluwer, pp. 175–193.

[21] Doumpos, M., Cohen, S., 2014. Applying data envelopment analysis on accounting data to assess and optimize the efficiency of greek local governments. Omega 46, 74 – 85.

[22] Doyle, J., Green, R., 1994. Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. Journal of the Operational Research Society 45, 567–578.

[23] Emrouznejad, A., Barnett, B., Tavares, G., 2008. Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. Socio-Economic Planning Sciences 42, 151–157.

[24] Fernandes, E., Pacheco, R. R., 2002. Efficient Use of Airport Capacity. Transportation Research : Part A. Policy and Practice 36, 225–238.

[25] Fung, M., Wan, K., Hui, Y., Law, J., 2008. Productivity changes in Chinese airports 1995–2004. Transportation Research Part E 44, 521–542.

[26] Gillen, D., Lall, A., 1997. Developing measures of airport productivity and performance: An application of data envelopment analysis. Transportation Research Part E: Logistics and Transportation Review 33, 261–273.

[27] Gillen, D., Lall, A., 2001. Non-parametric measures of efficiency of US airports. International Journal of Transport Economics 28, 283–306.

[28] Gouveia, M., Dias, L., Antunes, C., 2008. Additive DEA based on MCDA with imprecise information. Journal of the Operational Research Society 59, 54–63.

[29] Gouveia, M., Dias, L., Antunes, C., Boucinha, J., Inácio, C., 2015. Benchmarking of maintenance and outage repair in an electricity distribution company using the value-based DEA method. Omega 53, 104 – 114.

[30] Graham, D., 2008. Productivity and efficiency in urban railways: Parametric and non-parametric estimates. Transportation Research Part E: Logistics and Transportation Review 44 (1), 84 – 99.

[31] Kadziński, M., Ciomek, K., Rychły, P., Słowiński, R., 2015. Post factum analysis for robust multiple criteria ranking and sorting. Journal of Global Optimization (in press), 1–32, 10.1007/s10898-015-0359-3.

[32] Kadziński, M., Tervonen, T., 2013. Robust multi-criteria ranking with additive value models and holistic pair-wise preference statements. European Journal of Operational Research 228 (1), 169–180.

[33] Kadziński, M., Tervonen, T., 2013. Stochastic ordinal regression for multiple criteria sorting problems. Decision Support Systems 55 (1), 55–66.

[34] Kao, C., Liu, S.-T., 2014. Multi-period efficiency measurement in data envelopment analysis: The case of Taiwanese commercial banks. Omega 47, 90 – 98.

[35] Kocak, H., 2011. Efficiency examination of Turkish airports with DEA approach. International Business Research 4, 204–212.

[36] Lahdelma, R., Salminen, P., 2006. Stochastic multicriteria acceptability analysis using the data envelopment model. European Journal of Operational Research 170 (1), 241 – 252.

[37] LaPlante, A., Paradi, J., 2015. Evaluation of bank branch growth potential using data envelopment analysis. Omega 52, 33 – 41.

[38] Liu, J., Lu, L., Lu, W.-M., 2016. Research fronts in data envelopment analysis. Omega 58, 33 – 45.

[39] Liu, J., Lu, L., Lu, W.-M., Lin, B., 2013. Data envelopment analysis 1978–2010: A citation-based literature survey. Omega 41 (1), 3 – 15.

[40] Liu, J., Lu, L., Lu, W.-M., Lin, B., 2013. A survey of DEA applications. Omega 41 (5), 893 – 902.

[41] Martin, J., Roman, C., 2001. An application of DEA to measure the efficiency of Spanish airports prior to privatization. Journal of Air Transport Management 7 (3), 149 – 157.

[42] Meyer, P., Bigaret, S., 2012. diviz: a software for modeling, processing and sharing algorithmic workflows in MCDA. Intelligent Decision Technologies 6, 283–296.

[43] Mousseau, V., Figueira, J., Dias, L., da Silve, C. G., J.Climaco, 2003. Resolving inconsistencies among constraints on the parameters of an MCDA model. European Journal of Operational Research 147 (1), 72–93.

[44] 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.

[45] Panayides, P., Lambertides, N., Savva, C., 2011. The relative efficiency of shipping companies. Transportation Research Part E: Logistics and Transportation Review 47 (5), 681 – 694.

[46] Parker, D., 199. The performance of the BAA before and after privatisation. Journal of Transport Economics and Policy 33, 133–146.

[47] Pels, E., Nijkamp, P., Rietveld, P., 2001. Relative Efficiency of European Airports. Transport Policy 8, 183–192.

[48] Pels, E., Nijkamp, P., Rietveld, P., 2003. Inefficiencies and scale economies of European airport operations. Transportation Research Part E 39, 341–361.

[49] Podinovski, V., 2001. DEA models for the explicit maximisation of relative efficiency. European Journal of Operational Research 131 (3), 572 – 586.

[50] Podinovski, V., 2005. The explicit role of weight bounds in models of data envelopment analysis. Journal of the Operational Research Society 56, 1408–1418.

[51] PriceWaterhouseCoopers, Oliver Wyman, MKmetric GmbH, Deutsche Flugsicherung GmbH, 2010. Concept of the Central Airport in Poland (in Polish), Polish Ministry of Infrastructure, Warsaw, Poland.

[52] Roy, B., 2010. Robustness in operational research and decision aiding: A multi-faceted issue. European Journal of Operational Research 200 (3), 629–638.

[53] Salo, A., Punkka, A., 2011. Ranking intervals and dominance relations for ratio-based efficiency analysis. Management Science 57, 200–214.

[54] Santos, S., Amado, C., 2014. On the need for reform of the Portuguese judicial system Does Data Envelopment Analysis assessment support it? Omega 47, 1 – 16.

[55] Sarkis, J., 2000. A comparative analysis of DEA as a discrete alternative multiple criteria decision tool. European Journal of Operational Research 123, 543–557.

[56] Sarkis, J., 2000. Operational Efficiency of Major US Airports. Journal of Operations Management 18, 335–351.

[57] Sarkis, J., Talluri, S., 2004. Performance-based clustering for benchmarking of us airports. Transportation Research Part A 38, 329–246.

[58] Seiford, L., Zhu, J., 1998. Stability regions for maintaining efficiency in data envelopment analysis. European Journal of Operational Research 108, 127–139.

[59] Sexton, T., Silkman, R., Hogan, A., 1986. Data envelopment analysis: Critique and extensions. In: Silkman, R. H. (Ed.), Measuring the Efficiency: An Assessment of Data Envelopment Analysis. American Evaluation Association. Jossey Bass Inc., pp. 73–105.

[60] Shen, Y., Hermans, E., Brijs, T., Wets, G., 2013. Data envelopment analysis for composite indicators: A multiple layer model. Social Indicators Research 114 (2), 739–756.

[61] Stewart, T., 1996. Relationships between DEA and MCDM. Journal of the Operational Research Society 47, 654–665.

[62] Tervonen, T., van Valkenhoef, G., Basturk, N., Postmus, D., 2013. Hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis. European Journal of Operational Research 224 (3), 552 – 559.

[63] Thompson, R., Langemeier, L. N., Lee, C., Lee, E., Thrall, R., 1990. The role of multiplier bounds in efficiency analysis with application to Kansas farming. Journal of Econometrics 46, 93–108.

[64] Tsou, C.-M., Huang, D.-Y., 2010. On some methods for performance ranking and correspondence analysis in the DEA context. European Journal of Operational Research 203 (3), 771 – 783.

[65] Tsui, W., Balli, H., Gilbey, A., Gow, H., 2014. Operational efficiency of Asia-Pacific airports. Journal of Air Transport Management 40, 16 – 24.

[66] van Valkenhoef, G., Tervonen, T., Postmus, D., 2014. Notes on hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis. European Journal of Operational Research 239 (3), 865 – 867.

[67] Vincke, P., 1999. Robust solutions and methods in decision-aid. Journal of Multi-Criteria Decision Analysis 8 (3), 181–187.

[68] Wang, K., Huang, W., Wu, J., Liu, Y.-N., 2014. Efficiency measures of the Chinese commercial banking system using an additive two-stage DEA. Omega 44, 5 – 20.

[69] Wu, Y.-C. J., Goh, M., 2010. Container port efficiency in emerging and more advanced markets. Transportation Research Part E: Logistics and Transportation Review 46 (6), 1030 – 1042.

[70] Yang, H.-H., 2010. Measuring the efficiencies of Asia-Pacific international airports - parametric and non-parametric evidence. Computers & Industrial Engineering 59 (4), 697 – 702.

[71] Yoshida, Y., Fujimoto, H., 2004. Japanese-airport benchmarking with DEA and endogenous-weight TFP methods: Testing the criticism of over-investment in Japanese regional airports. Transportation Research Part E 40, 533–546.

[72] Yu, M.-M., Chern, C.-C., Hsiao, B., 2013. Human resource rightsizing using centralized data envelopment analysis: Evidence from Taiwan's airports. Omega 41 (1), 119 – 130.

[73] Zhao, Y., Triantis, K., Murray-Tuite, P., Edara, P., 2011. Performance measurement of a transportation network with a downtown space reservation system: A network-DEA approach. Transportation Research Part E: Logistics and Transportation Review 47 (6), 1140 – 1159.

[74] Zhu, J., 1996. Robustness of the efficient DMUs in data envelopment analysis. European Journal of Operational Research 90, 451–460.

[75] Zhu, J., 2001. Super-efficiency and DEA sensitivity analysis. European Journal of Operational Research 129, 443–455.

A. Alternative Fomulation of the MILPs for Computation of Extreme Efficiency Ranks

To enhance understanding of the underlying reasoning, in this section we present alternative formulations of the MILPs for computation of the extreme efficiency ranks presented in Section 2.4.1. Instead of minimizing the number of DMUs that can be simultaneously better than DMU_o, R_o^* can be obtained while subtracting from K the cardinality of the maximal subset of DMUs that are at the same time at most as good as DMU_o. For this purpose, the objective function in (7) should be replaced with:

$$\min R_o^* = K - \sum_{k=1,k\neq o}^{K} b_k,$$

while constraint [*] needs to be substituted with:

$$\sum_{n=1}^{N} u_n y_{nk} \leq \sum_{m=1}^{M} v_m x_{mk} + C(1 - b_k) \ (k = 1, \ldots, K, k \neq o).$$

Furthermore, instead of maximizing the number of DMUs that are simultaneously not worse than DMU_o, $R_{*,o}$ can be obtained while subtracting from K the cardinality of the minimal subset of DMUs that are at the same time worse than DMU_o. Then, the objective function in (8) should be replaced with:

$$\max R_{*,o} = K - \sum_{k=1,k\neq o}^{K} b_k,$$

while constraint $[_*]$ needs to be substituted with:

$$\sum_{m=1}^{M} v_m x_{mk} \leq \sum_{n=1}^{N} u_n y_{nk} + C b_k \ (k = 1, \ldots, K, k \neq o).$$

B. Interdependencies Between Robust Results and Stochastic Indices

The extreme, necessary, and possible results determined with LP influence the stochastic indices in the following way:

Remark B.1. For $DMU_o, DMU_k \in \mathcal{D}$:

1. $i : \{b_{i,*} > E_o^* \vee b_i^* < E_{o,*}\} \Rightarrow EAII'(DMU_o, b_i) = 0$ (i.e., for the efficiency subintervals outside the range delimited by the extreme efficiencies, the efficiency acceptability interval indices are 0, because a unit does not attain such efficiency scores for any feasible weight vector, including these sampled in the Monte Carlo simulation);
2. $\sum_{i:b_{i,*} \leq E_o^* \wedge b_i^* \geq E_{o,*}} EAII'(DMU_o, b_i) = 1$ (i.e., the sum of $EAIIs'$ corresponding to the efficiency intervals with non-empty intersection with $[E_{o,*}, E_o^*]$, is equal to one (see Proposition 2.1 and point 1));
3. $DMU_o \succsim_E^N DMU_k \Rightarrow PEOI'(DMU_o, DMU_k) = 1$ (if the necessary efficiency relation was valid, this needs to be confirmed by all feasible weight vectors, including these sampled in the simulation, and, thus, $PEOI'(DMU_o, DMU_k)$ is equal to one);
4. $\neg(DMU_o \succsim_E^P DMU_k) \Rightarrow PEOI'(DMU_o, DMU_k) = 0$ (i.e., if the possible efficiency relation was false, the truth of efficiency preference relation is not confirmed by any feasible weight vector, and, thus, $PEOI'(DMU_o, DMU_k) = 0$);
5. $l : \{l < R_o^* \vee l > R_{o,*}\} \Rightarrow ERAI'(DMU_o, l) = 0$ (i.e., for the ranks outside the interval delimited by the extreme ones, the efficiency rank acceptability indices are 0);
6. $\sum_{l=R_o^*}^{R_{o,*}} ERAI'(DMU_o, l) = 1$ (i.e., the sum of $ERAIs'$ for the ranks between the extreme ones, is equal to one (see Proposition 2.5 and point 5)).

Note that the inverse implications or relations are not necessarily true. In particular, the ranges of efficiencies or ranks determined exactly with LP may be wider than the respective ranges observed in the Monte Carlo sample of weight vectors. Consequently, the estimates $EAII'$ and $ERAI'$ may be equal to 0, whereas the true $EAII$ and $ERAI$ are greater than 0. Further, $PEOI'(DMU_o, DMU_k) = 1$ ($PEOI'(DMU_o, DMU_k) = 0$) does not imply that $DMU_o \succsim_E^N DMU_k$ ($\neg(DMU_o \succsim_E^P DMU_k)$) since the set of sampled weight vectors might not contain some feasible weights $(v, u) \in (S_v, S_u)$ such that $E_k(v, u) > E_o(v, u)$ ($E_o(v, u) \geq E_k(v, u)$). The only valid interdependencies between the estimates of stochastic indices and extreme, necessary, and possible outcomes are the following:

Remark B.2. For $DMU_o, DMU_k \in \mathcal{D}$:

1. $EAII'(DMU_o, b_i) > 0 \Rightarrow b_{i,*} \leq E_o^* \wedge b_i^* \geq E_{o,*}$ (i.e., when an efficiency acceptability interval index is positive, there is at least one feasible weight vectors for which a unit attains efficiency in the respective interval; this implies that the true best efficiency E_o^* of a unit is greater than the lower bound $b_{i,*}$ of the interval and its worst efficiency $E_{o,*}$ is less than the upper bound b_i^*);
2. $PEOI'(DMU_o, DMU_k) > 0 \Rightarrow DMU_o \succsim_E^P DMU_k$ (i.e. when the pairwise efficiency outranking index is greater than 0, the efficiency preference has been observed for at least one weight vector in the sample, and, thus, the possible efficiency preference relation holds);
3. $PEOI'(DMU_o, DMU_k) = 0 \Rightarrow \neg(DMU_o \succsim_E^N DMU_k)$ (i.e. when the pairwise efficiency outranking index is 0, the efficiency preference has not been observed for any weight vector in the sample, and, thus, it certainly does not hold for all feasible weight vectors; this, in turn, implies that the necessary efficiency preference relation is not valid);
4. $ERAI'(DMU_o, l) > 0 \Rightarrow R_o^* \leq l$ and $l \leq R_{o,*}$ (i.e., when an efficiency rank acceptability index for rank l is positive, there is at least one feasible weight vector for which a unit attains l-th position; this implies that the true best efficiency rank R_o^* of the unit and its worst rank $R_{o,*}$ are, respectively, not greater and not less than l).

C. Evolution of Robust Results with Incremental Specification of Weight Constraints

In this section, we consider a specification of weight constraints in the following iterations of DM's interaction with the proposed framework. We denote with $A_v^1 \subseteq A_v^2 \subseteq \ldots \subseteq A_v^s$ nested sets of weight constraints provided by the DM. These sets A_v^t, $t = 1, \ldots, s$, generate the respective sets of feasible weight vectors $(S_v, S_u)^t$. These are incrementally constrained, i.e., $(S_v, S_u)^1 \supseteq (S_v, S_u)^2 \supseteq \ldots \supseteq (S_v, S_u)^s$. For each iteration $t = 1, \ldots, s$, the following results can be derived:

- extreme efficiencies $E_o^{*,t}$ and $E_{o,*}^t$,
- possible $\succsim_E^{P,t}$ and necessary $\succsim_E^{N,t}$ efficiency preference relations,
- extreme efficiency ranks $R_o^{*,t}$ and $R_{o,*}^t$.

The evolution of the robust results with the increase of weight constraints is summarized in Proposition C.1.

Proposition C.1. For $DMU_o \in \mathcal{D}$ and $t = 1, \ldots, s-1$:

- $E_o^{*,t} \geq E_o^{*,t+1}$ and $E_{o,*}^t \leq E_{o,*}^{t+1}$ (i.e., in the following iteration, when the space of feasible weights is more constrained, the ranges of attained efficiencies may be narrowed down);
- $\succsim_E^{N,t} \subseteq \succsim_E^{N,t+1}$ and $\succsim_E^{P,t} \supseteq \succsim_E^{P,t+1}$ (i.e., the necessary and possible relations may be, respectively, enriched and impoverished);
- $R_o^{*,t} \leq R_o^{*,t+1}$ and $R_{o,*}^t \geq R_{o,*}^{t+1}$ (i.e., the ranking intervals may become narrower, but not wider).

D. Impact of Removing/Introducing Outlier DMUs on Robust Results

Traditionally, DEA methods have been focused on identifying the efficient frontier on which the DMUs are considered efficient. In this regard, much attention has been paid to identification of atypical DMUs that may greatly influence the frontier's shape [11, 74]. In general, there exist two basic approaches for detection of such outlier DMUs. On one hand, in a backward approach [6], DMUs with super-efficiencies greater than a pre-defined threshold are identified as outliers. On the other hand, in the forward search procedure [11], the subjectivity of using some arbitrary threshold can be avoided by using a dedicated distance function (see also [10, 12]).

In this subsection, we discuss the impact of removing/introducing some (outlier) DMUs on the robust results. In this perspective it is important to remind that all our results are derived from comparing DMUs' efficiencies pairwise rather than measuring their distance from an efficient frontier as in the traditional DEA models.

Let us consider the following subsets of DMUs: $\mathcal{D}' \subset \mathcal{D}'' \subseteq \mathcal{D}$. Thus, $\mathcal{D}''' = \mathcal{D}'' \setminus \mathcal{D}'$ contains the DMUs removed from $\mathcal{D}''$/introduced to $\mathcal{D}'$. We will denote the results obtained when analyzing a given subset of DMUs by using its symbol in the superscript (e.g., $\succsim_E^{N,\mathcal{D}'}$ indicates the necessary efficiency preference relation obtained for $\mathcal{D}'$). Then, the following proposition summarizes the interdependencies between the outcomes that can be obtained for $\mathcal{D}'$ and $\mathcal{D}''$.

Proposition D.1. For $DMU_o, DMU_k \in \mathcal{D}' \subset \mathcal{D}'' \subseteq \mathcal{D}$:

- $EE_o^{\mathcal{D}'} \geq EE_o^{\mathcal{D}''}$ (i.e., for each feasible weight an efficiency attained by DMU_o in $\mathcal{D}'$ is not worse than its respective efficiency in $\mathcal{D}''$; thus, after removing some DMUs, the expected efficiency EE of DMU_o cannot be deteriorated);

- $E_o^{*,\mathcal{D}'} \geq E_o^{*,\mathcal{D}''}$ and $E_{o,*}^{\mathcal{D}'} \geq E_{o,*}^{\mathcal{D}''}$ (i.e., the extreme efficiency scores of each DMU obtained within the constrained set $\mathcal{D}'$ are not less than its respective scores within $\mathcal{D}''$);

- $PEOI^{\mathcal{D}'}(DMU_o, DMU_k) = PEOI^{\mathcal{D}''}(DMU_o, DMU_k)$ (although the absolute efficiency scores attained by DMU_o and DMU_k may change after removing/introducing some other DMUs, the order between these scores remains the same for all feasible weight vectors; consequently, the value of pairwise efficiency outranking index for a given pair of DMUs does not depend on the remaining units, being the same in both $\mathcal{D}'$ and $\mathcal{D}''$);

- $DMU_o \succsim_E^{N,\mathcal{D}''} DMU_k \Leftrightarrow DMU_o \succsim_E^{N,\mathcal{D}'} DMU_k$ (the above justification proves that the status of $\succsim_E^N$ for a given pair of DMUs does not depend on other units; as a result, for all pairs of units contained in $\mathcal{D}' \times \mathcal{D}'$, the truth or falsity of $\succsim_E^N$ is the same in both $\mathcal{D}'$ and $\mathcal{D}''$);

- $0 \leq ER_o^{\mathcal{D}''} - ER_o^{\mathcal{D}'} \leq |\mathcal{D}'''|$ (i.e., for each feasible weight vector DMU_o is ranked not worse in $\mathcal{D}'$ than in $\mathcal{D}''$; in fact, it can be ranked better by at most $|\mathcal{D}'''|$, which is the cardinality of the removed subset of DMUs; thus, after removing some DMUs from $\mathcal{D}''$, the expected rank ER of DMU_o cannot be deteriorated, being at most by $|\mathcal{D}'''|$ better (lower) in $\mathcal{D}'$ than in $\mathcal{D}''$);

- $0 \leq R_o^{*,\mathcal{D}''} - R_o^{*,\mathcal{D}'} \leq |\mathcal{D}'''|$ and $0 \leq R_{o,*}^{\mathcal{D}''} - R_{o,*}^{\mathcal{D}'} \leq |\mathcal{D}'''|$ (i.e., after removing $|\mathcal{D}'''|$ units from $\mathcal{D}''$, the extreme ranks of DMU_o in $\mathcal{D}'$ cannot be deteriorated, being at most by $|\mathcal{D}'''|$ better in $\mathcal{D}'$ than in $\mathcal{D}''$).

Since the efficiency acceptability interval indices $EAIIs$ and efficiency rank acceptability indices $ERAIs$ depend on the entire set of DMUs and all feasible sets of weights, one cannot formulate any general remarks for their evolution after removing/introducing some DMUs.