

Falavigna, G.; Ippoliti, R.

Article

Model definitions to identify appropriate benchmarks in judiciary

Journal of Applied Economics

Provided in Cooperation with:

University of CEMA, Buenos Aires

Suggested Citation: Falavigna, G.; Ippoliti, R. (2022) : Model definitions to identify appropriate benchmarks in judiciary, Journal of Applied Economics, ISSN 1667-6726, Taylor & Francis, Abingdon, Vol. 25, Iss. 1, pp. 339-360,
<https://doi.org/10.1080/15140326.2021.2021128>

This Version is available at:

<https://hdl.handle.net/10419/314166>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Model definitions to identify appropriate benchmarks in judiciary

G. Falavigna & R. Ippoliti

To cite this article: G. Falavigna & R. Ippoliti (2022) Model definitions to identify appropriate benchmarks in judiciary, Journal of Applied Economics, 25:1, 339-360, DOI: [10.1080/15140326.2021.2021128](https://doi.org/10.1080/15140326.2021.2021128)

To link to this article: <https://doi.org/10.1080/15140326.2021.2021128>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 23 Mar 2022.



Submit your article to this journal [↗](#)



Article views: 1245



View related articles [↗](#)




View Crossmark data [↗](#)



Citing articles: 5 View citing articles [↗](#)

Model definitions to identify appropriate benchmarks in judiciary

G. Falavigna^a and R. Ippoliti^b 

^aResearch Institute on Sustainable Economic Growth (IRCrES), National Research Council of Italy (CNR), Moncalieri, Italy; ^bFaculty of Business Administration and Economics, Bielefeld University, Bielefeld, Germany

ABSTRACT

In this manuscript we present a comparative analysis of benchmarks based on technical efficiency scores computed using Data Envelopment Analysis with two different model specifications. In one case, we adopt the number of settled cases as output and human resources as input; in the other case, we adopt the same model definition but with judicial expenditure as additional key input. Our findings show that the model specification containing both judicial expenditure and human resources is more appropriate than the model based only on human resources. Moreover, we show that, without considering the additional variable costs generated within the production process, those courts incorrectly identified as benchmarks might mislead the policy makers dealing with the reform process.

ARTICLE HISTORY

Received 2 September 2021
Accepted 16 December 2021

KEYWORDS

Institutional efficiency;
judiciary; benchmark
analysis; public budget;
separability conditions

1. Introduction

Policy makers need to identify drivers of inefficiency and the key criteria that may steer the re-organisation process. In order to achieve these targets, a bottom-up approach can successfully pinpoint the crucial procedural issues and the interventions needed to improve the system under investigation, involving both operators and final users. However, this process has to be supported by efficiency benchmarks, able to provide a picture of the current organisational structure through Technical Efficiency (TE) scores. This is why Operational Research (OR) can be a valuable tool to help policy makers reform national public systems through validated techniques around which the interests of the stakeholders can converge, creating a common consensus on the policy reforms introduced. This takes on even greater significance if we consider the need to reduce public expenditure in the current age of austerity, which may lead, as one of its direct effects, to a reduction in public services, with negative repercussions on society.

Considering the Italian judicial system between 2014 and 2018 and its first instance courts as Decision Making Units (DMUs), we analyse the supply of justice to identify benchmarks and appropriate model definitions, proposing a comparative analysis of TE scores computed using Data Envelopment Analysis with two different model specifications. In one case, we adopt the number of settled cases as output and human resources

CONTACT R. Ippoliti  roberto.ippoliti@uni-bielefeld.de  Faculty of Business Administration and Economics, Bielefeld University, Bielefeld, Germany

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(i.e., judges and staff) as input; in the other case, we adopt the same model definition but with judicial expenditure data as additional key inputs (i.e., costs, allowances and fees). The former model specification is based exclusively on human resources and represents the approach suggested by the current literature, while the latter model is based on both human resources and additional variable costs, and it is proposed for the first time in this work.

The optimization problem dealt with here revolves around the assumption that the actual objective function of courts is to supply justice for society, that is to say, our first instance courts must be able to maximize the number of settled cases, guaranteeing the correct functioning of society. Based on the results gathered, we highlight that the model specification with both judicial expenditure and human resources as inputs is more appropriate than the alternative one, which is based exclusively on human resources. Moreover, we show that, without considering the additional variable costs generated within the production process, the DMUs incorrectly identified as benchmarks might mislead the policy makers dealing with the reform process.

The manuscript is organized as follows. [Section 2](#) examines current literature and then, assuming that policy makers are interested in maximizing the supply of justice not only depending on the available human resources but also with respect to the public expenditure generated by that production process, we identify the key elements to benchmark first instance courts. [Section 3](#) sets out the methodology adopted to estimate TE scores (Data Envelopment Analysis) and productivity over time (Malmquist indexes). [Section 4](#) highlights some data and descriptive statistics of the judicial system under investigation, while [Section 5](#) illustrates the results of the empirical analysis. Finally, [Section 6](#) offers some conclusions.

2. Theoretical background: judicial efficiency and budget constraints

A number of methods have been used to measure judicial efficiency and which approach may be best clearly depends on the targets pursued and on the stakeholders involved in analysing the judicial system. Society, for instance, may pay more attention to the time needed to settle a judicial case, i.e., how long someone has to wait for justice to be served (e.g., Christensen & Szmer, 2012). Conversely, policy makers may well try to make the supply of justice more efficient by improving the productivity of judges and the performance of courts. To do this, they need reliable benchmarks, such as, for example, clearance rates (e.g., Dakolias, 1999) and TE scores (e.g., Agrell, Mattsson, & Månsson, 2019), to stratify courts, identify the most efficient type of organisation and implement a reform in line with their findings.

In detail, scholars have suggested estimating TE scores by means of mathematical programming techniques (e.g., Data Envelopment Analysis, Free Disposal Hull, Directional Distance Function and Malmquist indexes), which have been successfully applied to the study of various national judicial systems (e.g., Ferro, Romero, & Romero-Gómez, 2018; Giacalone, Nissi, & Cusatelli, 2020; Mattsson & Tidana, 2018; Schneider, 2005; Silva, 2018). Ippoliti and Tria (2020) offers a comprehensive review of this literature, listing the inputs and outputs adopted, the judicial systems under investigation and the mathematical programming techniques used. Despite some heterogeneity, the number of settled cases seems to emerge as the main output. At times, this output is

presented as an aggregate measure (e.g., Falavigna, Ippoliti, & Ramello, 2018; Finocchiaro Castro & Guccio, 2015); in other cases, a more precise output measure is provided by disaggregating the supply of justice according to case matters (e.g., Kittelsen & Førsund, 1992; Peyrache & Zago, 2016). The contribution by Ferro, Oubiña, and Romero (2020) features three output variables, i.e., settled cases, average length of time to settle them and proportion of appealed sentences. Looking at the inputs, they are even more heterogeneous, with some papers focusing solely on human resources (e.g., Deyneli, 2012; Ferrandino, 2012) and other works also considering the demand for justice (e.g., Falavigna, Ippoliti, & Manello, 2019; Finocchiaro Castro & Guccio, 2018). It is worth noting that disaggregating the demand for justice and/or its supply on the basis of case matters is key to accounting for their varying degrees of complexity, an aspect that is likely to impact on the productivity of judges (Örkényi, 2021).

As mentioned, the definition of a suitable model depends on the main targets pursued by policy makers. If the objective is to boost the supply of justice and simultaneously keep public expenditure under control, the issue is whether the above model definitions are able to identify appropriate benchmarks. Could the reform process fail if policy makers relied on a model definition that does not go beyond human resources? Would it be preferable to combine the current approaches with a model that controls for the judicial expenditure ascribable to court procedures and affecting the public budget? These are the research questions that we aim to answer in the present work.

To fulfil their duty as suppliers of justice, courts have to bear additional variable costs, i.e., costs specifically related to judicial production. According to the civil/criminal procedure of each type of case, the judges may require the support of professionals to determine the extent of the damage caused (e.g., forensic tests or psychiatric evaluations), technicians to inspect any evidence gathered (e.g., handwriting analyses or ballistic reports), along with lawyers to provide legal aid. All of these costs are variable, i.e., they change depending on the supply of justice, while the internal organisation of each court has an impact on the amount of resources needed. Differently put, if the demand for justice is zero, the courts do not incur any of these costs. In addition, major differences are found across courts in terms of internal structure adopted. Thus, the reduction in expenditure, caused by budget constraints, that policy makers pursue leaves the courts with only two options, either re-organize to enhance efficiency or reduce the supply of justice to comply with the new financial limits imposed. Improving court performance while bringing down expenditure is clearly the better outcome and, in the current age of austerity, it is what policy makers want too. Yet, if no clear efficiency benchmarks can be relied on, the courts may have no choice but to reduce the supply of justice, with strongly negative effects on both society and market dynamics (Falavigna et al., 2019; Giacomelli & Menon, 2017). In view of the above, if the aim is to benchmark courts and keep public expenditure under control, one cannot disregard the variable costs generated by the courts' production process, since they are one of its main inputs. Furthermore, the financial means to cover the variable costs of justice come from the public budget, leaving fewer resources for other sectors (e.g., health care or welfare) or forcing the government to increase taxation. Of course, a third option is possible, i.e., implementing new judicial procedures able to lower the additional costs generated by the settlement of cases. This approach is viable only when precise benchmarks are available to accurately assess the

determinants of judicial inefficiency. This can be done by following the two-stage approach mentioned above but, as Ippoliti and Tria (2020) point out, the analysis of variable costs is not featured in any of the studies described. Consequently, to the best of our knowledge, no model definitions exist that include judicial expenditure as input of the judicial production process. An attempt in this direction can be found in Falavigna and Ippoliti (2021), but here judicial expenditure is considered an undesirable output of the production process of justice. Yet, if these costs are understood as being necessary to deliver justice (although not exclusively covering the processing of cases in courts), they should logically be included into the production process as inputs. Such a new model definition may perform better than the previous ones, since it combines commonly used estimations, mostly based on human resources, and the additional variable costs arising from the judicial production process.

Hence, the model definition presented in this contribution might not only be more appropriate for the current age of austerity, with its emphasis on budget constraints, but also ensure greater accuracy in identifying benchmarks. Now, as regards the specific case study analyzed here (i.e., the Italian judicial system), we test the following hypotheses:

H_{1a} model specification based on human resources and judicial expenditure is more appropriate than a model based exclusively on human resources;

H₂ TE scores estimated without considering (jointly) judicial expenditure and human resources can mislead policy makers in identifying benchmarks.

The next section illustrates the model definition and methodology adopted and describes our specific case study (i.e., the Italian judicial system).

3. Methodology

By using Data Envelopment Analysis (DEA), we test our hypotheses through a comparative analysis of two different model definitions, i.e., applying non-parametric frontier methodologies with different input-output space. We compare the TE scores estimated through DEA with human resources as input (i.e., judges and staff) and through DEA with judicial expenditure and human resources as inputs (i.e., judges, staff, costs, allowances and fees), to determine whether the adoption of these costs as inputs of the judicial production process might actually make a difference. Moreover, we propose a sensitivity analysis, comparing the collected results with models that consider an additional input: caseload (i.e., incoming cases and pending cases at the beginning of every year). Next, by means of the Malmquist indexes, we investigate the trend of judicial productivity over time, as well as its decomposition into efficiency change and technology change, to collect more robust evidence about the first hypothesis. Lastly, according to the two-stage approach of Simar and Wilson (2007), we use the collected TE scores as dependent variables in a truncated regression model to analyze the main determinants of inefficiency.

3.1. Data Envelopment Analysis (DEA)

To compute TE scores, the DEA methodology is used, following the well-known CCR approach (i.e., Charnes, Cooper, & Rhodes, 1978). Through a DEA model it is possible to build a deterministic non-parametric production frontier, comparing the performance of several DMUs, i.e., first instance courts in this case study, and computing the TE scores on the basis of the radial distance of the subjects from the frontier.

The literature suggests setting up the model according to the DMUs' ability to maximize the outputs (output-orientation), taking equal inputs, or to minimize the inputs (input-orientation), taking equal outputs. More specifically, the input-oriented framework, based on the input requirement set and its efficient boundary, aims to reduce the input amounts as much as possible, while keeping at least the present output levels. In this approach, the output levels remain unchanged and the input quantities are reduced proportionately until the frontier is reached. This is the orientation generally adopted by decision makers when they can control the inputs but not the outputs. Conversely, the output-oriented approach maximizes the output levels, without varying the present input amounts. Indeed, the input set remains unchanged, while the output levels increase until the frontier is reached (Daraio & Simar, 2007a). Keeping in mind that the actual objective function of courts is the supply of justice for society, the output-oriented framework is applied in this study, as proposed by Farrell (1957). Accordingly, we study the ability of first instance courts (i.e., our DMUs) to maximize their output (i.e., number of settled cases), given the adopted inputs, which are human resources in Model A (i.e., judges and staff) and human resources and judicial expenditure in Model B (i.e., judges, staff, costs, allowances and fees). More precisely, the output under investigation is the whole supply of justice by these courts, considering both criminal and civil cases. The proposed models provide policy makers with the opportunity to identify benchmarks for policy reforms according to two different perspectives, i.e., the aforementioned standard approach based on human resources and an approach focusing on both human resources and judicial expenditure. At the same time, a comparison of the two models offers the chance to determine which one might be more appropriate in supporting policy makers and the reform process of the judicial system.

3.1.1. Environmental variables and separability conditions

We have adopted case matters as environmental variables, testing the separability conditions on these characteristics and verifying whether they may affect the definition of the frontier or its distribution. Daraio and Simar (2005) developed a fully nonparametric methodology based on conditional FDH and conditional order-m frontiers, without any convexity assumption on the technology. Two years later, the same authors presented a generalization of that methodology, introducing a unified approach for considering together convexity and non-convexity in conditional non-parametric frontiers (Daraio & Simar, 2007b). However, before computing the conditional frontier, Daraio, Simar, and Wilson (2018) suggest applying a test on the separability conditions, to evaluate whether the environmental variables are to be considered in the first stage (i.e., in the frontier computation) or in the second stage of the analysis (i.e., in the regression model). Simar and Wilson (2007) described this test for the first Daraio, Simar, and Wilson (2015), (2018) illustrated its technical development. The main idea of the test is to compare

unconditional and conditional efficiency scores, where the conditioning variables are environmental ones (e.g., case matters). The separability conditions require the environmental variables not to influence the frontier but only the distribution of efficiency. Conditional estimates consider the environmental variables in the sampling procedure before the computation of the efficiency and, under the null hypothesis, unconditional and conditional efficiency scores are not very different. Indeed, the rejection of the null hypothesis means that separability is violated because scores computed in the unconditional setting diverge from those in the conditional one. In this case, environmental variables affect the shape of the technology and results obtained in the second stage are difficult to interpret, with problems that resemble endogeneity. For this reason, a satisfying result of the test is to accept the null hypothesis, meaning that environmental variables can be used as regressors in the second stage. Simar and Wilson (2020) advise splitting the original sample in order to maintain independence between the sample means under comparison. In this paper, we test the separability conditions for environmental variables following the approach suggested by Simar and Wilson (2020). The number of split samples is set to 10 and the number of bootstrap replications is 1,000. The Epanechnikov kernel function is used to estimate conditional efficiency scores. We have run 20 tests, i.e., five for each year and for each model considering both CRS and VRS. Accepting the null hypothesis means that there is no statistically significant difference between the means of conditional and unconditional efficiency estimates, which allows considering environmental variables as regressors in the second stage (see [Tables A1 and A2](#) in Annex A).

3.1.2. *DEA model and returns to scale*

As for returns to scale, we have adopted Variable Returns to Scale (VRS), which have been computed according to Banker, Charnes, and Cooper (1984).¹ The adoption of VRS has been tested and validated following the procedure suggested by Kneip, Simar, and Wilson (2016), Daraio et al. (2018) and Simar and Wilson (2020), rejecting the null hypothesis and accepting VRS as more appropriate than CRS. According to Daraio et al. (2018), the test requires the two sample means to be independent of each other, which is done by randomly splitting the original sample. In Simar and Wilson (2020), a Monte Carlo simulation is used to observe the rejection rate of different tests by comparing different sample splits, leading to the conclusion that the power of multiple-split tests is generally superior to that of single-split test. However, there is no rule to determine the number of splits, and it is up to researchers to decide how many sample splits ought to be applied. This decision is affected by the computational burden since, if the number of splits increases, the time needed for computation increases as well. In our test, we have used 10 splits and 1,000 replications, as suggested by Simar and Wilson (2020). Accordingly, 10 different tests have been applied, i.e., one for each year of the analysis (from 2014 to 2018) and for both models. [Table A3](#) in Annex A presents the results of these tests, which point to the rejection of the null hypotheses and the adoption of VRS.

¹In-depth descriptions of the input and output measures under different returns to scale can be found in Coelli et al. (1998).

The proposed DEA models measure the TE of a set of $j = 1, \dots, n$ observations (i.e., DMUs). These observations transform a vector of $I = 1, \dots, p$ input $\mathbf{x} \in \mathbb{R}_{++}^p$ into a vector of q outputs $\mathbf{y} \in \mathbb{R}_{++}^q$ using the technology represented by the following constant returns to scale production possibility set: $P_{crs} = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \geq X\lambda, \mathbf{y} \leq Y\lambda, \lambda \geq 0\}$, where $X = (x)_{ij} \in \mathbb{R}_{++}^{q \times n}$, $Y = (y)_{ij} \in \mathbb{R}_{++}^{p \times n}$, and $\lambda = (\lambda_1, \dots, \lambda_n)$ is a semipositive vector allowing a convex combination of inputs and outputs.

Thus, by solving the following linear programming, it is possible to estimate the output-oriented TE score of each DMU (Farrell, 1957):

$$\begin{aligned} & \underset{\varphi, \lambda}{\text{Max}} \varphi \\ & \text{subject to :} \\ & \varphi \mathbf{y}_o \leq Y\lambda \\ & X\lambda \leq \mathbf{x}_o \\ & N1'\lambda = 1 \\ & \lambda \geq \mathbf{0} \end{aligned} \tag{1}$$

Where $1 \leq \varphi \leq +\infty$, and the optimal solution is when φ is equal to 1. Therefore, if φ is higher than 1, the observation is radially inefficient and $(\lambda X, \lambda Y)$ outperforms $(\mathbf{x}_o, \mathbf{y}_o)$. $N1$ is an $N \times 1$ unitary vector that allows adding a convexity constraint. In addition, $\varphi - 1$ is the proportional increase in outputs that could be achieved by the i -th DMU, with input quantities held constant, and $1/\varphi$ defines a TE score that varies between 0 and 1 (Coelli, 1996).

Note that the available data cover the period from 2014 to 2018 and a specific frontier has been computed for each year (i.e., 5 different DEA models, for both Model A and Model B). Moreover, as explained in the next section, the sample allows assessing the Malmquist indexes between 2014 and 2018 with the aim to evaluate the productivity of DMUs and its changes over time.

Finally, in order to calculate the DEA models, we have worked with R-statistics 4.0.2 and the FEAR 3.1 package (Wilson, 2008), applying the bootstrap procedure to all the computations with a number of replications equal to 2,000 (Simar & Wilson, 1999, 2007).

3.2. Malmquist indexes

As suggested by Coelli, Rao Prasada, and Battese (1998) and Cooper, Seiford, and Tone (2007), the DEA methodology with output orientation is used in order to compute the Malmquist Indexes (MI), assuming CRS, i.e., assuming that the returns to scale of each observation are time invariant. This approach makes it possible to evaluate the efficiency change over time. Indeed, the MI is an index representing the Total Factor Productivity (TFP) growth of a DMU, in that it reflects progress or regress in efficiency along with progress or regress of the technology frontier over time under the multiple inputs and multiple outputs frameworks (Tone, 2004).²

²The calculation of the Malmquist indexes starting from the DEA approach is exhaustively discussed in Coelli (1996).

Through the Malmquist indexes, we assess changes in productivity and its decomposition following the main arguments put forward in the literature (Lovell, 2003; Mussard & Peypoch, 2006). The result is a Total Factor Productivity (TFP) index computing the ratio between outputs and inputs at different times.

According to Färe and Grosskopf (1996) and Färe, Grosskopf, Lindgren, and Roos (1992), in order to formalize the model, we can assume x to be the sole input and y the sole output, both of which are available over two time periods (t and $t + 1$). Then, TFP can be described as follows:

$$TFP = \frac{y^{t+1}/x^{t+1}}{y^t/x^t} \quad (3)$$

TFP can assume values from 0 to $+\infty$, with results higher than 1 denoting improvements in total productivity. Then, according to Färe et al. (1992), we propose the following decomposition into efficiency change (*eff*) and technology change (*tech*), where f_t and f_{t+1} represent the frontiers at time t and $t + 1$:

$$\begin{aligned} M_0(x^{t+1}, y^{t+1}, x^t, y^t) &= \left(\frac{\frac{y_{t+1}}{f_{t+1}(x_{t+1})}}{\frac{y_t}{f_{t+1}(x_t)}} \cdot \frac{\frac{y_{t+1}}{f_t(x_{t+1})}}{\frac{y_t}{f_t(x_t)}} \right)^{1/2} \\ &= \underbrace{\left(\frac{y_{t+1}}{f_{t+1}(x_{t+1})} \cdot \frac{f_t(x_t)}{y_t} \right)}_{\substack{\text{Efficiency} \\ \text{Change} \\ (\text{eff})}} \cdot \underbrace{\left(\frac{f_{t+1}(x_{t+1})}{f_t(x_{t+1})} \cdot \frac{f_{t+1}(x_t)}{f_t(x_t)} \right)}_{\substack{\text{Technology} \\ \text{Change} \\ (\text{tech})}}^{1/2} \\ &= tfp_{t, t+1} = M_{t, t+1} = eff_{t, t+1} h_{t, t+1} \end{aligned} \quad (4)$$

Simplifying, the Malmquist indexes are calculated as a ratio in which the numerator represents the production achieved at $t + 1$ and the denominator the level of output at time t . This decomposition is very interesting because it allows us to understand how each single DMU reaches the new frontier. In other words, we can determine whether a DMU has been able to achieve the new production level by better exploiting either its resources (i.e., efficiency change) or its technology (i.e., technology change). The *eff* component captures changes in efficiency from t to $t + 1$, that is to say, the ability of a DMU to reach the frontier at t using resources available at $t + 1$, while the other conditions at time t remain unchanged (Falavigna et al., 2018). The *tech* component describes shifts in the technology frontier from t to $t + 1$ (Färe & Grosskopf, 1996). This index captures the effect of the average technological progress (or regress) of the judicial system. Values higher than 1 mean that, between t and $t + 1$, technical progress was achieved.

It is worth underlining that TFP makes it possible to obtain only an approximation of technology. Note also that, according to the current literature (e.g., Badunenko & Kumbhakar, 2017; Gitto, 2017; Isik & Hassan, 2003), these efficiency and technology changes are the most commonly applied decompositions of Malmquist indexes.

Nevertheless, other factorisations are also suggested by scholars (i.e., Ray & Desli, 1997; Wheelock & Wilson, 1999), an in-depth explanation of which can be found in Falavigna et al. (2018).

Finally, in order to calculate the Malmquist indexes and their decompositions, we have worked with R-statistics 4.0.2 and the FEAR 3.1 package (Wilson, 2008), also in this case applying the bootstrap procedure to all the computations with a number of replications equal to 2,000 (Simar & Wilson, 1999).

3.3. Second stage: empirical strategy

According to Simar and Wilson (2007), in order to investigate the determinants of judicial inefficiency, the TE scores estimated in the first stage become the dependent variables of the regression models proposed in the second stage, controlling for several external and internal variables. Considering the geographical competence of first instance courts, which are our observations, the external environmental variables are:

- five geographical macro areas (NUTS 1), which are dummy control variables introduced in the model to account for social and cultural heterogeneity among our observations;
- years (i.e., 2014–2018), which are dummy control variables to capture time effects.

As for the first instance courts' internal variables, considering the disaggregated supply of justice by judicial macro areas, we focus on the procedural variables (i.e., the weight of these judicial areas over the total settled cases):

- *criminal procedures*, which is equal to the percentage of settled criminal cases;
- *mortgage foreclosure* and *bankruptcy*, which is equal to the percentage of insolvency cases featuring, respectively, a mortgage foreclosure procedure or a bankruptcy procedure;
- *labour and pension*, which is equal to the percentage of cases involving employees and employers (both public and private), as well as pension institutions;
- *voluntary civil process*, *ordinary civil procedures* and *special civil procedures*, which is equal to the percentage of civil cases falling into these three different case matters, according to the civil code;
- *other civil procedures*, which is a residual variable equal to the percentage of all the other civil cases.

In other words, this second stage aims to shed light on whether there is a statistically significant relation between court efficiency and case matters. Indeed, a DMU might be efficient just because cases are settled by means of more effective procedures, reducing total workload and increasing overall productivity. Depending on the results, there might be the opportunity to interpret judicial procedures in terms of efficiency drivers.

Table 1. Inputs and outputs adopted in the data envelopment analysis (Italy, 2018).

Type	Variable	Mean	Std. Dev.	Min	Max	Obs
Human resources	Judges	33	47	5	352	139
	Staff	95	109	14	853	139
Raw material*	Caseload	54,436	64,189	7,498	491,788	139
Judicial expenditure	Costs	376,560.37	621,676.68	8,176.31	4,303,297.90	139
	Fees	1,618,853.21	2,172,309.23	95,201.35	13,906,355.65	139
	Allowances	185,923.85	331,083.31	2,637.58	3,458,507.52	139
Supply of justice	Settled cases	26,406	32,010	3,212	245,472	139

Elaborations according to data made available by the Ministry of Justice and the Supreme Judicial Council

**This input has been introduced only in the sensitivity analysis, its results are proposed in the Annex B*

4. Data and descriptive statistics

Table 1 proposes some descriptive statistics about inputs and outputs based on the selected case study and the aforementioned model definition, considering our DMUs in 2018. The data were extracted from the databases of the Ministry of Justice and the Supreme Judicial Council, and disaggregated at the first instance level. Looking at judicial expenditure, we can observe that fees make up the most significant part of these variable costs for the supply of justice, with more than 1.6 million euro on average per court and a maximum value equal to more than 13 million euro. However, for what concerns standard deviation, we can also observe a certain level of heterogeneity among the observations, which might be due to differences in size among courts, as well as different internal organisational approaches.

Analysing the Italian geographical macro areas, Tables 2 and 3 display some additional descriptive statistics regarding our sample, highlighting both the average time needed to settle cases (Table 2) and the stratigraphy of pending cases (Table 3). Note that these are two key proxies adopted by the Italian Ministry of Justice to monitor the performance of courts, which is why they are used here to test the first hypothesis, i.e., whether Model B (with human resources and judicial expenditure) is more appropriate than Model A (with human resources only). Specifically, Table 2 shows the average delay in settling civil and insolvency cases according to macro areas. Case matters are a good proxy for the production lines of justice, while the related procedures represent the current technology adopted by judges in supplying justice. Based on the current literature (e.g., Falavigna & Ippoliti, 2021; Ippoliti & Tria, 2020), judicial procedures can affect the efficiency of the courts and, when looking at the insolvency procedures, Table 2 confirms significant differences among case matters. Considering the stratigraphy of pending cases in 2018 by macro areas, Table 3 shows the average number of pending cases older than 10 years (i.e.,

Table 2. Disaggregated descriptive statistics on judicial delay by case matter (Italy, 2018).

Case matter	North East	North West	Centre	South	Islands	Italy
Mortgage foreclosure	721	770	1,028	1,320	1,464	1,066
Bankruptcy	2,231	2,432	3,085	3,681	5,148	3,257
Labour and pension	313	297	467	592	732	509
Ordinary civil procedures	595	557	856	1,157	1,111	914
Special civil procedures	45	62	61	100	103	81
Voluntary civil process	87	114	114	109	105	108

Source of data: Italian Ministry of Justice (access: October 2020)

Table 3. Descriptive statistics on the stratigraphy of pending cases according to case matters (Italy, 2018).

Macro areas	Insolvency cases		Civil cases	
	(up to 2007)	(2008–2010)	(up to 2007)	(2008–2010)
North East	2.12%	4.46%	0.40%	0.81%
North West	3.18%	4.52%	0.35%	0.67%
Center	6.78%	6.57%	1.22%	1.59%
South	11.16%	5.54%	1.50%	3.50%
Islands	16.32%	7.33%	1.17%	2.47%
ITALY	7.69%	1.87%	0.96%	0.65%

Source of data: Italian Ministry of Justice (access: October 2020)

Table 4. Descriptive statistics on the variables introduced in the second stage (Italy, 2018).

Variable	Obs	Mean	Std. Dev.	Min	Max
North West	139	0.201	0.403	0.000	1.000
North East	139	0.158	0.366	0.000	1.000
Centre	139	0.201	0.403	0.000	1.000
South	139	0.281	0.451	0.000	1.000
Islands	139	0.158	0.366	0.000	1.000
Criminal procedures	139	0.318	0.056	0.201	0.459
Mortgage foreclosure	139	0.112	0.031	0.055	0.313
Bankruptcy	139	0.013	0.005	0.003	0.028
Labour and pension	139	0.085	0.037	0.026	0.209
Ordinary civil procedures	139	0.150	0.028	0.071	0.244
Special civil procedures	139	0.113	0.028	0.047	0.205
Voluntary civil process	139	0.081	0.037	0.028	0.284
Other civil procedures	139	0.128	0.063	0.034	0.358

Source of data: Italian Ministry of Justice (access: October 2020)

pending cases up to 2007) and between 7 and 10 years old (i.e., pending cases between 2008 and 2010). Observing the data on insolvency procedures, we can easily detect an efficiency gap between the North and the South of Italy. On average, the percentage of pending cases older than 10 years is 11.16% in the South of Italy and 16.32% in the Islands, while the figure drops to 2–3% in the North of Italy. These numbers are extremely relevant since, the older the pending case, the higher the probability that parties could successfully sue the Italian Ministry of Justice for excessive judicial delay. As for civil procedures, an efficiency gap is present in this case too, but it is significantly smaller. This result is rather unsurprising, in view of the fact that insolvency procedures are more complex and characterized by higher expected litigiousness.

Finally, considering the second stage, Table 4 presents some descriptive statistics about the proposed independent variables in 2018. By observing their values, we can better understand their significance. On average, 32% of the supply of justice concerns criminal procedures, while insolvency procedures amount to respectively 11% (mortgage foreclosure) and 1% (bankruptcy). Nevertheless, there is a great deal of heterogeneity among first instance courts and the composition of the supply of justice can affect their overall efficiency, since our DMUs can organize their internal sections and prioritize the settlement of selected case matters, which affects their ability to supply a higher level of justice.

5. Results

Based on the proposed empirical strategy we estimate the TE scores, comparing different model definitions to test hypothesis H_1 and H_2 . Afterwards, we estimate the Malmquist indexes to investigate judicial productivity, as well as efficiency change and technology change between 2014 and 2018, to collect more robust evidence on our first hypothesis.

5.1. First stage: DEA TE scores

Table 5 presents some descriptive statistics regarding the estimated TE scores in 2018, considering both Model A (i.e., DEA with human resources as input) and Model B (i.e., DEA with human resources and judicial expenditure as inputs). The table shows the top 10% DMUs in relation to both models, i.e., the most efficient courts in the 10th percentile. Note that the values range between 0 and 1, with 1 representing the efficiency benchmark in the comparative analysis (see Section 3.1). This means that, as the TE scores decrease, so does the distance from the efficiency frontier, indicating that the DMUs become more inefficient compared to the total population of first instance courts under investigation.

Looking at the top 10% DMUs, i.e., the most efficient 14 first instance courts, a different population distribution emerges, which is not surprising since there are two different model specifications. Nevertheless, is Model B more appropriate than Model A? Can we accept the validity of hypothesis H_1 , proposed above?

Comparing the two sub-samples of potential benchmarks in relation to some selected key judicial proxies, we cannot reject our first hypothesis. Regarding the stratigraphy of insolvency cases in 2018, the number of cases submitted between 2008 and 2010 is 16% lower for the benchmarks identified by Model B than for those identified by Model A. Also, the number of cases older than 10 years, i.e., pre-2007, is 21% lower. The stratigraphy of civil cases confirms this result, highlighting a lower number of cases in both periods equal to 23% for the benchmarks identified by Model B. As mentioned in the previous section, these proxies are used by the Italian Ministry of Justice to monitor the performance of courts since, the older the cases, the higher the chances that parties might sue the Ministry for excessive trial duration. On the other hand, considering the average time necessary to

Table 5. Technical Efficiency (TE) scores of top 10% DMUs (Italy, 2018).

Model A			Model B		
TE score	First instance court	Macro area	TE score	First instance court	Macro area
0.997	Verona	North East	0.994	Foggia	South
0.995	Tivoli	Centre	0.989	Ivrea	North West
0.994	Foggia	South	0.989	Bergamo	North West
0.990	Reggio Emilia	North East	0.987	Civitavecchia	Centre
0.987	Brescia	North West	0.985	Velletri	Centre
0.987	Velletri	Centre	0.985	Como	North West
0.984	Lecco	North West	0.985	Pordenone	North East
0.983	Civitavecchia	Centre	0.985	Reggio Emilia	North East
0.983	Bergamo	North West	0.984	Pescara	South
0.982	Biella	North West	0.983	Brescia	North West
0.981	Ivrea	North West	0.982	Rimini	North East
0.980	Torino	North West	0.981	Busto Arsizio	North West
0.980	Perugia	Centre	0.981	Torino	North West
0.980	Ravenna	North East	0.981	Modena	North East

settle cases, the benchmarks identified by Model B are characterized by better performance. Indeed, the time needed to settle insolvency cases is 6% shorter (i.e., almost 3 months shorter), while the time needed to settle civil cases is 8% shorter (i.e., almost 1 month shorter). What about the second hypothesis? Can these benchmarks mislead policy makers?

Taking the top 10% portion of the estimated TE scores as our benchmark, we can compare the two model definitions and observe how information gaps might mislead policy makers in the reform process. In other words, by comparing the two model definitions and the collected TE scores, we can determine the extent to which model definitions might support policy makers in identifying correct benchmarks and implementing a successful policy reform.

If we look at Model A, with human resources as input (i.e., the common model definition), some courts could indeed be identified as potential benchmarks, since their TE scores are among the top 10 % DMUs. However, these courts are not above the threshold if we consider the DEA model with financial expenditure as additional input (i.e., Verona, Tivoli, Lecco, Biella, Perugia and Ravenna). Hence, they are efficient in supplying justice if we focus on how their human resources are organized. Nevertheless, if we consider their variable judicial expenditure compared to that of other first instance courts, their performance could be improved, so as to achieve a higher level of justice (i.e., more settled cases) with the same costs. Examining the same issue, but from a different perspective, we can reflect on the interaction between the two model definitions. Indeed, if we look at Model B, with both human resources and judicial expenditure as inputs, some courts could be identified as potential benchmarks, since their TE scores are above the threshold. Yet, these courts are below the threshold if we consider the DEA model with human resources as input (i.e., Como, Pordenone, Pescara, Rimini, Busto Arsizio and Modena). Accordingly, they are efficient in supplying justice if we look at how their variable judicial expenditure is used. Nevertheless, if we consider their internal human resources organisation compared to that of other courts, their performance could be improved, so as to achieve a higher level of justice (i.e., more settled cases) with the same amount of judges and staff. These are the misleading results that would jeopardize the policy makers' work if a single approach were followed in the benchmark analysis or human resources were used as the sole input considered.

Without appropriate model definition, policy makers could mistakenly regard certain courts as an organisational benchmark, even though this is actually not the case. Based on the proposed scenario, our results show that the DMUs identified as *non-appropriate benchmarks* are equal to 43% of our sub-sample, suggesting that a benchmark analysis with information gaps is likely to mislead policy makers. Therefore, we cannot reject hypothesis H_2 , that is to say, TE scores estimated without considering judicial expenditure and how it affects the public budget can cause policy makers to identify incorrect benchmarks for policy reforms.

Lastly, Table A4 in Annex B presents the results of the proposed sensitivity analysis. In particular, we compare the collected results with alternative models that consider case-load as additional input (i.e., the total number of incoming cases and pending cases at the beginning of every year).

5.2. Productivity changes over time

So far, we have calculated the efficiency scores for every single year from 2014 to 2018, and these represent the ability of first instance courts to be efficient in terms of human resources (i.e., Model A) and judicial expenditure as additional input (i.e., Model B). However, these results do not provide evidence on performance improvements over time. To do this, we calculate the Malmquist productivity indexes between 2014 and 2018, estimating the TFP of the first instance courts, and we decompose them into efficiency change and technology change. Considering the top 10% DMUs in terms of the estimated efficiency change, the information in Table 6 further supports our conclusions about the appropriateness of Model B, while also providing a deeper understanding of judicial productivity over time.

Comparing the two sub-samples of potential benchmarks (i.e., top 10% DMUs) in relation to some selected key judicial proxies provides further evidence to support the idea that our first hypothesis cannot be rejected. Indeed, considering the top 10% DMUs' stratigraphy of insolvency cases in 2018, the relative frequency of cases submitted between 2008 and 2010 is 15% lower for the benchmarks identified by Model B than for those identified by Model A. The stratigraphy of civil cases confirms this result, even with a larger sub-sample. Indeed, if the top 20% DMUs are considered, the relative frequency of civil cases is still lower, i.e., equal to 5% (i.e., cases submitted between 2008 and 2010) and 7% (i.e., cases submitted up to 2007), for the benchmarks identified by Model B. Considering the average time necessary to settle cases, the benchmarks identified by Model B are characterized by better performance. Indeed, the time needed to settle insolvency cases is 6% shorter, while the time needed to settle civil cases is 17% shorter, which means almost 3 month shorter in both cases. Accordingly, we cannot reject hypothesis H_1 , even when performance improvements over time are taken into account. What about judicial dynamics?

Considering TFP in Table 6, Model B suggests that not all of the courts were able to improve their productivity over time (i.e., Malmquist index < 1), whereas the results of Model A show increases in the performance of our DMUs (i.e., Malmquist index > 1).

Table 6. Efficiency and technology change of top 10% DMUs (Italy, 2014–2018).

Model A				Model B			
First instance Courts	Efficiency change	Technology change	Malmquist index	First instance courts	Efficiency change	Technology change	Malmquist index
CAMPOBASSO	1.230	0.994	1.223	L'AQUILA	1.109	1.019	1.130
L'AQUILA	1.222	0.985	1.204	CAMPOBASSO	1.068	1.021	1.090
VALLO DELLA LUCANIA	1.055	1.029	1.085	GELA	1.041	0.997	1.038
AOSTA	1.053	1.029	1.083	SONDRIO	1.041	0.993	1.033
NOVARA	1.044	1.029	1.074	REGGIO CALABRIA	1.036	0.968	1.003
LOCRI	1.040	1.029	1.069	LECCE	1.032	0.966	0.996
PATTI	1.039	1.029	1.069	LOCRI	1.032	0.988	1.019
BIELLA	1.033	1.029	1.062	FIRENZE	1.030	0.961	0.990
BELLUNO	1.029	1.029	1.058	BRINDISI	1.029	0.963	0.991
PISTOIA	1.027	1.029	1.057	NAPOLI	1.029	0.967	0.995
PAOLA	1.025	1.029	1.054	BUSTO ARSIZIO	1.029	0.973	1.001
SONDRIO	1.025	1.036	1.061	BELLUNO	1.025	0.992	1.017
SIENA	1.022	1.029	1.051	PORDENONE	1.024	0.972	0.995
GELA	1.019	1.029	1.049	AOSTA	1.024	0.998	1.022

Note that the difference between the two models is in the specification of inputs, so different results in productivity are due to different returns. Let us consider, for instance, the court of Pordenone. If we analyze Model B (i.e., with additional economic input), we can see that this court was not able to effectively exploit its financial resources to manage cases (i.e., $TFP = 0.995$); however, if we focus on human resources (i.e., Model A), its productivity increased over time (i.e., $TFP = 1.042$). At first glance, this may seem contradictory, but the results must actually be read together. On the one hand, TFP indicates that employees in the court of Pordenone work productively (i.e., specialization economies) but, on the other hand, clear difficulties are detected in the management of fees, allowances and costs, that is to say, in the internal management of judicial expenditure. The well-known decomposition of the Malmquist indexes into efficiency and technology components provides a more detailed analysis of these results.

In all of the cases in which there was an increase in productivity (i.e., Malmquist index > 1), the prevalent component was efficiency change, which implies that a reduction in inputs often determined a better use of the remaining resources. It is significant that, also in the model with additional economic inputs (i.e., Model B), the efficiency change of some courts was higher than 1, suggesting that, although their overall productivity decreased, they were able to reach the same productivity levels as in 2014 exploiting the resources available in 2018 and leaving the other conditions at time 2014 unchanged. On the other hand, the technology component represents the shift from the frontier in 2014 to that in 2018 and, considering the geometric mean of the whole sample, the results suggest progress for Model A (i.e., $tech = 1.03$) and regress for Model B (i.e., $tech = 0.98$). The same conclusions are confirmed when looking at the top 10% DMUs presented in Table 6. Nonetheless, Table 6 shows that, as for Model A, only Campobasso and L'Aquila had technical regress (i.e., $tech < 1$) in the period considered but, at the same time, these are the only two observations showing progress in Model B. Although these results might depend on the additional inputs, only a qualitative investigation would reveal whether significant organizational changes have been adopted in these two DMUs.

5.3. Second stage: determinants of inefficiency

According to Simar and Wilson (2007), we employ several multivariate truncated regression models to investigate the determinants of judicial technical inefficiency. The proposed upper level for the truncation is equal to 1, meaning that we only analyse the inefficient DMUs (i.e., TE scores < 1). Additionally, in order to collect more robust results, we use the bootstrap option with 2,000 replacements. We start with the TE scores estimated with human resources as input (Model A), then we move to the TE scores estimated with judicial expenditure as additional input (Model B). Note that the variables “2014”, “Other civil procedures” and “Islands” are omitted from the models, i.e., they are the variables against which the models are assessed. Table 7 presents the results of the regression models.

As evidenced by our results, the composition of the supply of justice can have a significant impact on the efficiency of first instance courts when considering civil matters. Indeed, we can observe a statistically significant relation between TE scores and percentage of cases in “Ordinary civil procedures”, “Special civil procedures” and “Voluntary civil process”. *Ceteris paribus*, if the percentage of cases in the case matters

Table 7. Multivariate truncated regression models – bootstrap option (2,000 replacements) and upper level (1) (Italy, 2014–2018).

VARIABLE	Model A		Model B	
	eq1	Sigma	eq1	sigma
Criminal procedures	0.00941 (0.0177)		0.00848 (0.0174)	
Mortgage foreclosure	−0.0142 (0.0288)		−0.0248 (0.0278)	
Bankruptcy	0.0860 (0.164)		0.118 (0.159)	
Labour and pension	4.84e-05 (0.0255)		0.00133 (0.0261)	
Ordinary civil procedures	−0.206*** (0.0387)		−0.198*** (0.0381)	
Special civil procedures	0.231*** (0.0346)		0.242*** (0.0338)	
Voluntary civil process	−0.0841*** (0.0252)		−0.0511* (0.0261)	
North West	0.0172*** (0.00336)		0.0172*** (0.00340)	
North East	0.0264*** (0.00346)		0.0254*** (0.00341)	
Centre	0.0265*** (0.00324)		0.0238*** (0.00323)	
South	0.0183*** (0.00296)		0.0198*** (0.00304)	
Constant	0.942*** (0.0144)	0.0194*** (0.000564)	0.939*** (0.0140)	0.0195*** (0.000591)
Wald chi2(15)	301.88		303.56	
Prob > chi2	0.0000		0.0000	
Year (FE)	Yes		Yes	
Observations	695		695	

Bootstrapped standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

“Ordinary civil procedures” and/or “Voluntary civil process” increases, we can expect the technical efficiency of our DMUs to decrease, considering both Model A and Model B. If we look at “Special civil procedures”, we find a statistically significant positive coefficient, meaning that, when the percentage of cases in this case matter increases, we can expect a positive impact on the efficiency of our DMUs. Nevertheless, if we look at the other case matters (i.e., criminal, insolvency, labour and pension), we cannot observe statistically significant coefficients (i.e., p -values > 0.1).

Accordingly, we cannot reject the hypothesis that the composition of workload can directly affect the performance of judicial courts, influencing their technical ability to produce justice. In other words, the courts’ TE scores and their performance might depend on a higher number of cases settled in categories characterized by lower expected settlement times rather than on better internal management of human and financial resources. Accordingly, what might make the difference in increasing the performance of the DMUs is the demand for justice, instead of good managerial practices. Indeed, this result might be due to the specific procedures that judges have to apply in their institutional activity of enforcing the law in order to supply justice, which represent the technology of these production processes. In other words, by changing the civil and criminal codes that define the rules applied to evaluate judicial cases in certain case matters, the policy makers could have the opportunity to improve the technology of these

production processes, thus increasing the technical efficiency of courts. Nevertheless, not all the coefficients are statistically significant, offering no clear evidence regarding how to improve these specific technologies.

Finally, let us analyse the exogenous variables, i.e., the geographical macro areas. Considering both Model A and Model B, the data reveal a statistically significant relation between the areas in which the DMUs are located and their technical efficiency. These variables represent the socio-economic circumstances in which our judicial districts operate, e.g., the litigiousness of citizens who are resident in these areas and/or local rates of criminality, as well as how difficult it is for the Ministry of Justice to enrol judges to fill vacancies in these judicial courts. For example, the North of Italy is characterized by higher economic performance and income, less litigiousness and criminality, and limited vacancy problems; yet, moving to the Islands macro area, the situation is completely different, with lower economic performance and income, as well as higher levels of litigiousness, criminality and vacancies. Note that, according to our results, the macro area benchmark is “Islands”, i.e., all coefficients are positive and they suggest an increase in inefficiency moving from a court located in this macro area to another. This might be ascribed to the larger numbers of judge vacancies, i.e., the significantly lower inputs used in the production process, even though the demand for justice is extremely high.

6. Conclusions

This manuscript compares two different model definitions, which are able to estimate a judicial TE score for every first instance court, benchmarking DMUs and highlighting the main drivers of inefficiency. On the one hand, we have the classic model definition with human resources as input of this production process; on the other hand, we have a model definition with both human resources and judicial expenditure as inputs of the same process. Based on our results and considering some key proxies identified by the Italian Ministry of Justice, we cannot reject the hypothesis that the latter model definition is more appropriate than the former; and we cannot reject the hypothesis that incomplete model definitions might mislead policy makers in reforming the national justice system and in keeping public expenditure under control. Furthermore, our investigation represents a valid scientific basis allowing for the interests of the stakeholders to converge, and possibly generating wide consensus around the proposed policy reform.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

R. Ippoliti  <http://orcid.org/0000-0001-9199-452X>

References

- Agrell, P. J., Mattsson, P., & Månsson, J. (2019). Impacts on efficiency of merging the Swedish district courts. *Annals of Operations Research*, 1–27.
- Badunenko, O., & Kumbhakar, S. C. (2017). Economies of scale, technical change and persistent and time-varying cost efficiency in Indian banking: Do ownership, regulation and heterogeneity matter? *European Journal of Operational Research*, 260(2), 789–803.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
- Christensen, R. K., & Szmer, J. (2012). Examining the efficiency of the US courts of appeals: Pathologies and prescriptions. *International Review of Law and Economics*, 32(1), 30–37.
- Coelli, T. (1996). A guide to DEAP Version 2.1: A data envelopment analysis (computer) program. *CEPA working paper*, 96/08.
- Coelli, T., Rao Prasada, D. S., & Battese, G. E. (1998). *An introduction to efficiency and productivity analysis*. Norwell, MA, US: Kluwer Academic Publishers.
- Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *A comprehensive text with models, applications, references and DEA-solver software*. Springer Science+ Business Media.
- Dakolias, M. (1999). *Court performance around the world: A comparative perspective*. The World Bank.
- Daraio, C., & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: A probabilistic approach. *Journal of Productivity Analysis*, 24(1), 93–121.
- Daraio, C., & Simar, L. (2007a). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and application*. Berlin, Germany: Springer.
- Daraio, C., & Simar, L. (2007b). Conditional nonparametric frontier models for convex and nonconvex technologies: A unifying approach. *Journal of Productivity Analysis*, 28(1–2), 13–32.
- Daraio, C., & Simar, L. (2016). Efficiency and benchmarking with directional distances: A data-driven approach. *Journal of the Operational Research Society*, 67(7), 928–944.
- Daraio, C., Simar, L., & Wilson, P. W. (2015). *Testing the “separability” condition in two-stage nonparametric models of production* (No. 2015/21). LEM working paper series.
- Daraio, C., Simar, L., & Wilson, P. W. (2018). Central limit theorems for conditional efficiency measures and tests of the “separability” condition in nonparametric, two-stage models of production. *Econometrics Journal*, 21(2), 170–191.
- Deyneli, 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(3), 477–493.
- Falavigna, G., & Ippoliti, R. (2021). Reform policy to increase the judicial efficiency in Italy: The opportunity offered by EU post-Covid funds. *Journal of Policy Modeling*, 43(5), 923–943.
- Falavigna, G., Ippoliti, R., & Manello, A. (2019). Judicial efficiency and immigrant entrepreneurs. *Journal of Small Business Management*, 57(2), 421–449.
- Falavigna, G., Ippoliti, R., & Ramello, G. B. (2018). DEA-based Malmquist productivity indexes for understanding courts reform. *Socio-Economic Planning Sciences*, 62, 31–43.
- Färe, R., & Grosskopf, S. (1996). *Intertemporal production frontiers: With dynamic DEA*. Boston: Kluwer Academic Publishers.
- Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1992). Productivity changes in Swedish pharmacies 1980–1989: A non-parametric Malmquist approach. *Journal of Productivity Analysis*, 3(1–2), 85–101.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 120(3), 253–290.
- Ferrandino, J. (2012). The impact of revision 7 on the technical efficiency of Florida’s circuit courts. *Justice System Journal*, 33(1), 22–46.
- Ferro, G., Oubiña, V., & Romero, C. (2020). Benchmarking labor courts: An efficiency frontier analysis. *International Journal for Court Administration*, 11(2), 7.

- Ferro, G., Romero, C. A., & Romero-Gómez, E. (2018). Efficient courts? A frontier performance assessment. *Benchmarking: An International Journal*, 25(9), 3443–3458.
- Finocchiaro Castro, M., & Guccio, C. (2015). Bottlenecks or Inefficiency? An assessment of first instance Italian courts' performance. *Review of Law & Economics*, 11(2), 317–354.
- Finocchiaro Castro, M., & Guccio, C. (2018). Measuring potential efficiency gains from mergers of Italian first instance courts through nonparametric model. *Public Finance Review*, 46(1), 83–116.
- Giacalone, M., Nissi, E., & Cusatelli, C. (2020). Dynamic efficiency evaluation of Italian judicial system using DEA based malmquist productivity indexes. *Socio-Economic Planning Sciences*, 100952.
- Giacomelli, S., & Menon, C. (2017). Does weak contract enforcement affect firm size? Evidence from the neighbour's court. *Journal of Economic Geography*, 17(6), 1251–1282.
- Gitto, S. (2017). Efficiency change, technological change and capital accumulation in Italian regions: A sectoral study. *International Review of Applied Economics*, 31(2), 191–207.
- Ippoliti, R., & Tria, G. (2020). Efficiency of judicial systems: Model definition and output estimation. *Journal of Applied Economics*, 23(1), 385–408.
- Isik, I., & Hassan, M. K. (2003). Financial deregulation and total factor productivity change: An empirical study of Turkish commercial banks. *Journal of Banking & Finance*, 27(8), 1455–1485.
- Kittelsen, S. A., & Førsund, F. R. (1992). Efficiency analysis of Norwegian district courts. *Journal of Productivity Analysis*, 3(3), 277–306.
- Kneip, A., Simar, L., & Wilson, P. W. (2016). Testing hypotheses in nonparametric models of production. *Journal of Business and Economic Statistics*, 34(3), 435–456.
- Lovell, C. K. (2003). The decomposition of Malmquist productivity indexes. *Journal of Productivity Analysis*, 20(3), 437–458.
- Mattsson, P., & Tidå, C. (2018). Potential efficiency effects of merging the Swedish district courts. *Socio-Economic Planning Sciences*. Forthcoming.
- Mussard, S., & Peypoch, N. (2006). On multi-decomposition of the aggregate Malmquist productivity index. *Economics Letters*, 91(3), 436–443.
- Örkényi, L. (2021). Duration of court events required for litigation—an empirical study to provide a theoretical basis for an objective measurement system for judicial workload. *International Journal for Court Administration*, 12(1). doi:10.36745/ijca.370
- Peyrache, A., & Zago, A. (2016). Large courts, small justice!: The inefficiency and the optimal structure of the Italian justice sector. *Omega*, 64, 42–56.
- Ray, S. C., & Desli, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: Comment. *American Economic Review*, 87, 1033–1039.
- 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.
- Silva, M. C. A. (2018). Output-specific inputs in DEA: An application to courts of justice in Portugal. *Omega*, 79, 43–53.
- Simar, L., & Wilson, P. W. (1999). Estimating and bootstrapping Malmquist indices. *European Journal of Operational Research*, 115(3), 459–471.
- Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1), 31–64.
- Simar, L., & Wilson, P. W. (2020). Hypothesis testing in nonparametric models of production using multiple sample splits. *Journal of Productivity Analysis*, 53(3), 287–303.
- Tone, K. (2004). Malmquist productivity index. In *Handbook on data envelopment analysis* (pp. 203–227). Boston, MA: Springer.
- Wheelock, D. C., & Wilson, P. W. (1999). Technical progress, inefficiency, and productivity change in U.S. banking, 1984–1993. *Journal of Money, Credit, and Banking*, 31(2), 212–234.
- Wilson, P. W. (2008). FEAR: A software package for frontier efficiency analysis with R. *Socio-economic Planning Sciences*, 42(4), 247–254.

Annex A

Tables A1 and A2 report the results of the tests on the separability conditions, considering both the CRS and the VRS assumptions. Columns 2 and 3 show the test statistics obtained by averaging the Daraio et al. (2018) statistics across 10 sample splits (i.e., tau1) and the Kolmogorov-Smirnov statistic tau (i.e., tau2). The p-value columns (i.e., p-value1 and p-value2) display the corresponding p-values estimated using the bootstrap method described by Simar and Wilson (2020). According to these results, we cannot reject the null hypotheses ($p\text{-value} > 0.05$), i.e., we can adopt the case matters as regressors in the second stage.

Table A3 reports the results of the adopted tests to verify whether VRS or CRS is more appropriate. Columns 2 and 3 show the test statistics obtained by averaging the Kneip et al. (2016) statistics across 10 sample splits (i.e., tau1) and the Kolmogorov-Smirnov statistic (tau2). The respective p-values (i.e., p-value1 and p-value2) suggest rejecting the null hypotheses ($p\text{-value} < 0.05$), which points to the adoption of VRS.

Table A1. Separability conditions (CRS).

Year	tau1	tau2	p-value1	p-value2
Model A				
2014	−0.438	0.353	0.786	0.407
2015	−1.238	0.378	0.995	0.302
2016	0.463	0.510	0.098	0.111
2017	−0.818	0.387	0.942	0.296
2018	−0.472	0.353	0.829	0.372
Model B				
2014	0.172	0.275	0.368	0.624
2015	−0.034	0.334	0.560	0.411
2016	0.674	0.331	0.080	0.409
2017	−0.433	0.234	0.823	0.738
2018	0.426	0.561	0.099	0.061

Table A2. Separability conditions (VRS).

Year	tau1	tau2	p-value1	p-value2
Model A				
2014	0.335	0.299	0.231	0.525
2015	0.069	0.189	0.450	0.873
2016	0.604	0.390	0.091	0.277
2017	0.477	0.545	0.173	0.067
2018	−0.874	0.525	0.961	0.079
Model B				
2014	0.388	0.355	0.203	0.377
2015	0.509	0.475	0.160	0.175
2016	−0.158	0.155	0.674	0.965
2017	−0.761	0.328	0.941	0.479
2018	0.718	0.543	0.061	0.081

Table A3. Tests on the returns to scale assumptions (CRS vs VRS).

Year	tau1	tau2	p-value1	p-value2	RTS
Model A					
2014	22.071	1.000	0.000	0.002	VRS
2015	23.902	1.000	0.000	0.000	VRS
2016	19.685	1.000	0.000	0.000	VRS
2017	19.531	1.000	0.000	0.000	VRS
2018	21.342	1.000	0.000	0.000	VRS
Model B					
2014	5.014	1.000	0.000	0.000	VRS
2015	4.878	0.996	0.000	0.000	VRS
2016	5.925	0.999	0.000	0.000	VRS
2017	5.078	0.954	0.000	0.001	VRS
2018	8.338	1.000	0.000	0.000	VRS

Annex B

Tables A4 reports the TE scores with and without caseload, i.e., considering as additional inputs the incoming and pending cases at the beginning of every year. According to results, on average, the “North-East” represents the benchmark to lead a policy reform in all models, while the “Islands” is the geographical macro area with the lowest TE scores. However, considering the caseload, we can observe how the efficiency gap between the North and the South of Italy decreases.

Table A4. Sensitivity analysis: Models A and B without caseloads vs. Models A and B with caseloads.

Geographical macro area & Districts of second instance	Model A		Model B	
	Without caseload	With caseload	Without caseload	With caseload
North West	0.9555	0.9838	0.9604	0.9832
Brescia	0.9728	0.9881	0.9795	0.9868
Genova	0.9342	0.9801	0.9317	0.9802
Milan	0.9594	0.9802	0.9706	0.9803
Turin	0.9558	0.9872	0.9579	0.9859
North East	0.9642	0.9867	0.9666	0.9848
Bologna	0.9706	0.9871	0.9735	0.9863
Trento	0.9439	0.9880	0.9643	0.9861
Trieste	0.9547	0.9883	0.9554	0.9876
Venetia	0.9709	0.9847	0.9660	0.9810
Center	0.9593	0.9764	0.9595	0.9763
Ancona	0.9452	0.9780	0.9473	0.9779
Bologna	0.9506	0.9814	0.9667	0.9790
Florence	0.9599	0.9746	0.9612	0.9751
Perugia	0.9633	0.9773	0.9541	0.9778
Roma	0.9676	0.9764	0.9670	0.9756
South	0.9416	0.9647	0.9427	0.9656
Bari	0.9796	0.9813	0.9765	0.9795
Campobasso	0.9273	0.9598	0.9324	0.9606
Catanzaro	0.9262	0.9497	0.9292	0.9509
L'Aquila	0.9405	0.9783	0.9409	0.9794
Lecce	0.9611	0.9701	0.9607	0.9707
Naples	0.9512	0.9676	0.9534	0.9701
Potenza	0.9070	0.9452	0.9095	0.9453
Reggio Calabria	0.9270	0.9577	0.9260	0.9574
Salerno	0.9666	0.9676	0.9661	0.9688
Islands	0.9269	0.9610	0.9287	0.9617
Cagliari	0.9174	0.9612	0.9168	0.9624
Caltanissetta	0.8927	0.9545	0.8925	0.9543
Catania	0.9450	0.9568	0.9462	0.9565
Messina	0.9465	0.9540	0.9510	0.9573
Palermo	0.9316	0.9704	0.9361	0.9703
Italy	0.9492	0.9738	0.9512	0.9737