Achieving MAX-MIN Fair Cross-efficiency scores in Data Envelopment Analysis

Sandro Radovanović
University of Belgrade – Faculty of Organizational Sciences

Boris Delibašić
University of Belgrade – Faculty of Organizational Sciences

Aleksandar Marković
University of Belgrade – Faculty of Organizational Sciences

Milija Suknović
University of Belgrade – Faculty of Organizational Sciences

{sandro.radovanovic, boris.delibasic, aleksandar.markovic, milija.suknovic}@fon.bg.ac.rs

Abstract

Algorithmic decision making is gaining popularity in today's business. The need for fast, accurate, and complex decisions forces decision-makers to take advantage of algorithms. However, algorithms can create unwanted bias or undesired consequences that can be averted. In this paper, we propose a MAX-MIN fair cross-efficiency data envelopment analysis (DEA) model that solves the problem of high variance cross-efficiency scores. The MAX-MIN cross-efficiency procedure is in accordance with John Rawls's Theory of justice by allowing efficiency and cross-efficiency estimation such that the greatest benefit of the least-advantaged decision making unit is achieved. The proposed mathematical model is tested on a healthcare related dataset. The results suggest that the proposed method solves several issues of cross-efficiency scores. First, it enables full rankings by having the ability to discriminate between the efficiency scores of DMUs. Second, the variance of cross-efficiency scores is reduced, and finally, fairness is introduced through optimization of the minimal efficiency scores.

1. Introduction

In many business applications, the performance of an individual or company is estimated using the concepts of effectiveness and efficiency. While effectiveness presents the degree to which an individual is successful in producing the desired result, efficiency shows how much is achieved using different inputs. More specifically, efficiency can be presented as a ratio of generated outputs and selected inputs [1].

One of the most popular techniques for estimation of efficiency is DEA. It presents a non-parametric technique for relative efficiency estimation based on linear programming (LP). More specifically, one develops a LP model as a ratio of the weighted sum of outputs and the weighted sum of inputs. The optimization task is to find output and input weights such that the efficiency is maximized. [2]

Efficiency estimation defined in DEA is deemed as a utilitarian approach. One individual obtains the best possible efficiency score for him/herself by adjusting input and output weights. Those weights might do not lead to the best efficiency scores for the other individuals. In other words, one obtains an optimistic estimate of the efficiency score. To get an efficiency score that is considered fairer, one can calculate a cross-efficiency score and use it in decision-making. The cross-efficiency score presents the average efficiency score obtained by one individual using weights of inputs and weights of outputs obtained for all individuals in the dataset. [3]

With the rise of fairness issues and the need for justice in algorithmic decision-making tools, one tends to shift from utilitarian to (luck) egalitarian approaches [4], [5]. More specifically, one wants to be aware of inherited biases in the data and implement them in algorithms prior to decision-making [6]. In other words, missing the opportunity to tackle unwanted bias will result in algorithmic decision-making tools that inherit unwanted bias and most probably amplify it [7]. The problem one faces is an unknown source of unfairness (it may occur in data at hand, the algorithm itself, or decision rules made by decision-maker) [8], [9], as well as defining an appropriate notion of fairness in algorithmic decision-making tool.

In this paper, we adopt the Rawlsian approach to fairness in DEA cross-efficiency estimation. More specifically, one seeks a DEA model that benefits the worst individual in the dataset. We propose a method that seeks such input and output weights that the minimum efficiency score of all the individuals is maximized in the cross-efficiency setting. Further, due to the nature of the optimization, the proposed model will still seek the best efficiency score for each individual. The proposed approach remains a LP

HⵏCSS

problem, and thus can be efficiently solved using methods like Simplex or interior point method.

The effectiveness of the proposed approach is shown on a dataset regarding the efficiency of the healthcare systems. Different healthcare systems can seem inefficient because inputs or outputs are of different value in the context of a specific country (f.e. European countries are spending more on healthcare per capita due to availability of funding compared to Asian countries, but Asian countries compensate this with a greater quality of care). By using DEA to rank healthcare systems could yield a flawed ranking if cross-efficiency is used. The proposed approach for the calculation of cross-efficiency scores has lower variance, thus it presents a fairer and more stable efficiency estimation and consequently fairer ranking. To the best of our knowledge, adding fairness notions in DEA is a new area with very few papers regarding these issues.

The remainder of the paper is structured as follows. In Section 2 we provide background to the DEA method, while in Section 3 we present current achievements regarding cross-efficiency in DEA method and lessons learned from the literature review. Section 4 proposes the method and describes the experiment. In Section 5, we present the results and the discussion. Finally, we conclude the paper in Section 6.

2. Background

The usage of algorithms for algorithmic decision-making had a promise of removing biases in a decision-making process [6]. Instead, it led to many injustices in real-world applications. A typical example is COMPAS tool that assesses the probability of an individual to perform a felony in the future. It has been shown that formula used is twice as likely to make a false positive error for black defendants compared to white defendants, leading to even greater racial disparities [10]. Another notable example of unfairness in algorithmic decision-making is Google, which offered male users higher-paid jobs [11]. Algorithm Google used found and replicated the bias that existed in the data thus amplifying the unfairness.

Algorithms are unaware of the cultural and historical biases and injustices, and from the optimization point of view attributes like race, gender, or religion seems like a good property of an individual for making a distinction between good and bad outcome. While using such properties for decision-making is forbidden (it is a disparate treatment), omitting them can result in unfair and unjust decisions. More specifically, information about gender, race, or religion can be visible through the so-called proxy attributes. For example, males have a higher average

salary compared to females [11], or black people live in a certain part of the city [8]. To tackle this problem, one can restrict the utility (in this paper, efficiency) score of an alternative. This will reduce the amplification of the unfairness, thus reducing the disparate impact of the algorithmic decision-making.

To assess the efficiency score of an individual or decision-making unit (DMU) in general, one can use several approaches. One can use stochastic frontier analysis (SFA) where an outcome y_i of a DMU i is estimated using the inputs X_i and coefficients β associated with the inputs. More specifically, one uses the model (1):

$$y_i = \beta_0 + \sum_{q=1}^{m} (x_{i,q} \beta_q) + v_i - u_i \qquad (1)$$

where v_i presents a stochastic component used to explain the effects of economic adversities or plain luck. These effects are assumed to occur at random and the effects for all DMUs share a common probability distribution (i.e. Normal distribution). Finally, u_i presents the score needed for a DMU to achieve an efficient score. Therefore, $u_i \geq 0$ [12]. Once one has a dataset having DMUs, their input values, and output score, SFA can be used to estimate efficiency score using gradient descent.

Although SFA is used, especially in economic efficiency modeling, it has two downsides. First, it is a parametric model which means that one needs to assume the probability distribution function of random effects. Second, SFA can be used to access efficiency scores if only one output value exists. If multiple outputs jointly describe the output of the DMU, one would need to use multiple SFA models.

To incorporate multiple outputs into efficiency estimation, one can opt for the DEA model [13]. The basic output-oriented DEA mathematical model is presented in (2) [14].

$$\max e_d = \sum_{r=1}^{s} u_r y_{rd}$$
$$s.t.$$

$$\sum_{i=1}^{m} v_i x_{id} = 1 \qquad (2)$$
$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0,$$
$$j = 1, \dots, n$$
$$u_r \geq \varepsilon, \qquad r = 1, \dots, s$$
$$v_i \geq \varepsilon, \qquad i = 1, \dots, m$$

The description of the mathematical model is as follows. One seeks the best efficiency score e_d for each DMU $d, d = 1, \dots, n$ by calculating the weighted sum of outputs y_{rd} that d-th DMU produces. More

specifically, one tries to find the weights of output r denoted u_r such that the sum of $u_r y_{rd}$ is maximized. However, one needs to restrict this linear function to get an efficiency score. First, one needs to set that weighted sum of inputs ($\sum_{i=1}^{m} v_i x_{id}$) should be exactly one. This constraint uses as a trick to convert the ratio of the weighted sum of outputs and the weighted sum of inputs into a linear form. By setting the weighted sum of inputs to a constant, the mathematical model can focus on finding the best weighted sum of outputs. The second constraint is crucial for the efficiency score. By setting that the difference between the weighted sum of outputs and the weighted sum of inputs should be lower than zero, one ensures that the efficiency score is bounded between zero and one. Finally, one employs lower bounds on inputs and outputs weights by setting a constraint that each weight should be at least ε.

The mathematical model presented in (2) is solved for each of n DMUs. Therefore, one needs to solve n LP problems to obtain efficiency scores. DEA gained its popularity also from the fact that it can be obtained by solving LP tasks. One can use slack variables and calculate referent units by observing the solution of the dual formulation of the mathematical model presented in (2). Slack variables explain what needs to be changed to obtain an efficient score and referent units explain what DMU obtained better efficiency score with the same weights like the one being observed (thus, observed DMU can look up to them to get a better efficiency score). [13]

Finally, cross-efficiency is obtained by when DMU weights are used to evaluate the efficiency of all other DMUs. Then, for each DMU one calculates the average efficiency score obtained using weights from all DMUs (including itself) as presented in (3).

$$c_d = \frac{1}{n} \sum_{q=1}^{n} \sum_{r=1}^{s} u_{rq} y_{rd} \qquad (3)$$

A cross-efficiency score is being used as a self-evaluation obtained by solving the model in (2) is criticized as an unfair comparison. It can be used to get the full ordering of DMUs (and thus further decision-making) and more importantly it eliminates unrealistic weight combinations without the elicitation of weights from the domain experts and imposing the weight restrictions [15].

3. Literature Review

Throughout the years, many works on cross-efficiency have been done. By observing the literature, one can find different models with different assumptions and use cases.

Most commonly used is the benevolent formulation of cross-efficiency proposed by [16]. That mathematical model aims at maximizing the cross-efficiencies of all other DMUs with the idea that the best cross-efficiency model is the one that achieves the highest score of cross-efficiencies. The mathematical model is presented in (4).

$$\max e_k = \sum_{r=1}^{s} u_{rk} \left(\sum_{j=1, j \neq k}^{n} y_{rj} \right)$$
$$s.t.$$

$$\sum_{i=1}^{m} v_{ik} \left(\sum_{j=1, j \neq k}^{n} x_{ij} \right) = 1$$
$$\sum_{r=1}^{s} u_{rk} y_{rk} - \theta_{kk}^{*} \sum_{i=1}^{m} v_{ik} x_{ik} \leq 0 \qquad (4)$$
$$\sum_{r=1}^{s} u_{rj} y_{rj} - \sum_{i=1}^{m} v_{ij} x_{ij} \leq 0,$$
$$\qquad\qquad j = 1, \dots, n, j \neq k$$
$$u_{rk} \geq \varepsilon, \qquad r = 1, \dots, s$$
$$v_{ik} \geq \varepsilon, \qquad i = 1, \dots, m$$

The cross-efficiency estimation starts by calculating the best relative efficiency θ_{kk}^{*} that the DMU k can achieve (i.e. solving the formula (2) for DMU k). Then, because solving (2) for each DMU can result in many alternatives having optimal solutions (in other words there exists non-uniqueness of the solution) it is proposed to optimize a secondary goal in such a manner to have the same efficiency score for the DMU being observed, but with greater efficiency scores for other DMUs in total.

The downsides of this formulation are that one needs to solve multiple optimization models in a two-step procedure. First, one needs to solve n linear programming models to obtain initial efficiency scores, and then additional n linear programming models to adjust weights such that cross-efficiency scores are increased. In addition, the changes in the weights obtained from this procedure are very small. More specifically, if a unique optimal solution exists (and most often it does) there are no changes that will result in the same value of the efficiency score.

Other approaches do exist, such as deviation from the ideal point [17]. Deviation from the ideal point is presented in terms of convex optimization, thus making it hard to implement nice properties of the basic DEA model such as slack variables and referent units. Another interesting approach is [18] where there is a guarantee that efficiency scores after adjustment are still at the Pareto frontier but moved such that other DMUs obtain better efficiency scores.

One of the issues with cross-efficiency is the need to achieve the best possible efficiency score for the DMU at hand. One can imagine a situation where a small increase in efficiency scores for one DMU reduces the efficiency score of other DMUs by a larger quantity, thus making the overall cross-efficiency smaller. In addition, needing to obtain the same efficiency score for a DMU will yield a high variance in cross-efficiency scores. Therefore, the ranking of DMUs perhaps could be flawed.

In addition, one would like to foster the disadvantaged DMUs by giving them a greater score and provide them access to resources.

4. Methodology

One notion of justice in algorithmic decision-making is lent from the political philosophy. The notion used in this paper is John Rawls's theory of justice. The guiding idea is that justice is achieved by fairness, or more specifically, that 1) each individual (DMU) has the same full an adequate scheme of basic liberties, 2) there is equality of opportunity for obtaining the desired resources, and 3) resource allocation is divided to be to the greatest benefit of the least-advantaged individual (DMU) [19]. While usage of the same inputs and outputs will provide the same decision-making scheme, solving issues of equality of opportunity and fairer resource allocation in DEA is lacking. To the best of our knowledge, there are no papers regarding these issues. This paper aims at providing a fair ranking using cross-efficiency DEA by allocating the efficiency score to the greatest benefit of the least-advantaged DMU.

4.1. MAX-MIN cross-efficiency DEA

To enforce Rawls's theory of justice in DEA cross-efficiency ranking, we introduce several differences compared to traditional cross-efficiency estimation. First, instead of a two-phased procedure (calculating the cross-efficiency matrix and adjusting the weights), our approach aims at solving a single mathematical model that will calculate the cross-efficiency matrix such that the benefit of the worst off is maximized.

First, the original DEA model from (2) should be implemented to solve n linear programming tasks in one iteration. This is achieved by the summation of individual efficiency scores. Then, one should change the goal function to maximize the minimum efficiency score. The initial mathematical model is presented in (5).

$$\max \min \sum_{r=1}^{s} u_{rd}y_{rd} + u_{rj}y_{rj}, j \neq d$$
$$s.t.$$

$$\sum_{i=1}^{m} v_{id}x_{id} = 1, \qquad d = 1, \dots, n \qquad (5)$$
$$\sum_{r=1}^{s} u_{rd}y_{rd} - \sum_{i=1}^{m} v_{id}x_{id} \leq 0,$$
$$\qquad\qquad d = 1, \dots, n$$
$$u_{rd} \geq \varepsilon, \qquad r = 1, \dots, s$$
$$v_{id} \geq \varepsilon, \qquad i = 1, \dots, m$$

The MAX-MIN formulation will seek to find the largest average efficiency score of the DMU at hand d and another DMU j. Since there are n DMUs, this mathematical model will seek to find the largest average efficiency score from $n(n-1)$ combination of average efficiency scores. The remainder of the mathematical model remains as in (2) with a difference that the proposed model has many more variables (since n mathematical models are joined into one). Solving the MAX-MIN linear programming model can be formulated as in (6) [20].

$$\max \theta$$
$$s.t.$$

$$\sum_{r=1}^{s}(u_{rd}y_{rd} + u_{rj}y_{rj}) \geq \theta,$$
$$\qquad\qquad d = 1, \dots, m; j \neq d$$
$$\sum_{i=1}^{m}(v_{id}x_{id}) = 1, \qquad d = 1, \dots, m \qquad (6)$$
$$\sum_{r=1}^{s}(u_{rd}y_{rd}) - \sum_{i=1}^{m}(v_{id}x_{id}) \leq 0,$$
$$\qquad\qquad d = 1, \dots, n$$
$$u_{rd} \geq \varepsilon, \qquad r = 1, \dots, s$$
$$v_{id} \geq \varepsilon, \qquad i = 1, \dots, m$$

In other words, we introduce an additional variable θ that is maximized in the goal function. This variable is interpreted as the lower bound of the efficiency score.

Since this approach optimizes for the minimum average efficiency score it will seek that the minimum score is greater, but still optimizing the efficiency score of the DMU being observed. This is in accordance with the definition of justice by fairness by John Rawls. More specifically, efficiency scores are calculated to the greatest benefit of the least-advantaged DMU. Because of such an approach, we argue that the variance of the model is going to be lower. More specifically, during the optimization process variables are not going to be overfitted to the best possible efficiency score for one alternative but regularized for the efficiency score of the worst-off. This will lower

the efficiency score of the observed DMU, thus reduce variance (overfitting) of the model, reducing the need for human experts and weight restrictions. An additional benefit of the proposed approach is that the procedure of getting the cross-efficiency scores that are to be considered fair is solved using only one mathematical model.

However, the downside is that compensation exists. DMUs must adjust their weights in such a manner that it will not result in the best possible score for all observed DMUs. This can be justified as positive discrimination in algorithmic decision-making, and that achieving fairness and more equitable results have a cost [21], [22].

4.2. Data

To test the proposed mathematical model, we use data regarding healthcare systems [23]. The dataset explains the state of healthcare in 41 countries in the year 2015. Countries included in this research are OECD and BRICS countries. It is worth noticing that Brazil was excluded from the research due to data unavailability.

The choice of inputs and outputs was guided by the literature review conducted by [23]. More specifically, two outputs were commonly used for accessing the efficiency of the healthcare systems. Those are:

(O1) life expectancy – Expected number of years of life for a newborn if the mortality rate does not change. These values include both male and female populations, and

(O2) infant survival rate – Presents the ratio of children who survived the first year of life, and the number of children who died at birth or within the first year of life.

The inputs in this study are:

(I1) health expenditure per capita – This value explains health costs that include the cost of healthcare products and services. It is measured as dollars per capita using the current pricing index and the current purchasing power parity of the population.

(I2) number of medical doctors per 1000 inhabitants – Refers to all doctors that provide direct care to patients and those doctors that act as managers, or educators. The number of medical doctors is normalized to 1000 inhabitants.

(I3) number nurses per 1000 inhabitants – This number represents medical technicians who provide direct care to patients. The number of nurses is normalized to 1000 inhabitants.

(I4) number of hospital beds per 1000 inhabitants – This value presents the available number of hospital beds that are maintained and available to use. It is measured per 1000 inhabitants.

The values for the inputs and outputs are obtained from the World Bank (https://data.worldbank.org/) and the OECD database (https://stats.oecd.org/).

4.3. Experimental Setup

The effectiveness of the proposed approach is tested by comparing with the cross-efficiency scores obtained from the plain DEA model (as provided in (2)), as well as Doyle & Green cross-efficiency model (as provided in (4)). The choice of parameter ε is set to 0.001 for all of the proposed models. This parameter value aims at including all inputs and outputs during the efficiency calculation.

The cross-efficiency score is calculated using the average efficiency score as presented (3). Since cross-efficiency is used for DMU ranking, greater efficiency scores show better utility, thus better ranking. However, the value of the cross-efficiency score is irrelevant. The best DMU is going to be the one with the highest cross-efficiency score regardless if the score is close to one, or close to 0.5.

The issue with a cross-efficiency score is in the certainty of ranking. High variance in the decision model, especially the one with high degrees of freedom can artificially increase or decrease the ranking [24]. Therefore, we present standard deviations of the cross-efficiency scores as well. One can understand the standard deviations as the sensitivity analysis of the proposed approach. More specifically, by changing the values of the inputs and outputs by smaller intensities the efficiency score should not differ a lot. If the efficiency score of a DMU changes a lot, it is an indicator that DMU is unstable and could change the efficiency score with a small perturbation in underlying data. Additionally, since the proposed approach minimized the lowest average efficiency score, we present the minimum efficiency score obtained by the DMU.

5. Results and discussion

After performing the experiments on the previously described dataset, we obtained the results presented in Table 1. Column *DEA* represents classical DEA cross-efficiency score, *D&G* represents a Doyle & Green adaptation of classical DEA cross-efficiency scores explained in the previous section, and finally, column *MAX-MIN* is the proposed approach. The table is divided into three parts. In the first section mean cross-efficiency score is presented, while in the second standard deviation, and finally in the third part minimum efficiency score is obtained. The best

performances for each DMU are presented in bold letters.

As one can observe from Table 1 the cross-efficiency scores (first three columns) obtained with the Doyle & Green approach generated the highest average cross-efficiency scores. By observing the results, we can see that Slovenia (SVN), India (IND), and Indonesia (IDN) have score one, thus one cannot discriminate between these alternatives. Our proposed approach can discriminate between the proposed approach. However, as a result, our MAX-MIN approach has generally lower cross-efficiency scores. This is due to the regularization strategy during the DEA model learning.

Table 1. Comparison of the cross-efficiency results

DMU	Mean cross-efficiency score			Standard deviation			Minimum efficiency score		
	DEA	D&G	MAX-MIN	DEA	D&G	MAX-MIN	DEA	D&G	MAX-MIN
AUS	0.428	**0.528**	0.434	0.091	0.183	**0.052**	0.310	0.310	**0.336**
AUT	0.387	**0.454**	0.324	0.119	0.188	**0.040**	**0.282**	**0.282**	0.240
BEL	0.369	**0.465**	0.377	0.055	0.183	**0.035**	**0.283**	**0.283**	0.278
CAN	0.387	**0.521**	0.423	0.137	0.227	**0.083**	0.214	0.214	**0.246**
CZE	0.593	**0.664**	0.484	0.124	0.174	**0.063**	0.418	**0.446**	0.336
DNK	0.373	**0.451**	0.349	0.189	0.228	**0.085**	0.223	0.227	**0.237**
FIN	0.733	**0.799**	0.707	0.171	0.178	**0.081**	0.541	0.555	**0.567**
FRA	0.340	**0.441**	0.344	0.047	0.191	**0.028**	**0.268**	**0.268**	0.256
DEU	0.294	**0.368**	0.292	0.033	0.178	**0.023**	0.251	**0.253**	0.211
GRC	0.613	**0.660**	0.348	0.310	0.294	**0.098**	0.171	**0.179**	0.165
HUN	0.419	**0.519**	0.344	0.128	0.221	**0.056**	0.233	**0.253**	0.203
ISL	0.626	**0.701**	0.587	0.238	0.241	**0.101**	0.406	0.415	**0.436**
IRL	0.449	**0.555**	0.469	0.165	0.215	**0.088**	0.249	0.249	**0.281**
ITA	0.660	**0.763**	0.567	0.144	0.184	**0.067**	0.413	**0.423**	0.384
JPN	0.539	**0.590**	0.448	0.208	0.224	**0.110**	0.276	**0.298**	0.215
KOR	0.589	**0.658**	0.417	0.256	0.271	**0.117**	0.223	**0.243**	0.182
LUX	0.435	**0.524**	0.452	0.117	0.181	**0.065**	0.233	0.233	**0.262**
MEX	0.363	**0.622**	0.388	0.072	0.263	**0.070**	0.171	**0.199**	0.184
NLD	0.409	**0.507**	0.415	0.089	0.182	**0.050**	0.255	0.256	**0.283**
NZL	0.353	**0.488**	0.384	0.105	0.224	**0.068**	0.257	0.257	**0.284**
NOR	0.492	**0.556**	0.471	0.193	0.213	**0.084**	0.310	0.310	**0.337**
POL	0.519	**0.638**	0.418	0.173	0.253	**0.076**	0.259	**0.283**	0.224
PRT	0.651	**0.738**	0.523	0.123	0.167	**0.073**	0.344	**0.352**	0.321
SVK	0.372	**0.485**	0.318	0.104	0.226	**0.045**	0.232	**0.247**	0.211
ESP	0.774	**0.872**	0.633	0.195	0.191	**0.083**	0.441	**0.451**	0.410
SWE	0.618	**0.694**	0.549	0.273	0.269	**0.109**	0.336	0.336	**0.372**
CHE	0.263	**0.341**	0.284	0.094	0.193	**0.055**	0.153	0.153	**0.173**
TUR	0.447	**0.727**	0.414	0.180	0.280	**0.069**	0.236	0.248	**0.268**
GBR	0.489	**0.623**	0.499	0.140	0.219	**0.077**	0.280	0.280	**0.316**
USA	0.273	**0.371**	0.304	0.128	0.209	**0.083**	0.086	0.086	**0.106**
CHL	0.579	**0.810**	0.508	0.241	0.274	**0.082**	0.310	**0.339**	0.310
EST	0.770	**0.851**	0.605	0.178	0.185	**0.089**	0.534	**0.542**	0.438
ISR	0.723	**0.841**	0.606	0.168	0.183	**0.075**	0.432	**0.444**	0.403
SVN	0.997	**1.000**	0.998	0.012	**0.000**	0.055	0.928	**1.000**	0.930
CHN	0.486	**0.711**	0.381	0.230	0.298	**0.082**	0.193	0.207	**0.208**
IND	0.428	**1.000**	0.683	0.218	0.292	**0.148**	0.225	0.288	**0.298**
IDN	0.648	**1.000**	0.816	0.299	0.312	**0.114**	0.249	0.273	**0.391**
LVA	0.553	**0.653**	0.409	0.203	0.253	**0.082**	0.293	**0.312**	0.253
LTU	0.381	**0.459**	0.301	0.124	0.205	**0.051**	0.232	**0.246**	0.201
RUS	0.260	**0.341**	0.202	0.140	0.230	**0.040**	0.124	**0.134**	0.118
ZAF	0.226	**0.597**	0.312	0.113	0.330	**0.049**	0.095	0.101	**0.168**

As it is of immense importance for ranking using DEA approach, the variance of the cross-efficiency score is much lower. The proposed MAX-MIN approach has a lower standard deviation by a factor of even 5 that indicates that the efficiency score is stable

over the DEA models, thus making ranking stable regardless of the DMU being observed. In risk-averse

decision-making problems, one would like to be certain that the decision being made is not wrong, and we argue that our MAX-MIN approach is better in those cases.

If we observe the minimum efficiency score we can see that Doyle & Green and the proposed MAX-MIN score are always obtaining the best possible score. Doyle & Green obtained the best possible score for 22 countries, while the proposed MAX-MIN approach obtained the best minimum efficiency score for 19 countries. This is due to the compensation procedure we used for each DEA model. In other words, there is a trade-off between a good efficiency score and the best minimum efficiency score. Some DMU obtained a good efficiency score at the expense of the DMU with a lower efficiency score. In addition, there is a trade-off between standard deviation and mean (and minimum) efficiency score. The level of regularization we utilized resulted in generally lower efficiency scores (thus in lower values of minimum efficiency scores) but also resulted in a more stable ranking.

By inspecting the standard deviation of the cross-efficiency scores, we inspect the sensitivity of the proposed approach. We can note that the standard deviation is the lowest for the proposed approach majority of the DMUs. The decrease of standard deviation indicates that the proposed approach is not sensitive to smaller changes in the values of input and output attributes, and consequently more stable in efficiency scores. For some DMUs, Doyle & Green cross-efficiency procedure can result in standard deviations greater than 0.200. Even though that efficiency scores (for a single DMU) do not fit some known distribution, a high standard deviation indicates that the real efficiency score can be either very high or very low, making the average efficiency score less usable for decision-making.

Finally, we can say that MAX-MIN fairness resulted in a fairer ranking. Some DMU obtained unfair high (or low) efficiency scores due to unfair comparison by belonging to the majority group (or minority group i.e. outliers). For example, Greece (GRC), Japan (JPN), Sweden (SWE), Latvia (LVA), and Estonia (EST) performed well using DEA and D&G because those countries belong to a group of countries that are specific and differ a lot from the rest of the dataset. Their score is very high. However, when the efficiency score is calculated using weights of other DMUs, the efficiency score reduces results in a high variance estimation of the efficiency score. However, there are different examples. For example, Belgium

(BEL), Canada (CAN), Czech Republic (CHE), the USA, or Indonesia (IDN) benefited from the MAX-MIN cross-efficiency model. In those cases, both efficiency scores improved (compared to DEA model) and variance as well. For those countries, the proposed approach increased minimum efficiency scores as well. This situation can be explained that compensation for those was greater than the reduction of efficiency score of other DMUs. Therefore, the optimization procedure opted for increasing their efficiency score for a lower reduction in efficiency scores of other DMUs.

If we analyze the basic DEA scores, only eight countries we efficient. Those are Slovenia (SVN), Greece (GRC), Sweden (SWE), Japan (JPN), India (IDN), Spain (ESP), Chile (CHL), and Indonesia (IND). The reason why these countries resulted in efficient scores was the fact that they had a good combination of exactly one input and exactly one output. More specifically, the weight associated with one input is high, while weights associated with other inputs are at the constraint minimum ($\varepsilon = 0.001$). This is not the property one would like to use for decision-making. Cross-efficiency helps us dealing with that issue. Doyle & Green method regards only Slovenia, Spain, and Estonia as the efficient ones (using the best efficiency score, not the mean). The weights associated with the inputs and outputs are diversified in the sense that one DMU needs to have multiple inputs and multiple outputs with a weight greater than ε. However, some of them remain at 0.001. The proposed MAX-MIN approach requires that all weights are greater than 0.001.

Although it is not visible from the efficiency scores, the proposed approach has the advantage that it solves the problem at hand in a single optimization procedure. Other cross-efficiency methods must optimize the DEA mathematical model for each DMU, and then aggregate it. The downside of this approach is in larger optimization size with $(n(m \times s) + 1)$ variables and $(n^2 + n(s \times m))$ constraints. To make this benefit more clear, we measured the time needed for the execution of the process. The process is repeated 10 times using the same configuration. Basic DEA procedure needed on average 6.023 seconds, Doyle & Green 10.427 seconds, while the proposed approach 8.529 seconds.

6. Conclusions

In the spirit of recent events regarding social justice and fairness, we propose an approach for fair cross-efficiency estimation. The notion of fairness utilized in this paper is John Rawls's theory of justice. More specifically, fairness is implemented such that efficiency scores are estimated to be to the greatest

benefit of the least-advantaged DMU. More specifically, we propose a MAX-MIN formulation of the cross-efficiency DEA model. MAX-MIN formulation of the efficiency score calculates the sum of efficiency scores of two DMUs, namely the one being observed and any other DMU in such a manner that the minimum sum of efficiency scores is maximized. This allows regularization of the efficiency scores, thus preventing overfitting and preventing the need for human expert intervention, and providing weight restrictions to the model. Consequently, the variance of the cross-efficiency score is lower, making ranking based on the efficiency score better. In addition, in contrast to traditional cross-efficiency DEA models, the proposed one is learned using a single pass on the dataset. The proposed approach does have some downsides. First, the optimization problem is set such that DMUs must adjust their weights not to get the best possible efficiency score, but to compensate their score if they hurt the efficiency score of other DMUs. This might seem counterintuitive, but it acts as a counterweight to the overoptimistic estimate of the efficiency score (overfitting the efficiency to the underlying data). Second, the optimization problem is much higher in dimensionality. The number of variables and the number of constraints are increased. For large datasets (f.e. $m > 10, s > 10$, and $n > 1000$) the proposed approach will last longer than other cross-efficiency scores.

The proposed approach is tested on the dataset regarding estimation of the healthcare. The efficiency of the healthcare system has been a subject of research for many years. Policymakers want to design the healthcare system such that it is self-improving by observing the best practices. To do so, efficiency must be estimated. After applying the proposed approach, one can see several benefits. First, discrimination between the scores of DMUs exists and rankings are enabled. Second, the variance of cross-efficiency scores is reduced, and finally, fairness is introduced through minimal efficiency scores.

In the future work, we plan to introduce fairness in DEA scores and cross-efficiency scores regarding the disparate impact, equal opportunity [25], and equal odds such as in [5]. By doing so, one can correct cultural injustices toward historically disadvantaged and underserved communities. More specifically, one can enforce the absence of disparate impact in the results of the efficiency and cross-efficiency scores through constraints or regularization functions. Another line of research is concerned with welfare efficiency estimation. Instead of learning solely efficiency scores of DMU, one would like to create a trade-off between benefits for the community (or a country as a whole) and individual benefits [6].

Acknowledgments

References

[1] Cooper, W. W., Seiford, L. M., & Zhu, J. (Eds.). (2011). *Handbook on Data Envelopment Analysis*. Springer.

[2] Radojicic, M., Savic, G., & Jeremic, V. (2018). Measuring the efficiency of banks: the bootstrapped I-distance GAR DEA approach. *Technological and Economic Development of Economy, 24*(4), 1581-1605.

[3] Doyle, J., & Green, R. (1994). Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. *Journal of the Operational Research Society, 45*(5), 567-578.

[4] Binns, R. (2018). What can political philosophy teach us about algorithmic fairness?. *IEEE Security & Privacy*, *16*(3), 73-80.

[5] Radovanović, S., Petrović, A., Delibašić, B., & Suknović, M. (2020, August). Enforcing fairness in logistic regression algorithm. In *2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)* (pp. 1-7). IEEE.

[6] Kasy, M., & Abebe, R. (2021, March). Fairness, equality, and power in algorithmic decision-making. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 576-586).

[7] Kleinberg, J. (2018, June). Inherent trade-offs in algorithmic fairness. In *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems* (pp. 40-40).

[8] Valera, I. Discrimination in Algorithmic Decision Making. *Fundamental Questions*, 15.

[9] Petrović, A., Nikolić, M., Radovanović, S., Delibašić, B., & Jovanović, M. (2020). FAIR: Fair Adversarial Instance Re-weighting. *arXiv preprint arXiv:2011.07495*.

[10] Van Eijk, G. (2017). Socioeconomic marginality in sentencing: The built-in bias in risk assessment tools and the reproduction of social inequality. *Punishment & Society*, *19*(4), 463-481.

[11] Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated Experiments on Ad Privacy Settings. *Proc. Priv. Enhancing Technol., 2015*(1), 92-112.

[12] Kumbhakar, S. C., Parmeter, C. F., & Zelenyuk, V. (2020). Stochastic frontier analysis: Foundations and Advances I. *Handbook of Production Economics*, 1-39.

[13] Emrouznejad, A., & Yang, G. L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-economic planning sciences*, *61*, 4-8.

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

[15] Zhu, J. (2014). DEA Cross Efficiency. In *Quantitative Models for Performance Evaluation and Benchmarking* (pp. 61-92). Springer, Cham.

[16]   Doyle, J., & Green, R. (1994). Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. *Journal of the Operational Research Society*, *45*(5), 567-578.

[17]   Wang, Y. M., & Chin, K. S. (2010). Some alternative models for DEA cross-efficiency evaluation. *International Journal of Production Economics*, *128*(1), 332-338.

[18]   Wu, J., Chu, J., Sun, J., & Zhu, Q. (2016). DEA cross-efficiency evaluation based on Pareto improvement. *European Journal of Operational Research*, *248*(2), 571-579.

[19]   Kuppler, M., Kern, C., Bach, R. L., & Kreuter, F. (2021). Distributive justice and fairness metrics in automated decision-making: How much overlap is there?. *arXiv preprint arXiv:2105.01441*.

[20]   Boyd, S. & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

[21]   Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017, August). Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SigKDD International Conference on Knowledge Discovery and Data Mining* (pp. 797-806).

[22]   Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, *31*(4), 611-627.

[23]   Radojicic, M., Jeremic, V., & Savic, G. (2020). Going beyond health efficiency: What really matters?. *The International Journal of Health Planning and Management*, *35*(1), 318-338. Valcarce, D., Bellogín, A.,

[24]   Parapar, J., & Castells, P. (2018, September). On the robustness and discriminative power of information retrieval metrics for top-n recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems* (pp. 260-268).

[25]   Radovanović, S., & Ivić, M. (2021). Enabling Equal Opportunity in Logistic Regression Algorithm. *Management: Journal of Sustainable Business and Management Solutions in Emerging Economies*.