SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2025
ORGANIZATION AND MANAGEMENT SERIES NO. 226

REVISITING DEA APPLICATION IN THE JUDICIARY: ASSESSING MACRO-EFFICIENCY OF THE POLISH DISTRICT COURT SYSTEM IN THE 21ST CENTURY

Waldemar FLORCZAK

Jagiellonian University in Cracow; waldemar.florczak@uj.edu.pl, ORCID: 0000-0003-3408-7784

Purpose: The objective of this article is to investigate the technical efficiency of Poland's district court system from 2002 to 2021, employing a constrained variant of Data Envelopment Analysis (DEA) on appropriately processed time series data.
Design/methodology/approach: Different DEA models were employed to estimate the technical efficiency of this system, beginning with classical CCR DEA and super-efficiency DEA. After identifying crucial shortcomings in the outcomes produced by these two methods, a novel procedure based on constrained DEA was proposed to address and overcome these limitations in the context of evaluating judicial efficiency using DEA.
Findings: Applying super-efficiency DEA to a small sample does not yield sound or fully interpretable outcomes regarding the estimated scores of individual DMUs, despite seemingly increasing the variability of these scores. It is only when adequate restrictions are imposed on both the input and output weights—price and virtual alike—that one can obtain results free from the significant imperfections commonly associated with the classical DEA method in the context of judicial efficiency.
Research limitations/implications: Ignoring the inherent shortcomings of classical or super-efficiency DEA can lead to significant distortions in the estimated ratings assigned to individual DMUs.
Practical implications: Policy recommendations derived from traditional unconstrained DEA, intended to enhance the actual efficiency of specific DMUs, might ultimately be misleading.
Originality/value: The research contributes to the existing body of knowledge on judicial efficiency in at least two significant ways. First, unlike previous studies, it focuses on a single DMU (the first-instance court system) over a long 20-year period, allowing conclusions drawn from this investigation to be interpreted within the context of systemic macro-level efficiency over time. Secondly, and more importantly, the article highlights the inherent weaknesses of classical DEA, which cast reasonable doubt on the quality and interpretability of the results obtained using this method. To address these issues, the article introduces—for the first time in the field of judicial efficiency research—a procedure specifically designed to overcome these limitations.
Keywords: efficiency, DEA, weight constraints, judiciary.
Category of the paper: Research article.

1. Introduction

The aim of this paper is to illustrate the possibility to estimate the judicial technical efficiency using time-series data – as opposed to cross-sectional data typifying all the previous research on this subject matter – with Poland's district court system serving as the reference. The macro perspective adopted enables a holistic assessment of the combined efficacy of the legal, administrative, procedural, organizational, and economic frameworks—uniformly applied to all district courts—that collectively influence the system's functioning and thus determine its systemic efficiency across individual years in the analysis, something that cannot be achieved with the use of cross-sectional data because this macro-level context remains constant then.

A parallel methodological goal of the study is to draw attention to the immanent weaknesses of classical DEA, including its super-efficiency variant which cast reasonable doubt on the quality and interpretability of the results obtained using this class of methods. At the same time, the article proposes - for the first time in the field of judicial efficiency research - a procedure to address these weaknesses within the framework of existing modifications of classical DEA.

The set of research hypotheses subjected to empirical verification in this study is as follows:

H1: Neither classical DEA nor super-efficiency DEA produces fully interpretable weight estimates during the efficiency optimization process.

H2: Previous efforts to ensure unambiguous interpretability of results in the field of judicial efficiency research have proven insufficient.

H3: Policy recommendations derived from classical DEA in the context of judicial efficiency may ultimately be unreliable or misleading.

2. Methodological note and literature outline

The starting point of the classical DEA method (e.g., Panwar et al., 2022) is the following quotient of the weighted sum of outputs to the weighted sum of inputs, defining the technical efficiency of a given decision-making unit, DMU, (for ease of reading, the equivalents of general terms used in this study are provided in parentheses):

$$E_z = \frac{\sum_{t=1}^{T} w_{tz} Y_{tz}}{\sum_{s=1}^{S} v_{sz} X_{sz}} \tag{1}$$

where:

E_z – technical efficiency of the *z-th* decision-making unit (efficiency of the district court system in each year of the 2002-2021 period),

$z = 1, 2 \ldots, Z$ – index of the *z-th* decision-making unit (total Z = 20, i.e. number of years),

X_{sz} – the $s\text{-}th$ type of input in the $z\text{-}th$ decision-making unit (the amount of the $s\text{-}th$ input - out of the three considered in the study: the number of judges, outlays on the disctrict court system and the system load - in a given year of analysis),

s = 1, 2 …, S – index of the $s\text{-}th$ input (total of S = 3 types of inputs, listed above, common to all decision-making units),

Y_{tz} – $t\text{-}th$ type of output in the $z\text{-}th$ decision-making unit (amount of the $t\text{-}th$ type of output - out of the six included in the study: number of resolved civil, criminal, family, labour law, social security and commerce cases - in the given year of analysis),

t = 1, 2 …, T – index of the $t\text{-}th$ input (total of T = 6 output types, listed above, common to all decision-making units,

w_{tz}, v_{sz} – weights assigned to the $t\text{-}th$ outcome and $s\text{-}th$ input in the $z\text{-}th$ DMU, respectively.

A key property of the DEA method is that the weights appearing in formula (1) are determined through a linear programming procedure, rather than by arbitrarily assigning specific values to them. Thus, the most important—and at the same time, most disputable—issue of determining the values of these weights, which implicitly conveys the gradation, importance, or significance of individual inputs and outputs, is objectivized in a way that ensures the most favourable set from the perspective of individual decision-making units. This is achieved by maximizing outputs and minimizing inputs, represented by the following mathematical problem (e.g., Panwar et al., 2022):

$$max\ E_z = \frac{\sum_{t=1}^{T} w_{tz} Y_{tz}}{\sum_{s=1}^{S} v_{sz} X_{sz}} \tag{2}$$

with respect to the following constraints:

$$0 \leq \frac{\sum_{t=1}^{T} w_{tz} Y_{tz}}{\sum_{s=1}^{S} v_{sz} X_{sz}} \leq 1 \tag{3}$$

$$0 \leq w_{tz} \leq 1;\ 0 \leq v_{sz} \leq 1 \tag{4}$$

The bulk of the existent empirical applications in the public sector, including judiciary, use the above-reported constant returns to scale CRS-DEA scheme to assess the efficiency of various public units. A comprehensive review of research on judicial efficiency can be found in works by Voigt (2016) or Ippoliti and Tria (2020).

There is a great deal of variation in empirical studies as far as selection of input and output variables are concerned. In the nature of inputs, the "duty variable" - completely understandably - is almost always (exceptions is e.g. Tulkens, 1993) the number of judges, while other variables are selected most likely depending on the availability of data. In addition to the aforementioned variable, DEA studies of court efficiency take into account such variables as the number of cases outstanding/caseload (this is the variable most often present - along with the number of judges: e.g. Mattsson, Tidana 2019), the size of non-judicial staff (e.g. Angrell et al., 2019), office and IT equipment and work space (e.g., Angrell et al., 2019), or expenditures on court operations (e.g., Deynelli, 2012). It should be emphasized that in the vast majority of DEA applications in the judiciary the number of inputs simultaneously considered does not exceed three (Peyrache, Zago 2016 is an exception).

The vast majority of empirical applications include as outputs an aggregate of cases resolved in a given year, and therefore only one factor. Given the variation in the duration of litigation and labour input by subject of law, treating all cases coming before the courts as homogeneous is certainly a very far-reaching simplification. This raises the question of whether the efficiency scores achieved by individual courts can be reliably inferred from such far-fetched aggregate results, covering all types of legal cases. Exceptions in this regard include only the work of Kittelsen & Forsund (1992; seven categories of cases by subject of law); Tulkens (1993; four categories of cases); Mattsson & Tidana (2019; three categories of cases), Angrell et al. (2019; three categories of cases), and the work of Santos & Amado (2014; a very exhaustive breakdown of cases by subject of law: 43 categories).

In all the works cited in this micro-review, the outcome-maximising version of DEA is used. This approach is justified by the specifics of the operation of the judiciary, where an increasing burden on court systems can be observed in recent years. The only exception is the work of Schneider (2005). Furthermore, in cases where a non-discretionary variable is present in the set of inputs (e.g. caseload), it can only be treated on same terms as a discretionary variable if the chosen version of DEA is outputs maximising (see Cook et al., 2014).

Classical DEA exhibits low discrimination of efficiency scores between the analysed DMUs when the number of DMUs is small. Researchers typically address this issue by employing a variant of DEA that utilizes the concept of super efficiency (known as the SE-DEA model), the methodological details of which can be found e.g. in Lee et al. (2011). An example of application of the super-efficiency concept in which the aforementioned phenomena are present is e.g. Mattsson & Tidana (2019).

However, neither classical DEA-CRS nor its SE-DEA extension prevents the results from including inputs and/or outputs with zero weights, which, while consistent with the procedures (see constraint (4)), raises concerns about the validity of such outcomes (discussed further). To ensure that all inputs and outputs contribute meaningfully to the analysis —effectively ensuring that no weights are zero — a constrained SE-DEA model can be employed (e.g. Pedraja-Chaparro et al., 1997). This version of DEA, though not yet applied in the context of judicial efficiency, forces the optimal weights to adhere to specific assumptions based on logical considerations, ensuring that none of the weights are zero and thus eliminating the previously noted objections.

3. Data

In order to carry out the task at hand, it was necessary to collect data at the macroeconomic level, which involved using various sources and processing the information accordingly. Data on regional court activity by different categories of cases (including criminal, civil, family, labour, social security, and business cases) in 2011-2021 were taken from publicly available resources of the Ministry of Justice. The key source was the .xls-format database "Records of Cases in Common Courts in Poland," available on the Ministry's website (https://isws.ms.gov.pl/pl/baza-statystyczna/opracowania-wieloletnie). To expand the period of analysis and include the years 2002-2010, the document "Statistical Analysis of the Activities of the Judiciary in 2002-2011" (https://isws.ms.gov.pl/pl/baza-statystyczna/opracowania-jednoroczne-w-tym-pliki-dostepne-cyfrowo/rok-2011/) was additionally used. In addition, the collected data required further modifications to ensure it could be properly applied to efficiency analyses using the DEA method. In particular, this included the need for some identity-related additions concerning the basic relationships derived from the statistical court records, which were crucial for the calculations to be carried out correctly:

$$Caseload_t = Pending\ cases_{t-1} + Incoming\ cases_t \quad (5)$$

$$Pending\ cases_t = Caseload_t - Resolved\ cases_t \quad (6)$$

$$Case\ processing\ time_t = \frac{Pending\ cases_t}{Resolved\ cases_t} \cdot 12 \quad (7)$$

where the subscript "*t*" stands for *t-th* year.

The number of cases completed by subject of law were chosen as the output variables describing the effects of regional courts activity. The selection of such variables ensures that the conditions of homogeneity and completeness are met for the decision-making units under study (here: individual years of district courts activity), which is a key requirement for the correctness of DEA analyses. The results defined in this way reflect the effects of the courts' activities over time. Input variables were drawn from several sources. The non-discretionary input variable (caseload), reflecting the burden on the judicial system, was calculated based on formula (5). On the other hand, data on the average number of judges employed in the regional courts from 2006 to 2021 were obtained directly from the Ministry of Justice through the public information access procedure. For the years 2002-2005, for which direct data on the number of judges was missing, linear interpolation was used between 2001 (CSO data) and 2006.

To estimate the outlays from the state budget for the operation of regional courts adjusted for inflation, it was necessary to conduct the appropriate calculations and use several sources. First, a time series was created for expenditures on the general judiciary in current prices (which includes district courts) based on data from the Ministry of Finance available at https://www.gov.pl/web/finanse/sprawozdania-roczne and https://mf-arch2.mf.gov.pl/web/bip/

ministerstwo-finansow/dzialalnosc/finanse-publiczne/budzet-panstwa/wykonanie-budzetu-panstwa/sprawozdanie-z-wykonania-budzetu-panstwa-roczne.

Second, the results of calculations presented in the work of Siemaszko & Ostaszewski (2013) were used to determine the proportion of expenditures on district courts to total expenditures on the general judiciary. Third, budget expenditures for district courts were estimated in current prices by multiplying total expenditures on the general judiciary by the established fraction allocated to the district judiciary. Finally, the values of this series were divided by the Consumer Price Index (CPI) deflator for 2015 to obtain the real value of these outlays[1].

4. Empirical analyses

The second column of Table 1 presents the efficiency scores obtained using the classic DEA model. As anticipated, these scores are the least discriminatory—not primarily due to substantive factors, but rather for technical reasons. In such cases, the discriminative power of the DEA model can be enhanced by employing the concept of super-efficiency. In this setup, efficiency scores may exceed the value of 1 to varying extents, allowing for greater discrimination among "classically" efficient DMUs that would otherwise be capped at 1. Rescaling these scores to set the highest score to "1" enables a more "traditional" interpretation of the results obtained.

Table 1 also presents the efficiency scores obtained through the DEA super-efficiency procedure, including both original and rescaled scores. At first glance, it appears that the results are quite satisfactory. Despite the limited number of DMUs, the analysis has seemingly allowed for a robust discrimination of their efficiency scores, resulting in a diverse empirical set of efficiency scores that may give an illusory impression of greater realism. In fact, this is often where researchers investigating court efficiency through DEA tend to stop, with the majority ceasing their analysis at the classic DEA stage.

One might wonder why one shouldn't interpret the results immediately. What would be wrong with such an interpretation? The adage, "The less people know about how sausages and laws are made, the better they'll sleep at night", attributed to Otto von Bismarck, can be paraphrased for our context: "The less DEA practitioners (and even more so, DEA stakeholders) know about the parameters and weights derived from their DEA analyses, the better they sleep at night". Indeed, it seems that few care about what might be perceived as "trivial" matters, such as the actual parameter estimates obtained in the DEA model. However, these are not trivial issues; they are crucial for trusting the obtained efficiency scores. The fact that only one

[1] Data available upon request.

of the publications referenced in this paper (Santos, Amado, 2014) explicitly addresses the issue of parameter estimates and weights—though not exhaustively—illustrates the extent of both intentional and unintentional ignorance on this matter.

Table 1.
Efficiency scores and absolute weights obtained in the classical and super-efficiency DEA

DMU	Efficiency scores			Absolute weights estimates in SE-DEA								
	Classical DEA	SE-DEA	Rescaled SE-DEA	INPUTS			OUTPUTS					
				OUTLAY	*JUDGE*	*CASELOAD*	*COMMERCE*	*LABOUR*	*CIVIL*	*FAMILY*	*CRIME*	*INSURANCE*
D02	1.0000	**1.1043**	0.9761	0.0003651	0	0	0.0004914	0.0002414	0	0.0004841	0	0
D03	1.0000	**1.1275**	0.9967	0.0003124	0	0	0	0.0020802	0	0	0	0.0047333
D04	1.0000	**1.0846**	0.9587	0	0.0001322	0	0	0	0	0	0.0002298	0.0092038
D05	1.0000	**1.0460**	0.9247	0	0	9.547E-05	0	0	0	0.0007915	0	0.0031389
D06	1.0000	**1.0220**	0.9034	0	0	9.456E-05	0	0.0003931	0	0	0.0004415	0
D07	1.0000	**1.0275**	0.9083	0	5.171E-05	5.41E-05	0	0	0	0	0.0004389	0
D08	1.0000	**1.0170**	0.8990	0	1.62E-05	7.394E-05	0	0	0	0.0007865	0	0
D09	1.0000	**1.0469**	0.9254	0.0001663	1.995E-05	1.609E-05	0	0	0	0.0004117	0.0001957	0
D10	1.0000	**1.0155**	0.8977	0	0	7.37E-05	0	0	9.82E-05	0.0002155	0	0
D11	0.9884	**0.9884**	0.8737	0	0	7.163E-05	4.556E-05	0	9.259E-05	0	0.0001086	0
D12	1.0000	**1.0278**	0.9086	6.55E-05	0	4.94E-05	0	0	0.0001072	4.512E-05	3.37E-05	0
D13	1.0000	**1.0177**	0.8996	0.0002549	0	0	0.0001885	0	2.791E-05	0.0003605	3.735E-08	0
D14	1.0000	**1.0398**	0.9191	3.415E-05	0.0001236	0	0	0.0009049	0	0	0.000372	0
D15	1.0000	**1.0344**	0.9144	0	1.474E-05	5.411E-05	0.0004108	0	0	0	0.000141	0
D16	1.0000	**1.0015**	0.8853	0	0.00015	0	2.188E-05	0	0	0.0004584	0	0.008921
D17	1.0000	**1.0986**	0.9712	2.068E-05	0.0001198	0	0.0005435	0	0	0	0	0
D18	0.9952	**0.9952**	0.8797	0	8.501E-06	5.683E-05	0	0	0.0001086	0	0	0.0006618
D19	1.0000	**1.1313**	1.0000	9.276E-06	0.0001259	0	0	0	9.531E-05	0	0	0
D20	0.9852	**0.9852**	0.8709	0	8.42E-06	5.629E-05	0	0	0.0001075	0	0	0.0006554
D21	0.9776	**0.9776**	0.8641	0	4.223E-05	4.453E-05	0	0	8.646E-05	0.0001631	0	0.0008154

Source: Own computations using processed data (see section *Data*) and *pydea* software.

Let's examine the weights and parameters attached to the outputs in the super-efficiency variant of DEA (for want of space only absolute weights estimates are reported in Table 1). It is evident that many weights assume the value of "0", and occasionally, all weights except one are "0". Interpreting these weights in an economic sense leads to a fully logical yet hardly realistic conclusion that there is no contribution from the factors (inputs/outputs) whose weights equal "0". Such a conclusion, which is unavoidable in this scenario, undermines the fundamentals of DEA as a credible tool for efficiency estimation. This is because all inputs and outputs included in the DEA must have explicit contributing power, otherwise, they should not have been included in the analysis at all!

Thus, a question arises as to whether such outcomes have any explanatory merit. If not—and this seems to be the only logical conclusion—the next question is what can be done about it. A self-evident and very general answer would be: to ensure that the estimated weights make sense. This can be accomplished with the help of constrained DEA, where the optimization of the efficiency scores takes into account certain logically and empirically sound restrictions imposed on the weights during the optimization process.

It seems quite straightforward to "price" the unit value of a given resolved case (by subject of law) according to the time spent on its disposition. After all, "time is money", especially when the actual costs incurred in resolving an average case cannot be credibly determined. In addition, with the help of the super-efficiency concept and the constraints imposed on absolute and virtual weights—both on the side of outputs and inputs—one can obtain apparently sound results regarding the ratings of individual decision-making units, with no anomalies

(0-1 values) concerning the weights attached to the production factors. The entire procedure is outlined in the following sections.

Let us start with a proposal made by Santos & Amado (2014), where the authors established appropriate minority/majority relations between the output weights assigned to various cases by subject of law in line with their average time absorption, as shown in formula (7). Consequently, using their approach leads us to the following set of relations regarding the price weights assigned to the respective cases by subject of law[2]:

INSURANCE / *LABOUR* >= 1;

LABOUR / *CRIME* >= 1;

CRIME / *CIVIL* >= 1;

CIVIL / *FAMILY* >= 1;

CIVIL / *COMMERCE* >= 1

Although novel compared to all previous work in the investigated field, the work of Santos & Amando (2014) does not appear to address the critical issue of the inadmissibility of output weights as proved by the outcomes obtained in this investigation: there are still numerous zero weights assigned to individual outputs (not reported here).

It is only when additional constraints are imposed—this time on the so-called virtual weights, which have a straightforward interpretation as they represent the proportion of total time spent on resolving all types of cases that was absorbed by cases of a given subject of law—that the resulting outcomes regarding output weights begin to make sense. These constraints are explicitly given by the following formula:

$$s_i = \frac{d_i \cdot n_i}{\sum_{j=1}^{6} d_i \cdot n_i} \tag{8}$$

where:

$s_i -$ fraction of the courts' time devoted to resolving the cases of the i-th subject of law (i = 1, 2, ..., 6). By definition we have: $\sum_{i=1}^{6} s_i = 1$,

$d_i -$ average processing time of cases for *i-th* subject of law, determined according to formula (7),

$n_i -$ number of cases per *i-th* subject of law.

The empirical - and thus most realistic – ranges of virtual weights for the individual types by subject of law derived with the help of formula (8) are shown in Table 2.

[2] This notation is in line with commands fixed by *pydea* software with "self-speaking" weights names. It is tantamount to imposing the afore-mentioned constraints upon respective w_{tz} in formula (1). The reported minority/majority relationships between specific subjects of law has been derived by means of a comparative analysis carried out with the use of formula (7).

Table 2.
Constraints imposed on virtual weights

Virtual weights (minima)	Virtual weights (maxima)
VFAMILY >= 0.054	*VFAMILY* <= 0.115
VCOM >= 0.036	*VCOM* <= 0.140
VCRIME >= 0.087	*VCRIME* <= 0.28
VLABOUR >= 0.007	*VLABOUR* <= 0.064
VINSURANCE >= 0.006	*VINSURANCE* <= 0.021
VCIVIL >= 0.506	*VCIVIL* <= 0.748

Source: Own computations using processed data (see section *Data*).

Applying a variant of DEA with simultaneous restrictions on both price and virtual weights ultimately results in correct weights for the outputs (outcomes not reported here). However, the issue previously described persists concerning the weights assigned to inputs. This raises the question of whether such a situation is acceptable. Before addressing this question definitively, let us consider two rhetorical subsidiary questions that, while not challenging the essential role of judges in resolving court cases, aim to provide additional perspective:

1. Would judges, tasked with delivering judgments for all pending court cases, be willing to perform their duties without remuneration?
2. Would the institution of courts and judges be necessary at all if no legal cases were brought before them?

The negative answer is therefore self-imposing. The empirical realism of the input weights must therefore also be ensured in this case, as this is a necessary condition for the reliability of the scores, generated by DEA. This leads directly to a heuristic proposal that leverages the aforementioned observation to enable the determination of realistic, empirical price weights among the used inputs, with judges serving as the critical input.

The starting point involves determining the range of empirical variability for the ratios of outlays per judge and caseload per judge. The respective extreme values (based on data described is section *Data*) during the analyzed period (2002-2021) are as follows:

min (*CASELOAD/JUDGES*) = 1.408 max (*CASELOAD/JUDGES*) = 2.911

min (*OUTLAYS/JUDGES*) = 0.363 max (*OUTLAYS/JUDGES*) = 0.817

Thus, based on the assumption of the primary role of judges in the entire process of case disposition, the price weights assigned to the other two inputs should fall within the intervals defined by the previously reported extreme points, with the absolute weights assigned to the input “judges” serving as the reference category. If the weights were too large, it would imply that the relative role of the remaining inputs was greater than that of judges. Conversely, if the weights were too small, it would undervalue the empirical significance of those inputs. Consequently, the constraints imposed on the price weights of all the inputs are as follows:

ACASELOAD / AJUDGES >= 1.408 *ACASELOAD / AJUDGE* <= 2.911

AOUTLAYS / AJUDGES >= 0.363 *AOUTLAYS / AJUDGES* <= 0.817

Table 3.
Efficiency scores and virtual weights obtained in the constrained super-efficiency DEA variant with minority/majority relationships with price and virtual weights imposed on outputs and inputs

DMU	Efficiency scores		Virtual weights estimates in the constrained SE-DEA								
	Constrained SE-DEA	Rescaled	INPUTS			OUTPUTS					
			OUTLAY	*JUDGE*	*CASELOAD*	*LABOUR*	*COMMERCE*	*FAMILY*	*INSURANCE*	*CRIME*	*CIVIL*
D02	0.85719	0.84756	0.06415	0.21621	0.88624	0.06400	0.12231	0.10016	0.02100	0.18653	0.50600
D03	0.90008	0.88997	0.06451	0.19959	0.84692	0.06400	0.07500	0.05400	0.02100	0.28000	0.50600
D04	0.94970	0.93904	0.03045	0.19087	0.83164	0.06400	0.07500	0.05400	0.02100	0.28000	0.50600
D05	0.95654	0.94580	0.03332	0.19364	0.81848	0.06400	0.05943	0.06957	0.02100	0.28000	0.50600
D06	0.94202	0.93144	0.03525	0.19159	0.83472	0.06400	0.05685	0.07215	0.02100	0.28000	0.50600
D07	0.94804	0.93740	0.03444	0.18295	0.83741	0.05151	0.06481	0.08620	0.01148	0.28000	0.50600
D08	0.95752	0.94677	0.03327	0.16714	0.84395	0.03232	0.06827	0.10567	0.00774	0.28000	0.50600
D09	0.99351	0.98236	0.03052	0.15252	0.82350	0.03579	0.03626	0.05400	0.00806	0.28000	0.58589
D10	0.99434	0.98318	0.03022	0.14358	0.83190	0.02990	0.03948	0.05400	0.00769	0.27687	0.59205
D11	0.98538	0.97432	0.02867	0.13910	0.84707	0.02011	0.04922	0.05400	0.00600	0.28000	0.59066
D12	1.01135	1.00000	0.05810	0.12683	0.80384	0.00822	0.05323	0.05400	0.00600	0.19261	0.68595
D13	0.98806	0.97696	0.10028	0.21179	0.70002	0.00991	0.09686	0.09215	0.00600	0.22165	0.57342
D14	1.00837	0.99705	0.10127	0.20758	0.68285	0.03896	0.07114	0.06795	0.02100	0.28000	0.52096
D15	1.00589	0.99460	0.04830	0.21607	0.72977	0.00802	0.12098	0.05400	0.00944	0.17344	0.63411
D16	0.94520	0.93459	0.03871	0.16606	0.85320	0.00700	0.11491	0.10058	0.02100	0.16343	0.59309
D17	0.98458	0.97353	0.10212	0.19895	0.71459	0.00700	0.12833	0.05400	0.00600	0.14395	0.66072
D18	0.95819	0.94743	0.05341	0.22056	0.76967	0.00700	0.08001	0.05400	0.00600	0.14977	0.70322
D19	0.94634	0.93571	0.05178	0.19708	0.80784	0.00700	0.07327	0.05400	0.00600	0.12914	0.73059
D20	0.91228	0.90204	0.06339	0.22066	0.81211	0.00700	0.07621	0.05400	0.00600	0.12873	0.72805
D21	0.91949	0.90917	0.06536	0.22030	0.80190	0.00700	0.06101	0.05400	0.01311	0.14621	0.71866

Source: Own computations using processed data (see section *Data*) and *pydea* software.

Only now, once all the essential constraints have been imposed on the price and virtual weights regarding both outputs and inputs, can one arrive at the final results that are free from the articulated, crucial, and adverse properties which, if ignored, cast serious doubt on the credibility and interpretability of scores obtained using classic DEA. At this stage, no weight—neither on the side of outputs nor inputs—assumes the value of "0" or "1" (see Table 3; for want of space only virtual weights are reported). Moreover, because of the way the constraints were imposed on the weights, the range of estimated weight variability falls implicitly within the historical min-max intervals observed in the analyzed period. All of this leads us to claim that the final outcomes obtained in this empirical research are fully reliable and interpretable, enabling a sound general appraisal of the actual performance of the Polish district court system over a 20-year period.

5. Conclusions and final remarks

A natural question emerging from the DEA presented thus far—an answer to which holds primary practical and political importance—is whether there are actual differences in the efficiency scores obtained through the super-efficiency concept without consideration of the obtained weights, compared to those derived using the same super-efficiency concept while ensuring logical weights. Figure 1 presents the original and normalized scores of the two primary variants of DEA utilized in this investigation: unconstrained output-oriented CRS super-efficiency DEA and its constrained counterpart. Additionally, simple trend regressions for both variants are depicted.

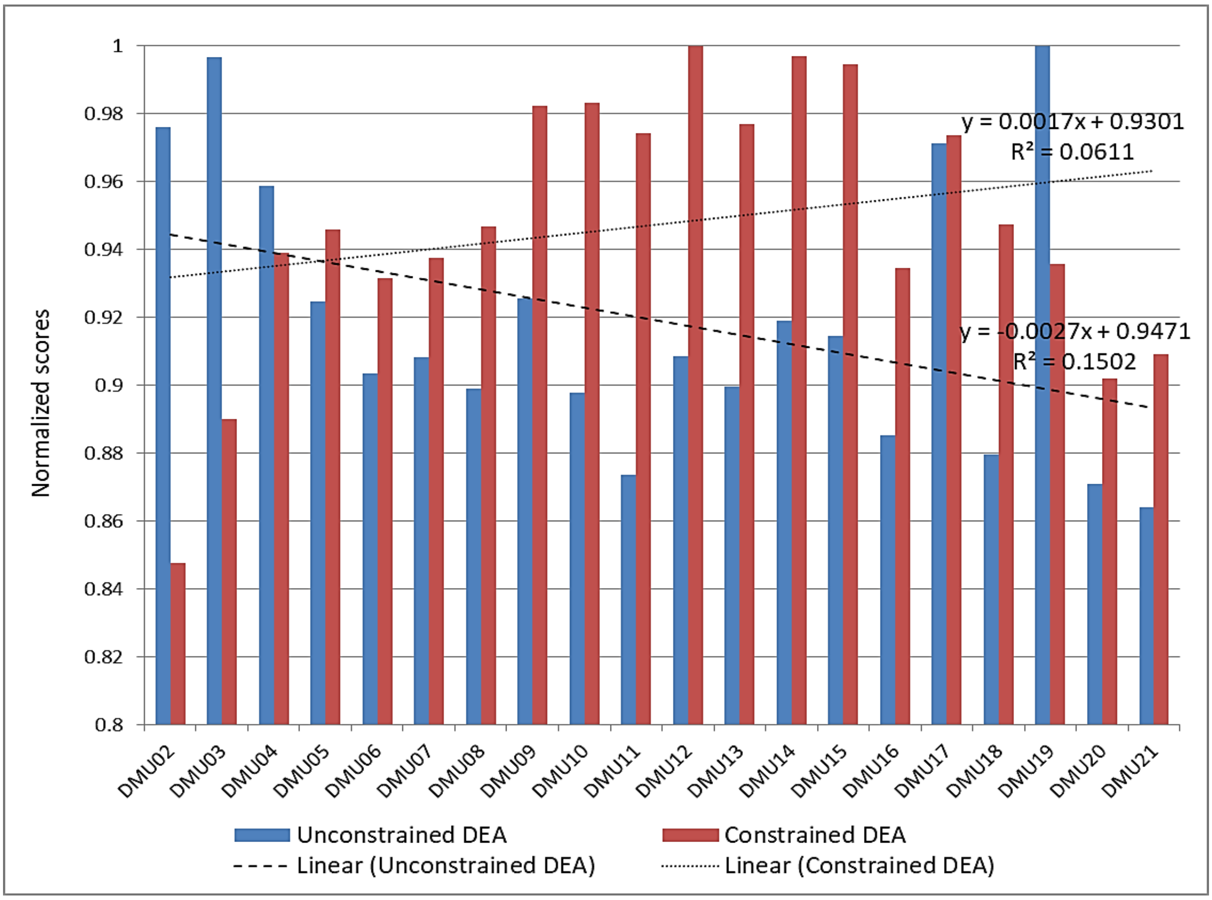

Figure 1. Bar-charts of normalized scores and simple trend regressions for constrained and unconstrained DEA outcomes.

Source: Own elaboration on the basis of research results.

The answer to the central question raised in the previous paragraph must be unequivocally affirmative! It suffices to use a simple "bare eye analysis" to draw this conclusion, which is further supported by formal correlation analysis. Here, both Pearson and Spearman correlation coefficients leave no doubt that the scores generated by the two versions of DEA have no correlation. The Pearson coefficient equals -0.26254390 (with a *t*-statistic of -1.15437 and a *p*-value of 0.2634), whereas the Spearman rank correlation coefficient is -0.04060150 (with a *z*-score of -0.176978 and a *p*-value of 0.8595). Thus, neither a linear nor a non-linear relationship links the obtained scores in any statistically significant manner.

Moreover, the state of affairs that emerges from the scores' estimates is qualitatively different. Many times, what might be regarded as a leading DMU/year in the unconstrained DEA appears to be quite the opposite in the constrained analysis. Even the overall tendencies that could be identified through trend analysis are quite different, showing a declining efficiency of the system in the unconstrained DEA, whereas a slightly increasing efficiency is observed in the constrained DEA. The variance of scores is clearly higher for the unconstrained DEA, which is due to the presence of extreme (and nonsensical) 0-1 weights assigned to numerous outputs and inputs. Nevertheless, no regularities in the score estimates can be identified in either variant, with peaks and troughs occurring in entirely different years. All in all, forming practical and political recommendations based on one variant or the other would lead to completely different consequences. As mentioned many times before, only the constrained DEA results should be considered reliable and binding.

Let us then take a look at the obtained results solely through the prism of constrained DEA. Erratic as the scores are, one can nevertheless identify at least three performance stages concerning the efficiency of the Polish district courts system. The first sub-period, encompassing the pre-accession (EU) years (2002-2003) and the early years after accession, is characterized by relatively low efficiency. The next sub-period consists of high-efficiency years (i.e., 2009-2015), clustered around the year 2012, for which the highest efficiency was estimated. Finally, starting from 2016 up to the final year under investigation (2021), there is a noticeable decline in the obtained efficiencies.

On one hand, it is evident that the estimated scores exhibit relatively low variability, with mean values of 0.958 and 0.947 for the original and normalized variants, respectively, and variances of 0.0390 and 0.0386. Therefore, one should not have expected spectacular increases in system efficiency across individual DMUs/years, even if their efficiency had reached the levels observed in the most efficient peer years[3]. On the other hand, it is important to recognize that the system's load in a given year is recursively dependent on the load observed in the preceding year (due to the accumulation of unsolved cases). Consequently, if the size of incoming cases were independent of the standing efficiency of the judiciary system, then even slight to moderate increases in efficiency in a given year would lead to increased efficiency in the following year, as a result of a reduced stream of caseload due to the decrease in pending cases from the previous year (see formula (5)), assuming the same involvement of discretionary inputs. If such improvements, even if minor, were maintained over a few consecutive years, the long-term effects could significantly enhance actual access to the courts of justice.

Although it might be methodologically quite risky to extend the analysis this far, one might conclude from the empirical outcomes of this investigation — ceteris paribus, accounting for various meso and micro idiosyncrasies that may characterize individual courthouses — that the macro-level institutional, procedural, organizational, and judicial setup was most conducive to the efficiency of the regional courts system in the years 2012, 2014, and 2015, with 2012 standing out as the indisputable leader.

Identifying the most historically efficient benchmark years is, at best, just a preliminary step toward more qualitatively oriented research, which would aim to answer the question of why the distinguished DMUs/years were more efficient than others. However, such an endeavour would require interdisciplinary expertise and the involvement of various stakeholders.

[3] Peer count analysis points to only three efficient DMU-s/years, being 2012, 2014 and 2015, respectively, of which year 2012 appears as a peer 18 times, year 2014 – 3 times, and year 2015 – 8 times.

This is important because DEA is a descriptive method that estimates relative efficiencies among homogeneous DMUs. While it addresses the actual variation in the effectiveness of individual DMUs within the specified period, it does not provide insight into the underlying reasons for such variation. Adequately estimating the efficiencies of a set of DMUs is one thing; adequately identifying — and possibly quantifying — the causes affecting those efficiencies at various levels is quite another. However, failing in the first step will inevitably lead to misconceptions regarding the second stage.

References

1. Agrell, P.J., Mattsson, P., Mansson J. (2019). Impacts on efficiency of merging the Swedish district courts. *Annals of Operations Research*, *288*, pp. 653-679. https://doi.org/10.1007/s10479-019-03304-0.
2. Cook, W.D., Tone, K., Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega*, *44*, pp. 1-4. https://doi.org/10.1016/j.omega.2013.09.004.
3. Deynelli, F. (2012). Analysis of relationship between efficiency of justice services and salaries of judges with two-stage DEA method. *European Journal of Law and Economics*, *34*, pp. 477-493. https://doi.org/10.1007/s10657-011-9258-3.
4. Ippoliti, R., Tria, G., (2020). Efficiency of judicial systems: model definitions and output estimation. *Journal of Applied Economics*, *23*(1), pp. 385-408. https://doi.org/10.1080/15140326.2020.1776977.
5. Kittelsen, S.A.C., Forsund, F.R. (1992). Efficiency analysis of Norwegian district courts. *Journal of Productivity Analysis*, *3*(3), pp. 277-306. https://doi.org/10.1007/BF00158357.
6. Lee, H.S., Chu, C.W., Zhu J. (2011). Super-efficiency DEA in the presence of infeasibility. *European Journal of Operational Research*, *212*(1), pp. 141-147. https://doi.org/10.1016/j.ejor.2011.01.022.
7. Mattsson, P., Tidana, C. (2019). Potential efficiency effects of merging the Swedish district courts. *Socio-Economic Planning Sciences*, *67*, pp. 58-68. https://doi.org/10.1016/j.seps.2018.09.002.
8. Panwar, A., Olfati, M., Pant, M., Snasel, V. (2022). A Review on the 40 Years of Existence of Data Envelopment Analysis Models: Historic Development and Current Trends. *Archives of Computational Methods in Engineering*, *29*, pp. 5397-5426. https://doi.org/10.1007/s11831-022-09770-3.
9. Pedraja-Chaparro F., Salinas-Jimenez J., Smith P. (1997). On the Role of Weight Restrictions in Data Envelopment Analysis. *Journal of Productivity Analysis*, *8*, pp. 215-230. doi: 10.1023/A:1007715912664

10. Peyrache, A., Zago, A. (2016). Large courts, small justice! The inefficiency of the optimal structure of the Italian justice sector. *Omega*, *64*, pp. 42-56. https://doi.org/10.1016/j.omega.2015.11.002.
11. Santos, S.P., Amado, C.A.F. (2014). On the need for reform of the Portuguese Judicial System – Does Data Envelopment Analysis assessment support it? *Omega – The International Journal of Management Science*, *47*, pp. 1-16. https://doi.org/10.1016/j.omega.2014.02.007.
12. Schneider, M.R. (2005). Judicial career incentives and court performance: An empirical study of the German labour courts of appeal. *European Journal of Law and Economics*, *20*(2), pp. 127-144. https://doi.org/10.1007/s10657-005-1733-2.
13. Siemaszko, A., Ostaszewski, P. (2013). Efektywność kosztowa sądownictwa powszechnego [Cost efficiency of the common courts]. *Instytut Wymiary Sprawiedliwości*, pp. 1-57, https://iws.gov.pl/wp-content/uploads/2018/08/Efektywno%C5%9B%C4%87-s%C4%85downictwa-powszechnego-oprac-12-1.pdf.
14. Tulkens, H. (1993). On FDH efficiency analysis: Some methodological issues and application to retail banking, courts and urban transit. *Journal of Productivity Analysis*, *4*(1/2), pp. 183-210. https://doi.org/10.1007/978-0-387-25534-7_17.
15. Voigt, S. (2016). Determinants of judicial efficiency: a survey. *European Journal of Law and Economics*, *42*, pp. 183-208. https://doi.org/10.1007/s10657-016-9531-6.