EKONOMISTA | **ARTYKUŁY**

2026, 3, 370–393
DOI: 10.52335/ekon/226201

Waldemar Florczak
Institute of Economics, Finance and Management, Jagiellonian University, Polska
https://orcid.org/0000-0003-3408-7784, waldemar.florczak@uj.edu.pl

How (not) to Apply Data Envelopment Analysis in the Judiciary: A Case Study of Efficiency in the Polish Regional Court System

Jak (nie) stosować metody DEA w sądownictwie: studium przypadku efektywności polskiego systemu sądownictwa okręgowego

Abstract

This paper aims to assess the technical efficiency of Poland's regional court system over a 20-year span in the twenty-first century by employing a constrained variant of Data Envelopment Analysis (DEA) applied to carefully processed time-series data. To estimate the technical efficiency of the system, various DEA models were utilised, beginning with the classical CCR DEA and super-efficiency DEA (SE DEA) approaches. However, significant shortcomings in the results produced by these methods were identified. To address these issues, a novel procedure involving constrained DEA was proposed, specifically designed to overcome these limitations and enhance the utility of DEA in evaluating judicial efficiency. Failure to account for the inherent limitations of classical or SE DEA models can lead to notable distortions in efficiency ratings assigned to individual Decision-Making Units (DMUs). Consequently, policy recommendations derived from such methods, intended to improve the actual performance of specific DMUs, may ultimately be unreliable or misleading.

Keywords: Efficiency, Data Envelopment Analysis, Weights Constraints, Regional Courts System.

JEL: C61, H00, H40, K00

Streszczenie

Celem artykułu jest ocena technicznej efektywności systemu sądów regionalnych w Polsce na przestrzeni 20 lat XXI wieku przy pomocy metody Analizy Obwiedni Danych (DEA) i z wykorzystaniem starannie przetworzonych szeregów czasowych. W celu oszacowania efektywności technicznej systemu zastosowano różne modele DEA, począwszy od klasycznej metody CCR DEA oraz podejścia DEA z super-efektywnością. Jednakże analiza wyników uzyskanych za pomocą tych metod ujawniła istotne niedoskonałości. W odpowiedzi na te ograniczenia zaproponowano nową procedurę opartą na ograniczonej DEA, która została opracowana w celu przezwyciężenia wskazanych mankamentów oraz zwiększenia użyteczności DEA w ocenie efektywności funkcjonowania wymiaru sprawiedliwości. Nieuwzględnienie inherentnych ograniczeń klasycznych modeli DEA lub metod DEA z super-efektywnością może prowadzić do znaczących zniekształceń w ocenie efektywności przypisywanej poszczególnym jednostkom decyzyjnym (DMU). W konsekwencji, rekomendacje polityczne wynikające z takich analiz, mające na celu poprawę rzeczywistej efektywności określonych jednostek decyzyjnych, mogą okazać się niewiarygodne lub nawet wprowadzać w błąd.

Słowa kluczowe: efektywność, Analiza Obwiedni Danych, ograniczenia wagowe, sądownictwo okręgowe.

JEL: C61, H00, H40, K00

1. Introduction

As an important component of the justice system, a well-functioning and independent universal judiciary has a major impact on the quality of social, public and economic life, but also forms the foundation of democracy, generating numerous externalities. Due to its importance to the state and society, the general judiciary should be an extremely efficient component of the public sector. In order to ensure that it has a high level of efficiency and introduce necessary reforms, it is necessary in doing so to perform a sound assessment of the prevailing trends, taking into account objective tools of quantitative analysis.

The purpose of this study is to estimate the technical efficiency of Poland's regional court system from 2002 to 2021, using the Data Envelopment Analysis (DEA) method—a key tool for evaluating efficiency in the public sector, including the judiciary (e.g., Voigt, 2016). Adopting a macro perspective enables a holistic assessment of the efficiency of systemic legal, administrative, procedural, organisational, and economic solutions that uniformly influenced all regional courts during the study period, shaping the functioning of the entire system.

In addition to systemic factors, court efficiency is also affected by factors at the meso level, which result from local differences in the external spatial environment of the courts, and at the micro level, related to the internal peculiarities of the operations of individual court units. Traditional studies of judicial efficiency using the DEA method, based on cross-sectional data, mainly focus on analysing the meso and micro levels. This is likely due to the lack of available data for long-term analysis. As the review articles by Voigt (2016) and Ipolliti (2020) point out, such studies typically rely on cross-sectional data. As a result, systemic factors in such analyses remain constant and are not even approximately quantified.

A complementary objective of this study is to highlight the inherent limitations of classical DEA, which raise legitimate concerns regarding the reliability and interpretability of its results. Simultaneously, this article introduces—for the first time in the context of judicial efficiency research—a methodological approach aimed at mitigating these shortcomings within the scope of existing modifications to classical DEA.

To estimate the technical efficiency of the system, various DEA models were utilised, beginning with the classical CCR DEA and super-efficiency DEA (SE DEA) approaches. However, significant shortcomings in the results produced by these methods were identified. To address these issues, a novel procedure involving constrained DEA was proposed, specifically designed to overcome these limitations and enhance the utility of DEA in evaluating judicial efficiency. Failure to account for the inherent limitations of classical or SE DEA models can lead to notable distortions in efficiency ratings assigned to individual Decision-Making Units (DMUs). Consequently, policy recommendations derived from such methods, intended to improve the actual performance of specific DMUs, may ultimately be unreliable or misleading.

The paper is structured as follows. Section 2 contains a concise description of the Polish regional court system. Next, the relevant literature is reviewed from the perspective of the paper's main research focus. Section 4 briefly presents the

methods employed in the study and provides a detailed discussion of the data used. Section 5 then presents the results in detail, with particular emphasis on specific idiosyncrasies related to the application of DEA in judicial efficiency analysis—an issue of critical importance that remains largely underexplored in the existing literature. Finally, the paper concludes with the main findings and policy-relevant recommendations, complemented by several additional remarks.

2. A brief overview of the Polish regional court system

Poland's common court system is one of the three pillars of power in the country, alongside the legislative and executive branches. Its activities are based on the provisions of the Constitution of the Republic of Poland and the Act on the System of Common Courts and other specific laws regulating court procedures. The main task of this system is to ensure justice, settle disputes, and enforce the law. It is based on the civil law paradigm and on the principles of judicial independence, independence of judges and the pursuit of justice. The judicial system is three-tiered, comprising district, regional and appellate courts, and its main areas of operation are civil, criminal, family, business, juvenile, and labour and social security cases (Ustawa z dnia 27 lipca 2001 r.; Sąd Okręgowy w Warszawie, 2013)

Regional courts are the second tier of Poland's general judicial system, and their activities focus on more complex cases and appeals against district court judgments. They function at the level of larger administrative units, such as counties, covering wider territories than district courts. Regional courts handle more legally complex and casuistic cases, both civil and criminal. Their jurisdiction includes, for example, crimes with high penalties, economic cases with high litigation value, or restructurings and bankruptcies of large companies. In addition, they are responsible for handling appeals against district court judgments, making them an important element in the appeals system. In district courts, judges tend to deal with a larger number of simpler cases, requiring them to have extensive experience in a wide range of law, but less often to specialise in narrow areas. In regional courts, judges often specialise in specific areas of law, such as commercial law, criminal business law or human rights cases. Because of the more complex nature of the cases, judges working in regional courts are expected to have a higher degree of legal knowledge and more professional experience.

In summary, regional courts differ from district courts in terms of their scope of operation, specialisation, level of cases handled, and their function in the appellate system. They have a more complex structure and more specialised staff, and play an important role in appeals and the handling of cases of special importance.

The functional competencies of regional courts include:

a) Civil cases: handling more complex cases, such as those with a high value of litigation, involving real estate, inheritance law, or complex contracts.
b) Criminal cases: handling cases of serious crimes such as homicide, economic crimes, and corruption, as well as adjudicating on appeals against district court judgments.

c) Family and juvenile cases: resolving more complex family cases, including those involving parental rights, adoption, as well as appeals in divorce and child custody cases.
d) Commercial cases: dealing with more complex commercial disputes, restructurings and bankruptcy proceedings of large business entities.
e) Labour and social security matters: handling appeals in labour cases and social security disputes.

Regional courts operate within designated areas of jurisdiction, which usually include several district courts, creating a more central system compared to lower-level courts. Each regional court consists of numerous divisions, and their number and type depend on the size of the court and its area of jurisdiction. The most common departments are:

a) Civil Division: handles civil cases with a higher value of litigation or a higher degree of complexity, such as real estate claims or property rights.
b) Criminal Division: handles criminal prosecutions for serious crimes, as well as hearing appeals against district court judgments.
c) Family Division: handles complex family cases, including child custody cases, adoptions, and appeals against judgments in family and custody cases.
d) Labour and Social Security Division: handles appeals against district court judgments on labour law and social security.
e) Commercial Division: handles more complex economic cases, including restructuring cases, bankruptcy cases, and larger disputes between entrepreneurs.

Despite their pivotal role in the Polish judicial system, regional courts also face challenges. Among the most significant are the growing number of appellate cases and the complexity of proceedings, while there has been no commensurate increase in the number of judges and court staff. Modernisations aimed at improving the efficiency of the judiciary, including the regional courts, do not always produce the desired results, especially when judged by indicators such as the average length of court proceedings.

The functioning of regional courts is monitored by the Ministry of Justice, which annually publishes statistics on the number of cases handled and the effectiveness of the courts. These indicators, despite their simplicity, do not always reflect real changes in the efficiency of the regional court system, which underscores the need for a deeper analysis of efficiency at the macro level. This is the purpose of this study.

3. Literature review and the positioning of the present research

Although there are various methods for estimating the performance of homogeneous entities engaged in market, social, or public activities—where the effects are not easily captured by direct market valuations (Smith & Street, 2005)—the DEA method and its modifications have dominated applications in this field. DEA stands out because of its key features, such as the capacity to account for multiple inputs and outputs common to all DMUs without requiring arbitrary assumptions about the production function. Furthermore, the optimisation techniques employed in DEA favour each of the DMUs being analysed, which makes the methodology

particularly advantageous from their standpoint. These characteristics have led to widespread DEA use in evaluating the efficiency of various entities in the public sector (Banker, 2014; Data Envelopment Analysis in Public Sector, 2014; Avikran, 2015), including within the judicial systems.

A comprehensive review of research on judicial efficiency can be found in works by Voigt (2016) and Ippoliti & Tria (2020). A literature search reveals that the vast majority of studies employ the classic version of DEA, using sufficiently large cross-sectional data at the micro-level (e.g., Ostrom & Hanson, 2003; Schneider, 2005), at the meso-level (for specific countries and selected judicial instances, such as Falavigna et al., 2015; Yeung, 2018; Gupta & Bolia, 2020), or even at the international level (e.g., Malcarne & Ramello, 2015; Voigt & El Bialy, 2016).

Only a few studies resort to the SE-DEA variant when the number of DMUs is low, resulting in little variation in performance indicators among the analysed units (El Bialy & Garcia-Rubio, 2011; Mattsson & Tidana, 2019). The performance-maximising DEA variant is typically chosen (with Schneider, 2005, being the exception), likely due to the increasing burdens on judicial systems in many countries. Furthermore, when non-discretionary variables such as caseload are included in the inputs, no additional modifications to the optimisation process are needed in the performance-maximising DEA variant (see Dyson et al., 2001; Cook et al., 2014). One study did use constrained SE-DEA (Santos & Amado, 2014) in response to the observation that average labour intensity of court cases varies systematically by subject matter, with consistent minority/majority relationships. However, even in that study, the authors did not present the optimal weights assigned to the inputs and outputs they considered.

Regarding the variables that define inputs and outputs in empirical studies, there is considerable variation. The most commonly used input variable is the number of judges, which is quite intuitive. However, exceptions exist in studies such as those by Lewin et al. (1982) and Tulkens (1993). The selection of other input variables largely depends on data availability. In addition to the number of judges, a variable reflecting the burden on the courts - namely caseload - is frequently included, appearing in almost all studies (e.g., Pedraja & Salinas, 1996; Schneider, 2005; Peyrache & Zago, 2016; Mattsson & Tidana, 2019).

Other input variables considered include the number of administrative staff (e.g., Kittelsen & Forsund, 1992; Pedraja & Salinas, 1996; Agrell et al., 2019), the availability of office and IT equipment and workspace (e.g., El Bialy & Garcia-Rubio, 2011; Yeung & Azevedo, 2011; Agrell et al., 2019), and expenditures on court operations (e.g., Deynelli, 2012). Finally, most applications of DEA in the judiciary rarely consider more than three input variables simultaneously, with the study by Peyrache & Zago (2016) being a notable exception.

Most of the empirical applications cited in this section, as well as in review papers such as Voigt (2016) and Ippoliti & Tria (2020), typically consider only the total number of cases resolved in a given year as the output measure. This approach raises concerns, as the duration of proceedings and resource/labour expenditures can vary significantly depending on the type of case. Treating all cases that come to court as uniform regarding the resources required is a considerable simplification that may lead to erroneous conclusions about the actual efficiency of the courts.

This issue prompts the question of whether the efficiency of individual court units can be accurately assessed based solely on such highly aggregated data, which encompasses all types of legal cases. Notably, there are exceptions in the literature in which there was a more detailed breakdown of cases by legal subject. For instance, Kittelsen & Forsund (1992) categorised cases into seven groups, Tulkens (1993) divided them into four categories, and both Mattsson & Tidana (2019) and Angrell et al. (2019) each classified cases into three categories. Additionally, Santos & Amado (2014) provided an extensive breakdown into 43 categories of cases.

To the best of the authors' knowledge, with the exception of one study by Santos & Amado (2014), no other research has addressed the evaluation of the substantive weights assigned to individual inputs and outputs in the context of court efficiency. Even in the aforementioned study (Santos & Amado, 2014), this assessment was not fully executed. The lack of attention to this aspect in DEA-based studies of judicial efficiency raises questions about the substantive reliability of the results obtained.

This concern arises because both the classical DEA and its super-efficiency variant produce efficiency scores ranging from zero to one. The characteristics of this method imply that among units deemed efficient (i.e., with a score of one), there may be those whose efficiency is attributable to only selected inputs or outputs—potentially even to just one. This property of the method casts significant doubt on the interpretation of results derived from classical DEA analyses.

This study proposes practical solutions to this problem by applying a modified version of DEA with constraints, involving the introduction of substantive minimum and maximum values for the weights, as well as minority/majority relationships between the weights. The work draws on a rich international body of research in the field under study while bringing several important innovations. First, it covers a long period of twenty years and focuses on a single subject—the regional court system both in its first-instance and second-instance roles. As a result, the conclusions of the study can be interpreted in the context of systemic efficiency rather than individual courts. Secondly, the analysis of the obtained weights in the classic version of DEA, leading to their substantive adjustment, allows a fully correct interpretation of the results regarding the efficiency gradation of the regional judiciary in Poland over time. Thirdly, to indicate the sensitivity of the results obtained through DEA, particularly in the context of estimates of the weights assigned to inputs and outputs, the results will be presented in a variant and sequential manner. All the aforementioned features, along with their pragmatic qualities, highlight the originality of this study within the context of the application of DEA methods to analyse judicial efficiency.

4. Methods and data

Efficiency is a key concept in economics, and is the subject of special attention among economists. As some authors note, the pursuit of maximum utilisation of scarce resources is the foundation of this discipline of science. For example, according to Krajewski and Milewski (2018, p. 11), economics […] *shows how people use scarce resources, how they use them to conduct economic activities, how they allocate*

them between different uses that compete with each other, and what guides them in making such choices. It also shows whether the use of scarce resources is efficient or not and analyses the factors on which this depends […].

The measurement of efficiency in public institutions relies mainly on parametric and non-parametric frontier methods. Parametric approaches, most notably Stochastic Frontier Analysis (SFA), specify a functional form for the production or cost frontier and decompose deviations from it into inefficiency and random noise. While SFA allows statistical inference and accounts explicitly for measurement error, its results are sensitive to distributional assumptions, functional form misspecification, and the need to predefine a single output–input technology. An example of this approach applied to the Polish justice system is the study by Bełdowski et al. (2020), while the work by Antonucci et al. (2014) illustrates its use in an international context.

In contrast, DEA is a non-parametric, linear-programming-based method that constructs an empirical efficiency frontier from observed data without imposing a functional form. DEA is particularly advantageous in the public sector, where technologies are complex, outputs are multidimensional, and prices are often unavailable. It accommodates multiple inputs and outputs measured in different units and provides transparent benchmarking information. Although DEA is deterministic and sensitive to outliers, its flexibility and minimal *a priori* assumptions make it especially suitable for evaluating heterogeneous public institutions, where SFA structural assumptions are often difficult to justify.

Measures to improve the efficiency of the functioning of individual DMUs, once the level of their inefficiency has been determined and the most efficient patterns identified, offer the possibility of increasing the efficiency of public institutions—including the judiciary—at minimal cost, and sometimes even without incurring additional expenses. The prerequisite for the success of such measures, however, is a detailed diagnosis of the current state, taking into account the most important quantitative trends over time.

Two-dimensional efficiency and productivity indicators - such as those published on the Ministry of Justice's websites (https://isws.ms.gov.pl/pl/baza-statystyczna/publikacje/ under the tab: Mierniki wymiaru sprawiedliwości (*Judicial System Indicators*), e.g. clearance rates, overall backlog rates, number of cases disposed of per judge or disposition time (see also Florczak, 2025b)—do not allow a reliable assessment of the functional efficiency of the system of common courts by different instances. This is because these indicators only partially reflect systemic efficiency. Even those indicators that refer to the *final results* of the system's operation—such as the indicator of the average duration of court proceedings, often discussed in the context of the functioning of the judiciary—do not provide the possibility of an objective and complete assessment of efficiency, since they do not take into account the relationship between outputs and inputs. Indeed, changes in these indicators can signal both an increase and a decrease in efficiency, but equally changes in the degree of strain on the system.

Therefore, in order to obtain a reliable assessment of the technical efficiency of institutions performing certain public utility functions, more sophisticated analytical methods are required, which (a) take into account the diversity of functions

and tasks carried out by such institutions, even when the effects of their activities cannot be quantified in monetary terms, (b) take into account the main inputs that are necessary to carry out the tasks entrusted to them, (c) do not assume knowledge of the form of the production function (the selection of which is always more or less arbitrary), according to which inputs are converted into outputs, (d) make it possible to reach an objective estimate of the efficiency of a given institutional unit against other similar units included in the research sample.

The DEA method meets all of these conditions, and can be used to assess the relative efficiency of homogeneous DMUs. A compact formal review of this method and its modifications, which are later used in the empirical investigations in this article, is presented below. The article then describes the reasons for this selection, with references to relevant literature.

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 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* DMU (efficiency of the regional court system in each year of the 2002-2021 period);
$z = 1, 2 \ldots Z$ – index of the *z-th* DMU (total Z =20, i.e. number of years);
X_{sz} – the *s-th* type of input in the *z-th* DMU (the amount of the *s-th* input - out of the three considered in the study: the number of judges, outlays on the regional court and the caseload - in a given year of analysis);
$s = 1, 2 \ldots S$ – index of the *s-th* input (total of S=3 types of inputs, listed above, common to all DMUs);
Y_{tz} – *t-th* type of output in the *z-th* DMU (amount of the *t-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);
$t = 1, 2 \ldots T$ – index of the *t-th* input (total of T=6 output types, listed above, common to all DMUs;
w_{tz}, v_{sz}, – weights assigned to the *t-th* outcome and *s-th* input in the *z-th* DMU, respectively.

A key property of the DEA method is that the weights appearing in the equation 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 objectivised in a way that ensures the most favourable set from the perspective of individual DMUs. This is achieved by maximising outputs and minimising 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:

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

where "i" denotes all DMUs in the sample

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

The above problem can be reduced to an equivalent linear programming problem, which determines the optimal weights. The optimisation is conducted separately for each DMU included in the sample:

$$max\, E_z = \sum_{t=1}^{T} w_{tz} Y_{tz} \tag{5}$$

with respect to the following constraints:

$$\sum_{t=1}^{T} w_{tz} Y_{ti} - \sum_{s=1}^{S} v_{sz} X_{si} \leq 0 \tag{6}$$

$$\sum_{s=1}^{S} v_{sz} X_{sz} = 1 \tag{7}$$

$$w_{tz}, v_{sz} \geq 0 \tag{8}$$

In the classical version of DEA, however, there are certain problematic assumptions and undesirable properties which hinder transparent interpretation of the results and can negatively affect their quality—especially when disregarded, as is often the case in studies applying DEA to the judicial sector (see Dyson et al., 2001). In particular, classical DEA exhibits a low level of discrimination of efficiency scores between the analysed DMUs when the number of DMUs is low. This issue frequently arises in studies on judicial efficiency, and researchers typically address it by employing a variant of DEA that utilises the concept of super efficiency (SE-DEA model). A concise representation of this extension of the DEA-CRS (Constant Returns to Scale) model is shown below (Lee et al., 2011):

$$max\, E_z = \sum_{t=1}^{T} w_{tz} Y_{tz} \tag{9}$$

with respect to the following constraints:

$$\sum_{t=1}^{T} w_{tz} Y_{ti} - \sum_{s=1}^{S} v_{sz} X_{si} \leq 0, \forall i \neq z \tag{10}$$

$$\sum_{s=1}^{S} v_{sz} X_{sz} = 1 \tag{11}$$

$$w_{tz}, v_{sz} \geq 0 \tag{12}$$

The difference between the classical DEA-CRS and its super-efficiency extension consists in excluding the *z*-th DMU from constraint (6) and using constraint (10) instead (see equations (6) and (10)). As a result, the efficiency scores of DMUs that were previously considered efficient in the classical DEA-CRS model (with scores of 1) are no longer capped at 1, allowing more meaningful discrimination among these DMUs.

However, neither the 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 constraints (8) and (12)), 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. This version of DEA, though never 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.

Generally, there are three distinct types of weights in constrained DEA analysis: absolute, price, and virtual weights. In this research, the latter two were used, resulting in the following presentation of the constrained SE-DEA model:

$$max\, E_z = \sum_{t=1}^{T} w_{tz} Y_{tz} \tag{13}$$

with respect to the following constraints:

$$\sum_{t=1, z\neq a}^{T} w_{ti} Y_{ti} - \sum_{s=1}^{S} v_{si} X_{si} \leq 0, \forall i \tag{14}$$

$$\sum_{s=1}^{S} v_{sz} X_{sz} = 1 \tag{15}$$

$$w_{tz}, v_{sz} \geq 0 \tag{16}$$

$$v_{sd} \leq v_{sz} \leq v_{se} \tag{17}$$

$$w_{tb} \leq w_{tz} \leq w_{tc} \tag{18}$$

$$f \leq w_{tz} Y_{tz} \leq g \tag{19}$$

The additional constraints (17)-(19) introduce the following modifications to the unconstrained weights in the classical SE-DEA:

(i) Subscripts *b* and *c* denote DMUs whose input weights cannot be greater or lower than those of the referenced *z-th* DMU, respectively (these are referred to as input *price weights*).

(ii) Subscripts *d* and *e* denote DMUs whose output weights cannot be greater or lower than those of the referenced *z-th* DMU, respectively (also referred to as output *price weights*).

(iii) Constants *f* and *g* put the lower and the upper limits upon the contribution/share of the *z-th* output in the production of total outputs (referred to as *virtual weights*).

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 commercial cases) in 2011–2021 was 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 *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 also 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:

$$Caseload_t = Pending\ cases_{t-1} + Incoming\ cases_t \tag{20}$$

$$Pending\ cases_t = Caseload_t - Resolved\ cases_t \tag{21}$$

$$Case\ processing\ time_t = \frac{Pending\ cases_t}{Resolved\ cases_t} \cdot 12 \tag{22}$$

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 court activity. The selection of such variables ensures that the conditions of homogeneity and completeness are met for the DMUs under study (here: individual years of regional court 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 equation (20). On the other hand, data on the average number of judges employed in the regional courts from 2006 to 2021 was 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 on 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 regional 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 regional courts to total expenditures on the general judiciary. Third, budget expenditures for regional courts were estimated in current prices by multiplying total expenditures on the general judiciary by the established fraction allocated to the regional 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.

Table A1 displays the values for all inputs and outputs used in the DEA analysis, along with the number of cases resolved and the average annual growth rates of these variables. Table A2presents the Pearson linear correlation coefficients between these variables, demonstrating minimal collinearity among the input variables. This indicates that each variable has the potential to independently influence performance, as supported by both theoretical and empirical evidence (see the literature review).

Table A3 shows the variations in the average time taken to process cases by area of law, computed using equation (22). There is a systematic annual increase in processing times (with the exception of criminal cases, which display substantial variability), even though the number of cases remains relatively constant. A key observation relevant to the analyses presented in this article is the regular gradation of average case resolution times by subject of law. Regardless of the year chosen, criminal cases took the shortest time, while family law or insurance cases took the longest. Additionally, with a few exceptions, the average duration of litigation can be hierarchically categorised by type of law (in ascending order): *CRIME* < *LABOR* < *COMMERCIAL* < *CIVIL* < *FAMILY* (*INSURANCE*). Regarding family and insurance cases, no consistent majority/minority relationship can be established, as the average litigation duration fluctuates for each type each year.

Continuing the analysis based on estimates of the average time spent resolving cases by area of law, we can determine the fractions of total time allocated to each case type over the study period. Table A4 displays the workload by area of law, enhancing our understanding of how different case types impact regional courts in Poland. These fractions are associated with the concept of virtual weights used in DEA analysis. By imposing restrictions on parameters that control the proportionality of inputs to outputs, we enhance the substantive consistency of the efficiency analysis. The following equation was used to calculate these values:

$$s_i = \frac{d_i \cdot n_i}{\sum_{j=1}^{6} d_i \cdot n_i} \tag{23}$$

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 (equation (22)),

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

The high variability of proportions across different areas of law reflects changes in both the average length of cases and their numbers over time. Before proceeding to the main focus of the article, it is important to emphasise that both the estimates of the average time taken to resolve cases by subject of law and their fractions of the total court work time define the actual, historical, and empirical ranges of variation for these characteristics over the analysed period. This observation is crucial and will be further developed and operationalised empirically to provide a clear interpretation of the results obtained through DEA methods in the context of the efficiency of the Polish regional court system.

5. Research results

This section adopts variants of DEA with fixed scale effects and a focus on maximising outputs. First, studies that consider the potential impact of the size of judicial units on their efficiency typically refer to cross-sectional analyses in which DMUs vary in size. However, this paper focuses on the entire regional court system in Poland over a twenty-year period, rather than on individual courts. Second, the number of regional courts during the study period exhibited minimal variability and was practically constant. Third, introducing a non-discretionary variable (caseload) allows the neutralisation of any economies of scale that might arise from the increasing number of cases handled by the courts. This approach technically enables the treatment of each year studied as a separate DMU in the analysis.

The motivation for the choice of the performance-maximising option was the theoretical and empirical reasons outlined in the literature review. This option is primarily justified by the high level of rigidity in the supply of judges and the constraints imposed by budgetary laws on the judiciary's financing. The court system lacks flexible resources; it cannot quickly increase the number of judges or adapt funding. Therefore, maximising work output is essential in light of rising workloads. Additionally, considering the trend of increasing average time taken to resolve court cases observed in most EU countries, including Poland (Study on the Functioning of Judicial Systems in the EU Member States, 2022), efforts to improve the judicial system should prioritise enhancing outputs rather than minimising inputs.

Table A5 shows the performance indicators obtained using the classic radial CCR method of DEA maximising outputs. Several conclusions can be drawn from table A5. First, each discretionary input—the number of judges and the budget expenditures for the functioning of the regional courts in Poland—provides qualitatively different information regarding obtained outcomes. The efficiency values obtained separately for these two inputs differ considerably by year. Consequently, both inputs should be included in the analysis simultaneously.

Second, the caseload variable is clearly a non-discretionary input, and thus analysing the scores it might generate in a separate DEA model with a single input serves no purpose. However, it is essential to include it in the DEA model estimating the efficiency of the court system, as already explained above.

Third, including more inputs while keeping the number of outputs constant leads to a decrease in discriminating power. For example, the number of efficient DMUs is three and seven in the one-input models, seven, fourteen, and fifteen in the two-input models, and seventeen in the three-input model (out of a total of twenty DMUs). This is a well-known feature of DEA: the number of efficient DMUs typically fluctuates around the product of the inputs and outputs used in the analysis. This characteristic often goes unnoticed when the number of analysed DMUs is in the tens or even hundreds, as is commonly the case with cross-sectional micro-unit data.

Fourth, all the input variables are undeniably important for both theoretical and empirical reasons. Therefore, the DEA model utilising all three inputs simultaneously should be considered the primary or "binding" variant. However, this variant is the least discriminating, not due to substantive issues but rather for technical reasons (as discussed in the previous paragraph).

Under such circumstances, one way to enhance the discriminatory power of the DEA model is to apply the concept of *super-efficiency*. In short, this method assigns a "new" efficiency score to each efficient DMU (in the classical DEA model) when that particular DMU is excluded from the analysis. By removing the efficient DMU, we allow for a recalculated efficiency score that can exceed the value of one. This approach introduces greater variation among the previously efficient DMUs, distinguishing between them. Afterward, rescaling these scores back to "one" for the highest score ensures that the results can be interpreted in the "traditional" way.

Table A6 presents the efficiency scores obtained through DEA using the super-efficiency procedure, along with the corresponding rescaled scores. At first glance, the results seem to be more than satisfactory. Despite the very low number of DMUs, it seems that their efficiency scores could be discriminated well, so that the resulting empirical set of efficiency scores is quite diversified and thus illusively more realistic. Actually and virtually, this is at best the direction chosen by all researchers investigating the efficiency of courts by means of DEA (with a large majority of them stopping at the stage of classical DEA). Moreover, if one were to exclude the first year of the analysis, which seems to be a clear outlier, and repeat the calculations from scratch, one would arrive at even more "reliable" results in terms of a lower variance in the efficiency scores thus obtained.

At this stage, one might be tempted to proceed directly to the interpretation of the results. However, such an approach would raise serious methodological concerns. In particular, interpreting efficiency scores without a careful examination of the underlying parameters and weights may lead to conclusions that lack empirical and economic validity. Despite the fact that this practice is widespread in the applied DEA literature, it substantially limits the reliability and interpretability of the resulting efficiency assessments. In truth, it appears that few are concerned with such "trivial matters" as the actual estimates of parameters produced by the DEA model. The problem is that these details are far from trivial; they are, in fact, vital if one is to have confidence in the resulting efficiency scores. The fact that almost none of the publications cited in this paper, save for one (Santos & Amado, 2014), pays explicit (though not exhaustive) attention to parameter estimates or weights

illustrates the degree of both intentional and unintentional neglect surrounding this issue.

Examining the weights or parameters assigned to outputs in the super-efficiency variant of DEA, incorporating all three inputs, both in their absolute and virtual values (for want of space only absolute weights are reported in table A7), reveals a notable pattern (this issue is also prominently evident in all the other analysed variants). It is evident that many weights take on a value of '"zero", and in some cases, all but one weights are zero. This implies that only one factor—either on the input or output side—appears to contribute to the overall efficiency. Interpreting these weights economically leads to logical, yet highly unrealistic, conclusions: factors with zero weights make no contribution, or, in an equally implausible scenario, only one factor influences the overall efficiency of a given DMU.

Such conclusions, though inevitable, undermine the fundamental credibility of DEA as a reliable tool for estimating efficiency. The rationale behind DEA assumes that all inputs and outputs should contribute meaningfully—especially with outputs representing a comprehensive set of DMU activities. If they do not do so, they should not have been included in the analysis in the first place. This issue is even more striking considering that the individual DMUs in this analysis represent specific years from 2002 to 2021. It defies logic to suggest that in one year, a particular output consumes all available inputs, while in another year, that same output uses none (e.g., see the weights assigned to output *LABOUR* in 2004 compared to neighbouring years or *INSURANCE* in 2002 and the following three years, table A7). Similarly, the notion that in one year an input accounts for all outputs, and in another year it contributes nothing (e.g., input *JUDGES* in 2019 vs. 2016 in table A7) is equally nonsensical.

These outcomes substantially limit the explanatory value of the unconstrained DEA results. If the estimated weights lead to economically implausible or substantively uninterpretable conclusions, the resulting efficiency scores cannot be treated as fully reliable. A necessary methodological response is therefore to impose substantively justified restrictions on the weights, so that the estimated parameters remain consistent with the empirical and institutional characteristics of the analysed judicial system. This can be achieved through the use of constrained DEA, where the optimisation of efficiency scores is performed with the incorporation of specific, substantively grounded, and logically and empirically sound restrictions placed on the weights during the optimisation process.

It seems relatively straightforward to assign weights to a resolved case (by subject of law) based on the time spent resolving it. After all, *time is money*, especially when the actual costs of resolving an average case cannot be credibly calculated. Additionally, by applying the concept of super efficiency and placing constraints on both absolute and virtual weights—on both the output and input sides—one can obtain seemingly sound results for evaluating individual DMUs, avoiding anomalies such as values of zero to one. The entire procedure is described and implemented step by step in the following sections.

We begin with a proposal by Santos & Amado (2014), who, based on actual data, established appropriate minority/majority relations between output weights assigned

to different cases by subject of law and in line with the average time each case absorbs, as shown in equation (22) and summarised in table A3. Consequently, this approach results in the following set of relationships for assigning price weights to various cases by subject of law (see considerations in the previous section; the data is formatted for easy readability in the *pydea* software; the names of the weights are self-explanatory, with the prefix *A*(?) indicating absolute weights: this is equivalent to applying the previously mentioned constraints to the respective in equation (1)):

ALABOUR / *ACRIME* >= 1
ACOMMERCE / *ALABOUR* >= 1
ACIVIL / *ACOMMERCE* >= 1
AFAMILY / *ACIVIL* >= 1
AINSURANCE / *ACIVIL* >= 1

Although novel compared to all previous work in the field, this approach does not appear to address the critical issue of inadmissibility of output weights, which – in the study under consideration - turned out to exhibit the same uninterpretable properties as discussed before (outcomes not reported here). It is only when additional constraints are applied—specifically on the virtual weights, which have a clear interpretation as they represent the proportion of total time spent on resolving cases by particular legal subjects—that the resulting output weights begin to make sense (these constraints are explicitly outlined by equation (23)).

Similar to the minority/majority relationship concerning absolute weights, realistic ranges for the variability of virtual weights assigned to different case types, categorised by subject of law, must also be determined based on logical criteria. The most empirically grounded—and therefore most realistic—values for these categories are presented in table A4. Consequently, the actual ranges of variability for the virtual weights for each legal subject type are as follows (prefix *V*(?) indicating virtual weights):

VFAMILY >= 0.119	*VFAMILY* <= 0.220
VCOMMERCE >= 0.027	*VCOMMERCE* <= 0.092
VCRIME >= 0.097	*VCRIME* <= 0.188
VLABOUR >= 0.009	*VLABOUR* <= 0.037
VINSURANCE >= 0.151	*VINSURANCE* <= 0.379
VCIVIL >= 0.272	*VCIVIL* <= 0.537

Utilising a DEA variant with simultaneous constraints on both price and virtual weights ultimately results in correct output weights (not reported here). However, the issue previously noted regarding input weights still persists. This outcome calls into question the methodological acceptability of the obtained results. Before we continue, let us consider two rhetorical questions that do not diminish the critical role of judges in resolving court cases:

1. Would judges, entrusted with deciding all pending cases, willingly carry out their duties without compensation?
2. Would the institution of courts or judges even be necessary if no cases were brought before them?

The answer to both questions is clearly negative, reaffirming that the criticisms and doubts raised about the admissibility of output weights in the classic/

unconstrained DEA model—and the validity of the rankings it generates—apply equally to input weights. Thus, ensuring empirical realism in the input weights is just as essential for the reliability of DEA-generated efficiency scores.

Unlike output variables, where both price and virtual weights can be estimated based on actual time spent, applying a similar approach to inputs is not feasible. Instead, the proposal relies on the undeniable observation that judges play a direct and central role in the case disposition process. However, as highlighted by the two questions raised above, judges only act when their activity is complemented by the presence of two additional inputs: caseload, which is a non-discretionary input, and financial outlays. This observation leads to a heuristic suggestion that leverages this insight, allowing the determination of realistic, empirical price weights between the inputs used, with judges serving as the key input.

The initial step involves establishing the empirical range of variability for the outlays-to-judges and caseload-to-judges ratios with their respective extreme values for the period under analysis (2002–2021) as follows:

min (*CASELOAD/JUDGES*) = 0.429
max (*CASELOAD/JUDGES*) = 0.612
min (*OUTLAYS/JUDGES*) = 0.544
max (*OUTLAYS/JUDGES*) = 1.001

Consequently, based on the assumption that judges play the primary role in the entire case disposition process, it follows that the price weights attributed to the other two inputs should fall within the intervals established by the previously mentioned extreme points, with the absolute weights assigned to the input "judges" serving as the reference category. Both excessively high and excessively low values of input weights lead to economically implausible interpretations and undermine the empirical credibility of the estimated efficiency scores. If the weights are too high, this would suggest that the relative importance (or price) of the other inputs surpasses that of judges. If the weights are too low, this would undervalue the importance (or price) of those inputs from an empirical standpoint. Therefore, the constraints applied to the price weights of all inputs are as follows:

ACASELOAD / *AJUDGES* >= 0.429
ACASELOAD / *AJUDGE* <= 0.612
AOUTLAYS / *AJUDGES* >= 0.544
AOUTLAYS / *AJUDGES* <= 1.001

Only after imposing all the necessary constraints on the price and virtual weights for both outputs and inputs can we arrive at final results that are devoid of the previously identified critical and detrimental properties. At this stage, no weight—whether associated with outputs or inputs—takes on the value of "zero" or "one (as shown in table A8: for want of space, this time only virtual weights are reported). Furthermore, due to the manner in which the constraints were applied, the variability of the estimated weights inherently aligns with the historical minimum and maximum intervals observed during the analysed period. Thus the final results produced in this empirical study are entirely reliable and interpretable, facilitating a robust evaluation of the actual performance of the Polish regional court system over a span of twenty years.

6. Conclusions and final remarks

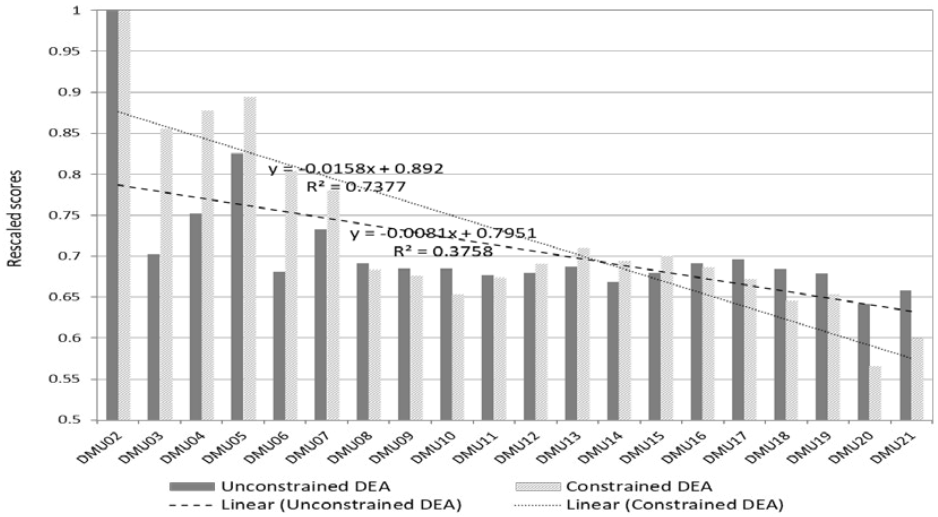
A pertinent question arising from the DEA analysis thus far—one that has significant practical and political implications—is whether there are any actual differences in the efficiency scores derived using the super-efficiency concept without paying attention to the obtained weights, compared to those scores calculated with a focus on obtaining logical weights. From a methodological standpoint, only the latter results can be considered sufficiently robust for reliable assessments of judicial system efficiency. Failing this, the scores produced in the classical, unconstrained DEA (which operate under fundamentally unrealistic assumptions regarding weights) could lead to highly misleading conclusions, as demonstrated by the findings from this research.

To address the question posed, a comparison was conducted between the variants deemed most appropriate regarding input specification, specifically those incorporating all three inputs, and their efficiency values presented in the final column of table A6 and the third column of table A8). Both formal analyses utilising various statistical tests and measures (as shown in table A9) and a descriptive interpretation of the content in figure 1 will yield conclusions. Should any patterns emerge between the substantially correct variant (constrained DEA) and the highly questionable one (unconstrained DEA), it may be possible to theorize or hypothesize that the rankings of DMUs generated by the unconstrained SE-DEA could be adjusted to produce more reliable assessments.

It appears that the overall trend in efficiency changes over time is consistent regardless of the method employed, revealing a significant decline in the efficiency of the regional court system from 2002 to 2021. This is arguably the most crucial general conclusion applicable to the two options analysed. Indeed, considering the results of the conducted tests—summarised in table A9—it is challenging to identify any other notable patterns. The only consistent finding, as indicated by Pearson's linear correlation coefficients, is that the direction of changes in the efficiency ratios (whether increasing or decreasing in individual years) remains the same for both variants. Taken together, the following suggest that the considered variants do not generate predictable or persistent discrepancies in year-specific efficiency-score estimates:

a) the absence of statistically significant differences between the mean efficiency scores obtained using the two approaches;
b) the statistically significant differences in their variances across the analysed sub-periods;
c) the lack of evidence of a non-linear relationship;
d) the weak linear correlation between the two sets of estimates. This observation is also supported by a visual inspection of figure 1.

Figure 1.
Bar-charts of rescaled efficiency scores for unconstrained and constrained DEA outcomes



Source: Own work on the basis of research results.

Moreover, the non-parametric Spearman's rank correlation test indicates that the rank order of the efficiency score estimates for each DMU differs statistically between the two approaches. Admittedly, such conclusions can only be drawn once the whole sample has been divided into two equal sub-samples, but the lack of stability of the conclusions over time undermines their factual validity. This finding is further supported by graphical analysis (figure 1), which highlights the absence of a consistent relationship between the efficiency scores produced by the two methods. Consequently, it is impossible to make even a rough inference about the actual (or substantively correct) status of cases regarding the technical efficiency of the judiciary based on the results derived from the unconstrained SE-DEA. Therefore, any further conclusions should solely rely on the outcomes obtained from the constrained SE-DEA variants.

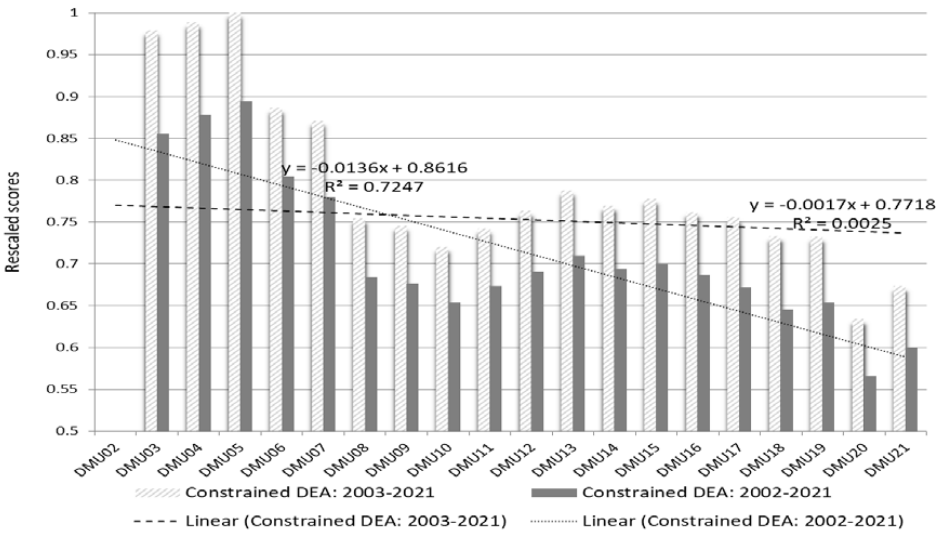
In all analyses conducted so far, the efficiency of the system was found to be highest in the first year of the sample, raising questions about the homogeneity of the year 2002 in relation to the other years. This issue pertains not merely to the formal statistical records of judicial functioning in that year but also to the legal, legislative, and institutional conditions in place at the time, which may have significantly differed from those that were implemented in subsequent years.

This may be because the operation of the Polish legal and judicial system at the beginning of the twenty-first century was significantly shaped by Poland's accession to the European Union in May 2004. This event initiated a series of changes aimed at harmonising legislation, enhancing the independence of the judiciary, fostering international cooperation, protecting human rights, and modernising the training

system for judges and prosecutors. This led to higher standards governing the judiciary in Poland, aligning the Polish judicial system more closely with the standards observed in other EU member states. The adaptation of existing legal, procedural, and institutional frameworks occurred gradually, beginning several years prior to formal accession. As such, it is plausible that 2002 represented a transitional period, with the system in place at that time not yet having "matured" into the solutions that would define the subsequent years. This context may help explain the atypical results obtained for that year. Without delving into a theoretical exploration of this uncertainty, one could address it through a straightforward methodological procedure using the same analytical tool, namely SE-DEA, while excluding the year 2002 from the analysis. This approach allows evaluation of the extent to which the inclusion of the year 2002 influences the efficiency score estimates throughout the entire sample.

Figure 2.

Bar-charts of rescaled scores and simple linear trend regressions for constrained DEA outcomes spanned on full (2002–2021) and abbreviated (2003–2021) samples



Source: Own work on the basis of research results

To achieve this, it was essential not only to restrict the sample to the years 2003–2021, but also to make slight adjustments to the previously established price weights for the inputs (not reported here), as certain threshold values have changed. Figure 2 illustrates the progression of these scores for the primary variants that incorporate all three inputs. It seems that excluding the year 2002 from the sample offers a clearer perspective on the dynamics and levels of efficiency/inefficiency within the regional court system in Poland. First, the efficiency of individual DMUs is higher in the analyses that omit 2002, with differences ranging from ten percent

to thirty percent, depending on the specific variant. Second, the rate of decline in system efficiency is more gradual in the versions that exclude 2002. Third, the ranking order based on efficiency scores across the years remains nearly identical in the corresponding variants, demonstrating consistency regardless of the approach used when 2002 is left out. In summary, removing 2002 from the analysis yields a slightly less harsh evaluation of the Polish regional court system's efficiency. However, this evaluation still remains quite critical.

Based on the results obtained in this study, the Polish regional court system appears to be experiencing a persistent and substantial decline in technical efficiency over the analysed period. Despite efforts to synchronise the Polish justice system with European Union legislation, which initially led to a relatively high level of systemic efficiency around the time of the country's accession to the EU, there was a consistent and troubling decline in efficiency in the years that followed. This downward trend persisted irrespective of the political formations in power and the various reform attempts made to address the issues within the system.

The evidence suggests that the reforms and legislative changes, although aimed at enhancing efficiency and aligning with EU standards, have not resulted in the desired improvements. Instead, the ongoing deterioration reflects deeper, systemic problems that remain unaddressed. These findings indicate the need for further, in-depth institutional and policy-oriented research aimed at identifying the structural determinants of the observed inefficiencies.

The inefficiency of the Polish regional judiciary is particularly troubling, not only due to the critical role that the judiciary plays in ensuring the effective functioning of society and the economy, as highlighted in the introduction of this study, but also for at least three additional reasons. Firstly, from an economic standpoint—this study's primary perspective—it is difficult to justify the current situation. The number of incoming cases has remained almost constant, while the number of employed judges has shown a slight upward trend, and the real budgetary expenditure on the regional courts has more than doubled (see table A1). Therefore, it is hard to argue that the system is overburdened. From this economic perspective, it is difficult to accept the fact that court proceedings are taking longer. One possible interpretation is that the observed increase in case duration has not been accompanied by proportional improvements in measurable system outputs. However, this raises the question of whether such "professional rationality" aligns with the principles of economic efficiency and the broader societal interest at a macro level. This situation prompts a deeper reflection on whether internal incentives within the judiciary may be contributing to inefficiency, rather than solving it, thereby calling for more comprehensive reforms aimed at aligning judicial practices with economic and social priorities.

Secondly, the inefficiency of the regional courts in the Polish judiciary is significantly more pronounced than that observed in district courts, which follows from other research carried out by the author of this paper (Florczak, 2025a, 2025b). This is particularly concerning, given the pivotal role regional courts play not only within the judicial system's overall structure and daily operations, but also in the career development of judges. Regional courts serve as a crucial stage in the judicial

career promotion process. Judges working in regional courts tend to enjoy higher prestige, better salaries, and greater judicial power, particularly in handling appeal cases, while also having a lighter workload in terms of time required for their duties.

Thirdly, it is important to note that DEA, even when modified to include the necessary logical constraints (constrained DEA), remains a conservative method for estimating technical efficiency. This is because, for each DMU, the method maximises the ratio of the sum of weighted outputs to the sum of weighted inputs, resulting in the highest—and most optimistic—efficiency score possible within the imposed constraints. Also, identifying a given DMU as the most efficient does not imply that it was, or is, functionally perfect or ideal. There may still be a significant, yet difficult to quantify, gap between technical efficiency and an ideal functional benchmark, a gap that undoubtedly exists in reality.

To empirically illustrate this point, let us refer to the survey evaluation of the Polish judiciary's performance (CBOS, 2022, p. 4). According to this study, the percentage of negative opinions for the year 2005 was the highest across the entire survey sample (years 2005–2022), despite the fact that based on the objective quantitative indicators discussed in this article, 2005 was one of the more efficient—if not the most efficient—year for the functioning of Poland's regional judiciary. This discrepancy does not undermine the conclusion regarding the superiority of objective measures over subjective ones in assessing the efficiency of public institutions. However, it is also fair to assert that even in 2005, the judicial system could have performed better, particularly from the perspective of court users, as the rate of negative evaluations of the justice system was significantly lower in other years of the CBOS survey.

Identifying empirically the most historically efficient benchmark years can, at best, serve as a first step towards more qualitatively-oriented research, aimed at determining what factors contributed to making these distinguished DMUs or years more efficient than others. However, such a task would require interdisciplinary expertise and the involvement of various stakeholders. This is because DEA is inherently a descriptive method used to estimate relative efficiencies among homogeneous DMUs. While it provides insight into variations in the efficiency of individual DMUs over a given period, it does not explain the underlying causes for such variations. Consequently, the present study confines itself to estimating and comparing the actual levels of technical efficiency achieved by the Polish district court system across successive years of its operation, thereby identifying periods of higher efficiency as well as years in which the system deviated most strongly from its historically best-performing benchmarks. By employing modified variants of classical DEA models proposed in this research, it becomes possible to obtain a more reliable assessment of the historical diversification of system efficiency. In contrast, reliance on unconstrained radial DEA models—still predominant in practice—may generate distorted efficiency rankings and, as a result, lead to potentially misleading policy recommendations, particularly when such rankings are subsequently used as a basis for investigating the determinants of observed efficiency differences.

Accurately estimating the efficiencies of a set of DMUs is one challenge, but identifying—and potentially quantifying—the factors influencing those efficiencies

at different levels is another. However, if the first step—correctly estimating efficiencies—is mishandled, any conclusions drawn about the second stage, regarding the causes, will be flawed.

Thus the final conclusion is that the methodological considerations raised in this study regarding the application of DEA to the analysis of judicial efficiency extend beyond the specific empirical context examined in this research.

References

Act of July 27, 2001, *Law on the System of Common Courts*, Journal of Laws 2001 No. 98, item 1070 (Pol.), consolidated text: Journal of Laws 2024, item 334, as amended.

Agrell, P. J., Mattsson, P., & Månsson, J. (2019). Impacts on efficiency of merging the Swedish district courts. *Annals of Operations Research, 288*, 653–679. https://doi.org/10.1007/s10479-019-03304-0

Antonucci, L., Crocetta, C., & D'Ovidio, F. D. (2014). Evaluation of the Italian judicial system. *Procedia Economics and Finance, 17*, 121–130.

Avkiran, N. K. (2015). *Productivity in the service sector with data envelopment analysis*. SSRN Working Paper. https://doi.org/10.2139/ssrn.2627576

Banker, R., Emrouznejad, A., Lopes, A. L. M., & Rodrigues de Almeida, M. (Eds.). (2014). Data envelopment analysis in the public sector. *Socio-Economic Planning Sciences, 48*(1), 1–104. https://doi.org/10.1016/j.seps.2013.12.005

Bełdowski, J., Dąbroś, Ł., & Wojciechowski, W. (2020). Judges and court performance: A case study of district commercial courts in Poland. *European Journal of Law and Economics, 50*(1). https://doi.org/[do uzupełnienia]

CBOS. (2022). *Społeczne oceny wymiaru sprawiedliwości* [Social evaluations of the justice system] (Komunikat z Badań Nr 95/2022).

Cook, W. D., Tone, K., & Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega, 44*, 1–4. https://doi.org/10.1016/j.omega.2013.09.004

Data envelopment analysis in the public sector. (2014). *Socio-Economic Planning Sciences, 48*(1), Special issue.

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*, 477–493. https://doi.org/10.1007/s10657-011-9258-3

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., & Sarrico, C. S. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research, 132*, 245–259. https://doi.org/10.1016/S0377-2217(00)00149-1

El Bialy, N., & Garcia-Rubio, M. A. (2011). *Assessing judicial efficiency of Egyptian first instance courts: A DEA analysis*. MAGKS Joint Discussion Paper Series in Economics. https://www.econstor.eu/bitstream/10419/56541/1/657907855.pdf

Falavigna, G., Ippoliti, R., Manello, A., & Ramello, G. (2015). Judicial productivity, delay and efficiency: A directional distance function (DDF) approach. *European Journal of Operational Research, 240*, 592–601. https://doi.org/10.1016/j.ejor.2014.07.014

Florczak, W. (2025a). Revisiting DEA application in the judiciary: Assessing macro-efficiency of the Polish district court system in the 21st century. *Scientific Papers of Silesian University of Technology. Organization and Management Series, 226*, 257–270. https://doi.org/10.29119/1641-3466.2025.226.20

Florczak, W. (2025b). Limitations of two-dimensional indicators and unconstrained radial DEA models in evaluating judicial efficiency: Insights from Poland's appellate court system. *Ekonomia i Prawo. Economics and Law, 24*(4), 405–440. https://doi.org/10.12775/EiP.2025.20

Gupta, M., & Bolia, N. B. (2020). Efficiency measurement of Indian high courts using DEA: A policy perspective. *Journal of Policy Modeling, 42*, 1372–1393. https://doi.org/10.1016/j.jpolmod.2020.06.002

Ippoliti, R., & Tria, G. (2020). Efficiency of judicial systems: Model definitions and output estimation. *Journal of Applied Economics, 23*(1), 385–408. https://doi.org/10.1080/15140326.2020.1776977

Kittelsen, S. A. C., & Forsund, F. R. (1992). Efficiency analysis of Norwegian district courts. *Journal of Productivity Analysis, 3*(3), 277–306. https://doi.org/10.1007/BF00158357

Krajewski, E., & Milewski, R. (2018). *Podstawy ekonomii* [Fundamentals of economics]. PWN.

Lee, H. S., Chu, C. W., & Zhu, J. (2011). Super-efficiency DEA in the presence of infeasibility. *European Journal of Operational Research, 212*(1), 141–147. https://doi.org/10.1016/j.ejor.2011.01.022

Lewin, A. Y., Morey, R. C., & Cook, T. J. (1982). Evaluating the administrative efficiency of courts. *Omega, 10*(4), 401–411. https://doi.org/10.1016/0305-0483(82)90019-6

Malcarne, A., & Ramello, G. B. (2015). Judicial independence, judges' incentives and efficiency. *Review of Law & Economics, 11*(2), 149–169. https://doi.org/10.1515/rle-2015-0024

Mattsson, P., & Tidåsen, C. (2019). Potential efficiency effects of merging the Swedish district courts. *Socio-Economic Planning Sciences, 67*, 58–68. https://doi.org/10.1016/j.seps.2018.09.002

Ostrom, B. J., & Hanson, R. A. (2003). *Efficiency, timelines and quality: A new perspective from nine state criminal trial courts*. National Institute of Justice & State Justice Institute. https://www.ojp.gov/pdffiles1/nij/178403-1.pdf

Panwar, A., Olfati, M., Pant, M., & Snášel, 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*, 5397–5426. https://doi.org/10.1007/s11831-022-09770-3

Pedraja, F., & Salinas, J. (1996). An assessment of the efficiency of Spanish courts using DEA. *Applied Economics, 28*, 1391–1403. https://doi.org/10.1080/000368496327651

Peyrache, A., & Zago, A. (2016). Large courts, small justice! The inefficiency of the optimal structure of the Italian justice sector. *Omega, 64*, 42–56. https://doi.org/10.1016/j.omega.2015.11.002

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, 47*, 1–16. https://doi.org/10.1016/j.omega.2014.02.007

Sąd Okręgowy w Warszawie. (2013, December 12). *The judicial system in Poland*. Biuletyn Informacji Publicznej Sądu Okręgowego w Warszawie. Updated May 7, 2024.

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), 127–144. https://doi.org/10.1007/s10657-005-1733-2

Siemaszko, A., & Ostaszewski, P. (2013). *Efektywność kosztowa sądownictwa powszechnego* [Cost efficiency of the common courts]. Instytut Wymiaru Sprawiedliwości. https://iws.gov.pl/wp-content/uploads/2018/08/Efektywno%C5%9B%C4%87-s%C4%85downictwa-powszechnego-oprac-12-1.pdf

Smith, P. C., & Street, A. (2005). Measuring the efficiency of public services: The limits of analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 168*, 401–417. https://doi.org/10.1111/j.1467-985X.2005.00355.x

Study on the functioning of judicial systems in the EU Member States. (2022). European Commission for the Efficiency of Justice (CEPEJ). https://commission.europa.eu/system/files/2022-05/part_1_-_eu_scoreboard_-_indicators_-_deliverable_0.pdf

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), 183–210. https://doi.org/10.1007/978-0-387-25534-7_17

Ustawa z dnia 27 lipca 2001 r. – Prawo o ustroju sądów powszechnych, Dz.U. 2001 nr 98 poz. 1070, tekst jedn.: Dz.U. 2024 poz. 334 z późn. zm.

Voigt, S. (2016). Determinants of judicial efficiency: A survey. *European Journal of Law and Economics, 42*, 183–208. https://doi.org/10.1007/s10657-016-9531-6

Voigt, S., & El Bialy, N. (2016). Identifying the determinants of judicial performance: Taxpayers' money well spent? *European Journal of Law and Economics, 41*, 283–319. https://doi.org/10.1007/s10657-014-9474-8

Yeung, L., & Azevedo, P. (2011). Measuring efficiency of Brazilian courts with data envelopment analysis (DEA). *IMA Journal of Management Mathematics, 22*(4), 343–356. https://doi.org/10.1093/imaman/dpr002

Yeung, L. (2018). *Measuring efficiency of Brazilian courts: One decade later.* SSRN Working Paper. https://doi.org/10.12660/rda.v279.2020.81376

Appendix

Additional materials related to this article are available in the online supplementary appendix at the journal's website