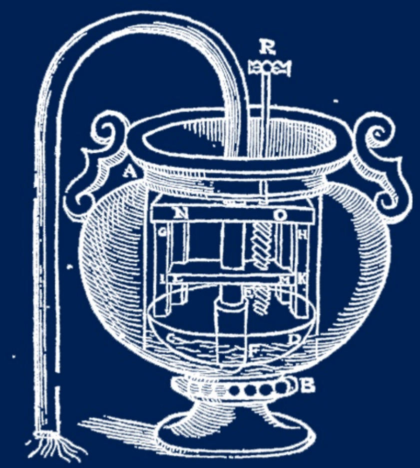
CNR-IRCrES Working Paper

Institutional efficiency and budget constraints: a Directional Distance Function approach to lead a key policy reform



6/2021

Greta Falavigna
Roberto Ippoliti

Consiglio Nazionale delle Ricerche
IRCrES

ISSN (online) 2421-7158

Direttore Emanuela Reale

Direzione CNR-IRCrES
Istituto di Ricerca sulla Crescita Economica Sostenibile
Via Real Collegio 30, 10024 Moncalieri (Torino), Italy
Tel. +39 011 6824911 / Fax +39 011 6824966
segreteria@ircres.cnr.it
www.ircres.cnr.it

Sede di Roma Via dei Taurini 19, 00185 Roma, Italy
Tel. +39 06 49937809 / Fax +39 06 49937808

Sede di Milano Via Corti 12, 20121 Milano, Italy
Tel. +39 02 23699501 / Fax +39 02 23699530

Sede di Genova Corso Ferdinando Maria Perrone 24, 16152-Genova, Italy
Tel. +39 010 6598798

redazione@ircres.cnr.it
www.ircres.cnr.it/index.php/it/produzione-scientifica/pubblicazioni

CNR-IRCrES Working Paper 6/2021

Institutional efficiency and budget constraints: a Directional Distance Function approach to lead a key policy reform

GRETA FALAVIGNA*, ROBERTO IPPOLITI**

* CNR-IRCrES, Consiglio Nazionale delle Ricerche, Istituto di Ricerca sulla Crescita Economica Sostenibile, Italia
** Faculty of Business Administration and Economics, Bielefeld University, Bielefeld (North Rhine-Westphalia) – Germany

Corresponding author: greta.falavigna@ircres.cnr.it

ABSTRACT
This manuscript focuses on the Italian judicial system and on how to shape a policy reform aimed at increasing court efficiency, taking the financial negative externalities generated by this production process into account. On the one hand, the authors identify the benchmarks and main drivers of judicial inefficiency, while, on the other hand, they show how incorrect model definition may mislead policy makers tackling this reform process, based on an analysis of the Directional Distance Function with and without bad outputs. According to the results, incorrect model definition causes a type I error equal to 10.37% and a type II error equal to 3.66%. Policy implications concern the opportunity to adopt the proposed model and the collected benchmarks to reform the judicial system, improving its technical efficiency and maintaining the public budget under control.

KEYWORDS: Directional Distance Function; Institutional efficiency; Budget constraint; Policy reform

DOI: 10.23760/2421-7158.2021.006

HOW TO CITE THIS ARTICLE
Falavigna G., Ippoliti R. (2021). *Institutional efficiency and budget constraints: a Directional Distance Function approach to lead a key policy reform* (CNR-IRCrES Working Paper 6/2021). Istituto di Ricerca sulla Crescita Economica Sostenibile. Disponibile da http://dx.doi.org/10.23760/2421-7158.2021.006

INDICE

1. Introduction

The recent global financial crisis has led to an increase in public debt across the world (Kumar and Woo, 2015), with expected negative impacts on long-term economic growth and welfare (Checherita-Westphal and Rother, 2012; Chudik et al., 2017), as well an intertemporal transfer of income and conflicts over fiscal policy among different generations (Arai et al., 2018; Barseghyan and Battaglini, 2016; Müller et al., 2016). In order to avoid these negative outcomes and support fiscal sustainability, the European Union (EU) member states adopted fiscal rules to limit their debt, i.e., the Maastricht Treaty convergence criteria, which aimed to keep public debt within 60% of GDP and government deficit within 3% of GDP (De Grauwe, 1996). Accordingly, several EU countries initiated comprehensive reforms of their national public systems to collect resources for social interventions during the crisis and, among them, Italy is one of the most interesting case studies. Indeed, Italy's public debt is huge and, as a consequence, the country is in even greater need of key structural reforms, both to reduce its debt and to gather additional resources to support welfare interventions at this critical moment. However, what principles should guide such policy reforms?

Policy makers are tasked with correctly identifying the main determinants 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 main 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 clear 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.

In this work, we focus on a specific case study, i.e., the Italian judicial system, and on how to shape a policy reform aimed at increasing the efficiency of the courts, considering the additional variable costs created by the production of this specific good (i.e., the justice). On the one hand, we identify the main drivers of inefficiency, while, on the other hand, we show how model definition may mislead policy makers in the reform process.

The paper is organized as follows. Section 2 presents an overview of the current literature on judicial efficiency, describing the main models adopted by academia in its estimation. Then, assuming that policy makers are interested not only in supplying justice but also in minimizing public expenditure, we identify the key elements in the model definition for the benchmarking of judicial courts. Section 3 sets out the methodology adopted to estimate *judicial* technical efficiency scores (i.e., Directional Distance Function, with and without bad outputs), and applies it to the case study under investigation (i.e., the Italian judicial system). Finally, Section 4 illustrates the results of the empirical analysis, while Section 5 offers some conclusions that may have particular significance from the policy makers' point of view.

2. THEORETICAL BACKGROUND: JUDICIAL EFFICIENCY AND BUDGET CONSTRAINTS

Several approaches have been proposed to measure judicial efficiency: the time needed to settle a case (e.g., Mitsopoulos and Pelagidis, 2007; Christensen and Szmer, 2012), the number of cases completed by a court (e.g., Beenstock and Haitovsky, 2004; Ramseyer, 2012); technical efficiency scores (e.g., Santos and Amado, 2014; Ippoliti *et al*., 2015a) and clearance rates (e.g., Buscaglia & Ulen, 1997; Dakolias, 1999). The methodologies applied in the benchmarking analysis range from multivariate regression models (e.g. Beenstock and Haitovsky, 2004) to mathematical programming techniques, like Free Disposal Hull (e.g., Tulkens, 1993), Directional Distance Function (e.g., Falavigna et al., 2015), Data Envelopment Analysis (e.g., Schneider, 2005), and Malmquist indexes (e.g., Mattsson et al., 2018). Obviously, the adoption of a specific approach depends on the expected target and on the main stakeholders interested in analyzing the judicial system. On the one hand, society may be mostly concerned with the time needed to settle a judicial case, that is to say, how much time one has to wait for justice to be served. On the other hand, in order to improve the supply of justice, policy makers are likely to focus on the productivity of judges and the performance of courts, which means that they have to use clear benchmarks, including, for example, simple indexes (i.e., clearance rates) or more sophisticated ones (i.e., technical efficiency scores). This allows them to stratify courts and, accordingly, implement a reform based on the most efficient organisational structure identified.

The clearance rate has been proposed by the European Commission for the Efficiency of Justice (CEPEJ) as a valid criterion to compare national judicial systems across the EU, and it has the advantage of being easily calculated without large amounts of information. This index is equal to the number of resolved cases divided by the number of incoming cases, and it represents the ability of the national system to satisfy the demand for justice[1]. However, it cannot be used as a leading measure to compare courts within a specific national system, since it does not consider vital information such as, for example, the number of judges involved in this specific production process. For this reason, in the last few years, academics have proposed the estimation of technical efficiency scores through the adoption of mathematical programming techniques, among which Data Envelopment Analysis (DEA), Free Disposal Hull (FDH), and Directional Distance Function (DDF).[2] In particular, DEA has been successfully applied to the analysis of national judicial systems, both in its one-stage form (e.g. Kittelsen and Førsund, 1992; Pedraja-Chaparro and Salinas-Jimenez, 1996) and in its two-stage form (Schneider, 2005; Deyneli, 2012)[3].

Following Ippoliti and Tria (2020), Table 1 presents a review of the current literature, showing the inputs and outputs adopted, as well as the judicial systems analyzed and the mathematical programming techniques used. As the readers can easily see, the number of settled cases is identified as the main output, although this is usually presented as an aggregate measure. Only few studies have tried to adopt a more precise output measure by disaggregating the supply of justice according to case matters (i.e., Kittelsen and Førsund, 1992; Santos and Amado, 2014). At the same time, even greater heterogeneity can be observed when the inputs are considered too. Some authors have exclusively used judges and staff as research inputs (e.g., Pedraja-Chaparro and Salinas-Jimenez, 1996; Deyneli, 2012), whereas other researchers have included both pending and/or incoming cases (e.g., Schneider, 2005; Falavigna *et al*., 2015; Castro Finocchiaro and Guccio, 2014; Ippoliti *et al*., 2015a, 2015b), suggesting that the demand for justice might affect court productivity. Yet, the model definition should be based on the main targets pursued by policy makers. In other words, if the expected outcome is to increase the supply of justice while keeping public expenditure under control, are these model definitions correct? Is

[1] An index equal to 1 means that the system is able to satisfy the demand for justice (incoming cases), while an index lower than 1 means that the system backlog increases (pending cases at year end). Obviously, an index higher than 1 means that the system is able to satisfy the demand for justice (incoming cases) and, at the same time, to reduce its current backlog (pending cases at year start).

[2] These models have been adopted to investigate the performance of public institutions (e.g., Valdmanis, 1992; Bjurek and Hjalmarsson, 1995; Anderson et al., 2011).

[3] According to Simar and Wilson (2007), the one-stage DEA procedure aims to estimate technical efficiency scores proposing relevant benchmarks, while the two-stage DEA procedure uses the estimated scores to study the determinants of inefficiency.

inappropriate model definition likely to hinder the reform process? Is it admissible to opt for a model definition that controls for the negative externalities caused by judicial procedures and affecting the public budget? These are exactly the research questions posed in this work.

In order to supply justice, courts incur additional variable costs, i.e., costs that are specifically linked to judicial production. Depending on the nature of each judicial case and its related civil/criminal procedure, the judges may need professionals to properly assess damage caused (e.g., forensic tests or psychiatric evaluations), technicians to evaluate evidence (e.g., handwriting analyses or ballistic reports), or lawyers to guarantee legal aid. All of these costs are variable, which means that they change according to the supply of justice and, obviously, the resources needed are affected by the internal organisation of every single court. In other words, with no demand for justice, none of these costs have to be borne; furthermore, significant differences exist across courts, based on the internal structure they have adopted. Consequently, if policy makers aim for a reduction in expenditure due to budget constraints, every court has two options: it can either re-organize itself to be more efficient or reduce the supply of justice to operate within the new financial constraints. Obviously, the former outcome, i.e., an improvement in court performance combined with a reduction in expenditure, is preferable and, considering the current age of austerity, it is also the outcome sought by policy makers. However, without clear efficiency benchmarks, the courts may be forced to adopt the second option, i.e., a reduction in the supply of justice, with a significant negative impact on both society and market dynamics (Giacomelli and Menon, 2017; García-Posada and Mora-Sanguinetti, 2015; Falavigna et al., 2019).

Table 1. Inputs and outputs adopted in mathematical programming techniques for the analysis of judicial efficiency

Study	Analyzed judicial system	Output	Input	Technical notes
Lewin *et al.* (1982)	USA - North Carolina (Criminal courts)	settled cases; pending cases (less than 90 days);	days of court held; number of district attorneys and assistants; size of the caseload; number of misdemeanours in the caseload and white population size;	DEA model
Kittelsen and Førsund (1992)	Norway (Civil and criminal courts)	settled cases (7 categories);	judges; staff;	DEA model
Tulkens (1993)	Belgium (Justices of the Peace)	settled cases (civil and commercial); settled cases (juvenile offences); family arbitration sessions held;	staff;	FDH model
Pedraja-Chaparro and Salinas-Jimenez (1996)	Spain (Administrative Litigation Division of High Courts)	settled cases (with sentence); settled cases (in other ways, i.e. without sentence);	judges; staff;	DEA model
Schneider (2005) †	Germany (Labour Courts)	settled cases; published decisions;	judges; pending cases;	DEA model
Yeung and Azevedo (2011)	Brazil (Civil courts)	settled cases at first level (over workload); settled cases at second level (over workload);	judges over workload; staff over workload;	DEA model
Ferrandino (2012)	USA - Florida (Criminal, civil and family courts)	settled cases;	judges;	DEA model
Deyneli (2012)	Europe (Civil and criminal courts)	settled cases (civil); settled cases (criminal); population;	judges; staff;	DEA model
Castro Finocchiaro and Guccio (2014)	Italy (Civil courts)	aggregated settled cases (with sentence); aggregated settled cases (in other ways, i.e. without sentence);	judges; staff; pending cases;	DEA model
Ippoliti et al. (2015a, 2015b)	Europe (Civil courts)	settled cases;	judges; staff; pending cases; incoming cases;	DEA model

Santos and Amado (2014)	Portugal (Civil and criminal courts)	settled cases according to proceedings (43 outputs);	judges; staff;	DEA model
Castro Finocchiaro and Guccio (2015, 2018)	Italy (Civil and criminal courts)	settled cases;	judges; staff; workload (i.e. pending plus incoming cases);	DEA model
Falavigna et al. (2015)	Italy (Tax courts)	settled cases (good output); delay (bad output);	judges; pending cases; incoming cases;	DDF model
Melcarne and Ramello (2016)	Europe (Civil courts)	settled cases;	judges; staff; pending cases; incoming cases;	DEA model
Peyrache and Zago (2016)	Italy (Civil and criminal courts)	settled cases;	judges; staff; pending cases;	DEA model
Falavigna et al. (2018)	Italy (Tax courts)	settled cases;	judges; pending cases; incoming cases;	DEA model and Malmquist Index
Silva (2018)	Portugal (civil courts)	settled cases (three categories);	number of administrative workers; incoming cases (three categories);	DEA model
Mattsson and Tidanå (2019)	Sweden (Civil and criminal courts)	settled criminal cases; settled civil cases; settled matters;	judges; law clerks; other personnel; area of the court (square meters);	DEA model
Mattsson et al. (2018)	Sweden (Civil and criminal courts)	settled criminal cases; settled civil cases; settled matters;	judges; law clerks; other personnel; area of the court (square meters);	DEA model and Malmquist Index
Falavigna et al. (2019)	Italy (Civil and criminal courts)	settled cases (civil cases); settled cases (criminal cases);	judges; staff; pending cases; incoming cases;	DEA model

Note: workload is defined as incoming plus pending cases at the beginning of every year.

Based on the proposed framework, if we are interested in benchmarking courts and in keeping public expenditure under control, we need to consider the variable costs generated by this production process, which can be regarded as a financial negative externality for society. Indeed, judicial production costs are directly covered by the public budget, subtracting key resources from other sectors (e.g., health care or welfare) or causing the government to introduce new taxes. Obviously, there is also a third way, which revolves around the introduction of new judicial procedures able to reduce the additional costs arising from the resolution of cases. Nevertheless, the latter approach requires key benchmarks to correctly estimate the determinants of judicial inefficiency. This is possible through the aforementioned two-stage approach but, as Table 1 clearly shows, none of the studies considered have integrated variable costs into their analysis. Therefore, to the best of our knowledge, there are no model definitions that adopt variable costs as negative bad outputs of the judicial production process. This is a vital issue since, depending on the model definition and the policy makers' main targets, incorrect reference values could cause the structural reform of the national judicial system to fail. Appropriate input selection and output definition are crucial and, considering the current gap in literature, the most important concern is to identify the best model definition and mathematical programming technique to obtain key benchmarks. Accordingly, we propose an innovative model definition that could be better suited to the current age of austerity and related policies regarding budget constraints. Focusing on our specific case study (i.e., the Italian judicial system), we test the following hypothesis:

H_1 technical efficiency scores estimated without considering the negative externalities affecting the public budget can mislead policy makers.

The following section presents the model definition and methodology adopted, as well as our specific case study (i.e., the Italian judicial system).

3. METHODOLOGY AND DATA

By using the Directional Distance Function (DDF), we test the first hypothesis through a comparative analysis of two different model definitions with different input-output space. In detail, we compare scores estimated through DDF without bad outputs and through DDF with bad outputs, to determine whether adopting additional costs as bad outputs in the justice production process might actually make a difference.

In the first subsection, a simple description of the DDF meaning in comparison to DEA is presented, while in the second one, data and statistics are discussed. Mathematical details on Directional Distance Function can be found in Appendix A.

3.1. From DEA to DDF: a simple conceptualization

In order to easily evaluate and compare the efficiency and/or productivity of observations in a sample, literature proposed to build a frontier on which efficient subjects are found. From the mathematical point of view, the problem is to solve a linear programming for each observation with the aim to obtain a score representing its ability in producing in the most economical way possible.

For instance, let we consider the pharmaceutical industry. Firms produce a drug to care patients with a specific pathology. The DEA methodology suggests to identify from the one hand the main factors needed for the production (i.e., inputs), and, from the other hand, the real quantity of drug produced (output). In this specific case, we can reasonably suppose that main production factors are two: the number of employees and the quantity of active principles; while the output is the quantity of drugs produced.

The DEA methodology allows to assign an efficient score to each firm of the pharmaceutical industry, giving the possibility to compare different performance and to highlight which activities are less efficiency. Clearly, the meaning of these efficiency scores is technical, because it considers only input necessary in the production process, without taking into consideration some internal organizational factors that can affect the production of the drug. One of the advantages of these methodologies is the possibility to set the problem following two orientations:

- Input-oriented: the linear program is defined in order to identify scores on the base of the ability of firms to minimize the use of inputs, taking equal produced output. In the case of pharmaceutical industry, firms on the frontier are able to produce the same quantity of drug, but using lower quantities of materials and lower number of employees.
- Output-oriented: the mathematical problem is defined in order to identify scores on the base of the ability of firms to maximize the production, taking inputs equal. Considering the pharmaceutical industry, firms on the frontier are able to produce the highest quantities of drugs, using the same production factors (inputs).

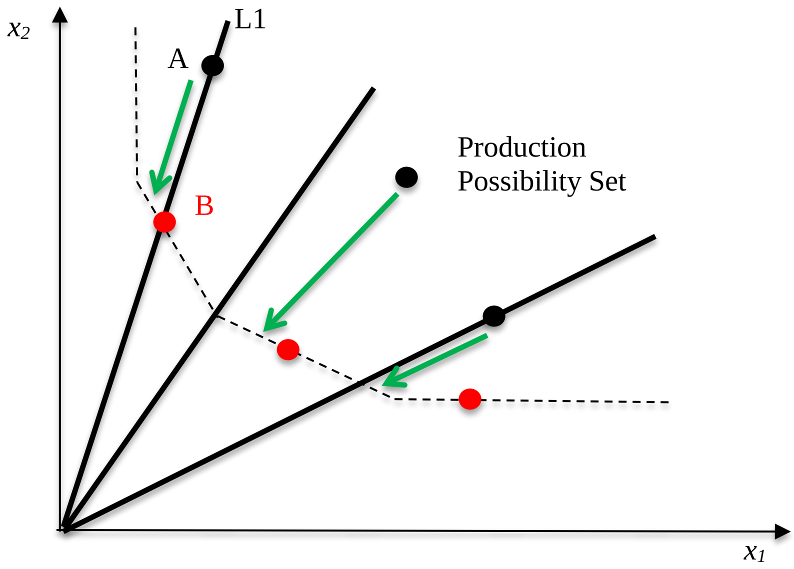
Figure 1(a) represents the conceptualization of the DEA methodology input-oriented. On the cartesian axes the two inputs are represented, while the dotted line depicts the best practice frontier, that is the representation of the most efficient firms (in the picture, they are the red dots). Indeed, these firms are able to produce the same quantity of drug but using the minimum amount of inputs. Let we consider firms A and B in the figure, the black line (L1) starts from the origin of axes, crosses the frontier in the point B and it goes on until it crosses point A. The black line represents the same combination of inputs: increasing the distance from the origin, also the amount of input required for the production increased. Firm B crosses the black line before than firm A, winning a position on the optimal frontier because firms A and B produce the same amount of drug, but firm B uses the lowest quantities of employees and active ingredient in comparison to other firms (non-efficient firms are depicted in black dots). Finally, the area above the frontier represents the production possibility set, that is to say all possible combinations of production.

Figure 2(b) describes the conceptualization of DEA output-oriented. Let we consider the same example of pharmaceutical industry, but, for simplifying the graphical visualization, firms produce two drugs (for instance, the same drug but in two different formats: syrup and tablets). Dotted line always represents the maximum combination of the two outputs that can be produced taking inputs equal. Firms on the frontier are more efficient because they produce the maximum quantity of drugs, using the same inputs of other firms. Let we consider now the black line L2, firms C and D use the same amount of materials and employees but, company C is able to produce more drugs than firm D. In this case, the production possibility set, is represented by the area under the frontier: all observations in this area are inefficient (black dots), while companies on the dotted line are efficient (red dots).

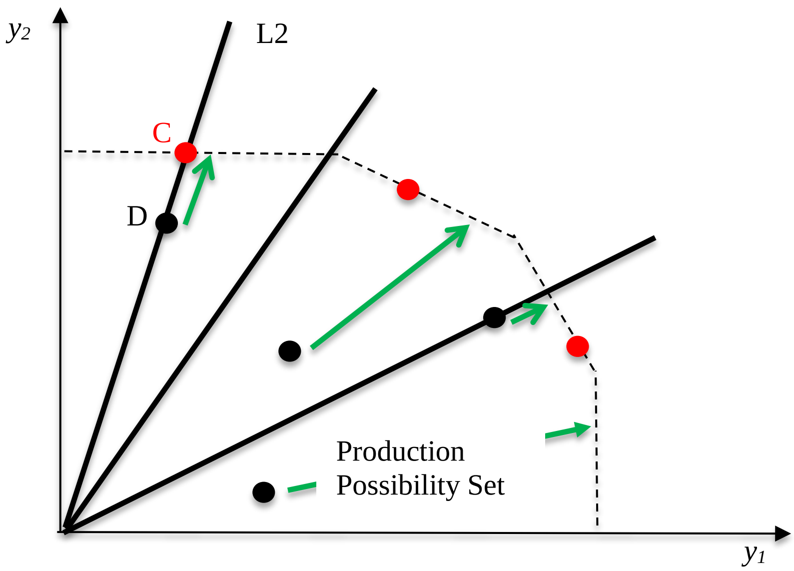
In both models (input and output oriented), the inefficiency can be measured calculating the radial distance (green arrows) between the observation (black dots) and the frontier (dotted line)[4].

Figure 1. DEA conceptualization

(a): input-orientation (x_1 and x_2)



(b): output-orientation (y_1 and y_2)



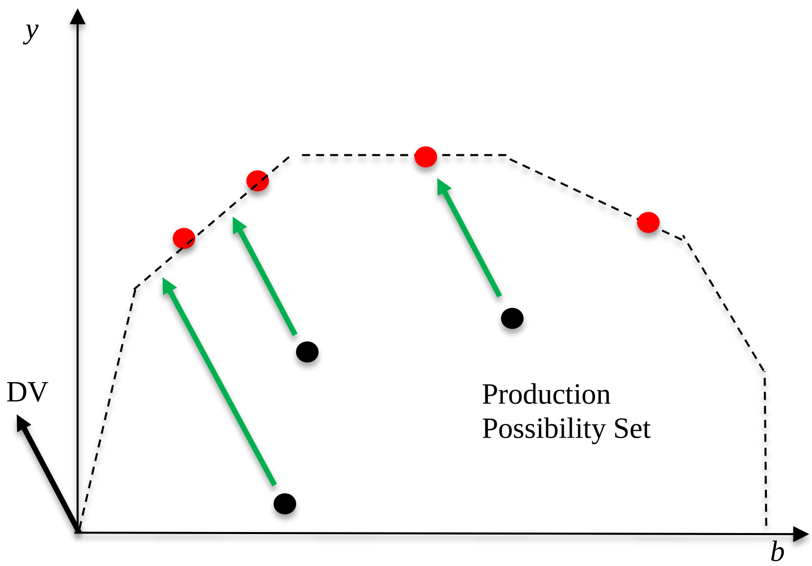
[4] The radial distance is the Euclidean distance between an observation and the origin of axes.

As explained in the Appendix A, the directional distance function (DDF) is a re-formulation of the linear programming of the DEA model, where a constraint is modified in order to give the possibility to consider different typologies of output.

Indeed, considering the pharmaceutical industry, together with drugs, also a negative externality is generated by the production process, that is the emission in the environment of pollutants. This is an output, exactly as drugs, but it is not desirable.

The directional (output) distance function allows to consider also this kind of production, that is called bad or undesirable. Figure 2 represents the production possibility set and the frontier of the DDF, where both good and bad outputs are considered (respectively, for pharmaceutical industry, drugs and pollutants). Red dots represent efficient firms that are able to maximize the good output minimizing the bad one and taking inputs equal. Efficient firms in the pharmaceutical industry are companies producing the maximum quantity of drugs, polluting less than other firms and using the same amounts of materials and the same number of employees. In figure 2, the production possibility set is represented by the area under the efficient frontier (dotted line). The main difference with the DEA model is how the inefficiency (that is the distance from the frontier) is calculated. If in the case of DEA, the inefficiency is calculated as the radial distance between the observation and the frontier, in the DDF model, it is necessary to define a "direction", that is a vector indicating in which manner the inefficiency (the distance of observation from the frontier) has to be calculated. In figure 2 the arrow DV represents this direction and green arrows show the measures of inefficiency for each firm.

Figure 2. DDF conceptualization (*y*: good output and *b*: bad output)



Even if the directional distance function has been proposed for considering together different typologies of outputs, the model can be applied even if only the good output is present. The meaning of the scores is the same of DEA model, but the computation of the inefficiency (the distance from the frontier) changes, because in the case of DDF the direction for reaching the frontier is defined by a pre-defined "direction" vector, while with the DEA, the direction is the radial distance.

In the present study, we present a new application of DDF to judiciary, where additional variable costs borne by courts to supply justice are considered as a bad output. The number of judges and the demand of criminal and civil justice are considered as inputs, whereas the good production is represented by the supply of criminal and civil justice. In details, we proposed two model definitions:

- in the first model, judicial systems are compared considering only the good output, i.e., the scores identify judicial systems on the base of their ability to maximize the supply of justice, taking equal the number of judges and the demand of justice (DDF TE score without Bad Outputs);
- in the second model, we consider also the costs of justice as bad output, i.e., the model compares judicial systems on the base of their ability to maximize the supply of justice, expending the minimum possible and using the same number of judges and demand of justice (DDF TE score with Bad Outputs).

3.2. The Italian judicial system: data and descriptive statistics

The Italian Ministry of Justice is in charge of administering civil and criminal justice, which is divided into two main tiers and one lowest level. At the lowest level are the so-called *Justices of the Peace* (i.e., *Giudici di Pace*), with specific civil and criminal competences. On a higher level, the first tier includes first instance courts (i.e., *Tribunali Ordinari*) which, gathering together the aforementioned justices of the peace, are part of the first instance districts (i.e., *Circondari Giudiziari*). In the period considered (i.e., between 2005 and 2010), there were 165 first instance districts, which represent the observations of our study. The second tier comprises 26 second instance districts (i.e. *Distretti di Corte di Appello)*, each with a variable number of first instance districts and responsible for appeals against first instance judgments. Finally, there is also a court of last resort (i.e. *Corte Suprema di Cassazione*), with seat in Rome and acting as the highest appellate court in all civil and criminal cases.

The most recent reform of the Italian judicial system, carried out in 2013 (*Law no. 148/2011* and *Decrees no. 155/2012* and no. *156/2012*), introduced a new judicial geography by redefining the territorial competence of first instance districts Indeed, the country's judicial geography went back to "*the unification of Italy"* in 1861, without significant changes for the last 150 years. The reform had two key targets. On the one hand, the policy makers aimed to lower public expenditure so as to realign the Italian system to the EU convergence criteria, reducing the number of judicial districts and their offices (i.e., courts and branch offices, as well as justices of the peace offices). On the other hand, based on the assumed specialisation economy of judges, the policy makers expected to increase the efficiency of the judicial system, merging the suppressed courts with others (i.e., increasing the size of the remaining courts). However, the territorial competence of the courts was redefined according to some national reference values identified after a long (and exhausting) political bargaining process, which strived above all to limit displeasure among voters and stakeholders and disregarded other reasons. Indeed, in selecting the courts to be suppressed, the policy makers did not rely on key efficiency reference values but exclusively considered size (e.g., number of citizens within the territory of each court) and administrative criteria (e.g., whether the seat of the court was in a province capital). At the end of this process, 31 courts and 220 branch offices were suppressed, along with 842 justices of the peace offices, merging their territorial competence with that of other courts (Ippoliti, 2015). This is a key example of wrong reference values adopted by policy makers and resulting in an ineffective reform. Without efficiency benchmarks able to guide the process, the end result was the merging of courts with wrong organisational models, decreasing the judicial efficiency of the districts and ultimately increasing the efficiency gap between the North and the South of Italy.

Table 2 proposes some descriptive statistics about inputs and outputs based on the selected case study and the aforementioned model definition. In detail, our data refer to civil and criminal justice and to first instance courts between 2005 and 2010 (see Figure S.2 in the Appendix B). The data were extracted from the databases of the Ministry of Justice and of the High Council of the Judiciary (i.e., *Consiglio Superiore della Magistratura*, or *CSM*). Comparing demand and supply, the level of inefficiency of these courts can easily be observed, since demand is twice as high as supply, meaning that the judicial system is not able to satisfy the demand for justice and that its backlog increases every year. Looking at the financial negative externalities, we can observe additional average costs for the supply of justice higher than 1 million euro, with a maximum value of 18 million euro. However, we can also observe a certain level of heterogeneity among the observations, which might be due to different internal organisational approaches (i.e., between statistics). At the same time, we can imagine an effective reorganisation process (i.e., within statistics), with reduced negative externalities impacting the public budget.

Table 3 shows some additional descriptive statistics regarding our sample, highlighting the average time needed to settle judicial cases. Specifically, the table shows the disaggregated supply of civil justice according to case matters and macro areas. Case matters are a good proxy for the production lines of justice, while the related civil procedures represent the current technology adopted by judges in supplying justice. Based on the proposed hypothesis, judicial procedures can affect the efficiency of the courts and, when looking at litigious and non-litigious household dissolutions, Table 3 confirms significant differences among case matters. On average, the settlement of a household dissolution case requires 835 days in the first step (i.e., litigious separation) and another 844 days in the second step (i.e., litigious divorce), which adds up to more than 4 years to conclude a dispute between husband and wife. As for a non-litigious case, only 258 days are needed on average. Undoubtedly, this abnormal length is partially due to litigiousness between the parties and/or the lawyers' opportunistic behaviour to maximize profit from their cases (Felli et al, 2007), but the current procedures play a critical role, producing the financial negative externalities under investigation in this work.

Table 2. Inputs and outputs adopted in the Directional Distance Function (Italy, 2005-2010)

Variable		Mean	Std. Dev.	Min	Max	Observations
Judges (input)	overall	27.686	44.415	3.000	365.000	N = 989
	between		44.473	4.667	351.167	n = 165
	within		1.689	12.519	41.519	T-bar = 5.994
Demand for criminal and civil justice (input)	overall	52,157.320	72,591.260	3,963.000	541,817.000	N = 989
	between		72,623.200	4,236.833	526,858.300	n = 165
	within		4,095.295	5735.815	96,583.820	T-bar = 5.994
Supply of criminal and civil justice (good output)	overall	23,689.180	33,884.662	1,538.000	272,063.000	N = 989
	between		33,844.122	1,760.500	256,224.500	n = 165
	within		2,720.458	5,406.843	51,441.340	T-bar = 5.994
Additional costs (bad output)	overall	1,064,658	1,724,232	0	18,400,000	N = 989
	between		1,670,964	86,957	14,600,000	n = 165
	within		438,497	-2,932,465	5,877,700	T-bar = 5.994

Finally, analysing the Italian geographical macro areas, Table 4 shows the disaggregated additional costs that courts have to bear in order to supply justice. It can easily be observed that lawyers' fees are the main component of this negative externality, since they amount to almost half of the total costs. Other significant costs concern the fee/allowance of magistrate auxiliaries, honorary judges, and experts.

Table 3. Disaggregated descriptive statistics on average disposition time by civil case matter (Italy, 2005-2010)

Case matter	**North West**	**North East**	**Centre**	**South**	**Island**	**Italy**
Insolvency application	177	130	151	276	286	209
Insolvency	2778	2127	3460	6395	4427	4045
Regular Execution	233	141	286	328	405	281
Real Estate Execution	1083	832	1322	1793	1898	1406
Consensual Separation	97	78	108	141	95	108
Litigious Separation	571	668	1373	867	722	835
Consensual Divorce	127	115	286	130	96	150
Litigious Divorce	560	644	1562	743	820	844
Private labour	427	607	721	894	954	714
Public labour	649	813	930	820	933	816
Pension	457	537	541	887	699	637
Ordinary jurisdiction	686	824	991	1207	1136	973
Special Procedure	54	37	76	95	75	70
Other	268	259	232	302	372	285

Table 4. Disaggregated descriptive statistics on the average variable costs borne by courts to supply justice (Italy, 2005-2010)

Cost item	**North West**	**North East**	**Centre**	**South**	**Islands**	**Italy**
Travel costs	5,973.77	6,856.78	11,602.83	20,895.93	29,181.63	14,702.42
Activities costs	23,481.94	18,325.44	22,578.63	29,125.64	56,521.74	29,053.89
Extraordinary costs (criminal cases)	3,909.38	775.53	283.75	2,296.48	1,998.54	2,045.99
Postal charges	425.82	1,793.31	1,295.61	700.67	2,035.69	1,109.10
Demolition costs (abusive works)	149.22	85.99	55.81	38.51	112.91	86.93
Detention costs	17,923.62	8,405.14	2,496.15	7,451.29	3,333.80	8,615.09
Printing costs	330.29	3,369.68	6,125.79	1,475.69	7,585.52	3,251.81
Other costs	3,180.42	5,225.69	3,377.83	3,927.24	10,761.48	4,872.58
Travel allowance	323.79	282.94	852.32	1,566.31	4,635.71	1,402.62
Custody allowance	53,280.78	27,169.42	142,237.50	77,193.64	48,594.92	71,333.14
Honorary judges and experts allowance	87,169.27	90,482.73	103,695.85	118,571.59	108,754.75	102,504.62
Justices of the peace allowance	4,781.33	2,967.27	6,676.46	22,122.24	30,881.37	13,508.69
Other allowances	2,769.04	1,076.49	780.42	1,449.42	2,860.70	1,803.95
Private investigators' fees	3.31	-	0.44	14.71	3.28	5.40
Magistrate auxiliaries' fees	131,208.20	101,841.81	159,344.60	120,966.91	273,513.93	150,518.61
Technical consultants' fees	2,315.23	4,815.51	6,185.61	2,085.43	1,913.17	3,275.55

Lawyers' fees	449,986.93	318,115.83	539,697.35	435,634.85	753,305.37	488,092.69
Other	1,600.97	654.99	4,639.75	27,815.02	7,140.68	10,001.67
Social security costs	12,938.66	9,291.38	18,202.54	16,319.41	26,456.24	16,302.99
Value-added tax	125,367.52	90,610.56	166,133.97	135,339.53	04,659.39	142,183.85
Total costs	927,119.48	692,146.48	1,196,263.21	1,024,908.57	1,574,250.81	1,064,657.95

4. RESULTS

Based on the proposed empirical strategy, we estimate the TE scores, comparing different model definitions to test hypothesis H_1. Table 5 presents some descriptive statistics regarding the estimated TE scores at the level of Italy's macro areas, considering both the most suitable model (i.e., DDF with bad outputs) and the alternative one (i.e., DDF without bad outputs). Note that the values range between 0 and 1, with 0 representing the efficiency benchmark in the comparative analysis (see Section 3.1). This means that, as the TE scores increase, so does the distance from the efficiency frontier, indicating that the DMUs become more inefficient compared to the total population of courts under investigation.

Table 5. Average Technical Efficiency (TE) scores at the level of macro areas (Italy, 2005-2010)

Macro area	DDF TE Score (without Bad Outputs)	DDF TE Score (with Bad Outputs)	Number of courts
North West	0.2137	0.3362	40
North East	0.2859	0.4013	25
Centre	0.3229	0.4537	30
South	0.3280	0.4658	45
Islands	0.1889	0.3386	25
Italy	**0.2720**	**0.4032**	**165**

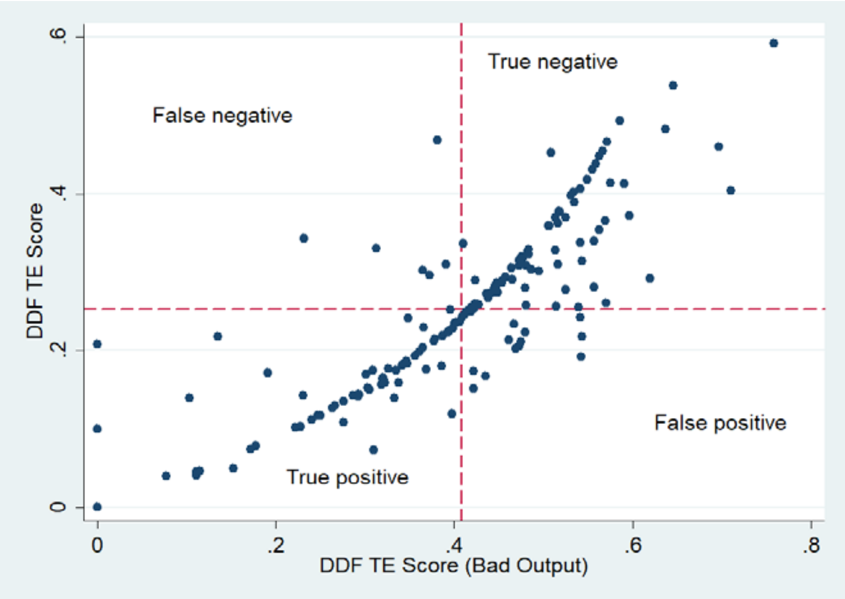
By comparing the two methodologies, we can see that the benchmark changes when the financial negative externality affecting the public budget is taken into account. If we look at the DDF model without bad outputs, the Islands macro area is clearly the national benchmark, which the policy makers could use as reference organisation for an effective structural reform. Nevertheless, if we consider financial distress (i.e., DDF with bad outputs), the North West macro area is the benchmark. This means that the Islands are very efficient in producing justice, given the demand for justice and number of judges, but the North West macro area is able to use the available financial resources in the most effective way[5]. Therefore, we cannot reject hypothesis H_1, that is to say, technical efficiency scores estimated without considering the negative externalities affecting the public budget can mislead policy makers into identifying incorrect benchmarks for policy reforms. Figure 3 shows this potential mistake, highlighting type I and type II errors in adopting the national average TE scores, with and without bad outputs, as leading reference values to reform the current judicial system.

In Figure 3, the DDF TE scores without bad outputs are arranged along the vertical axis, while the DDF TE scores with bad outputs are arranged along the horizontal axis. Based on the results collected, courts are plotted considering the year 2010 and adopting as reference values the national average TE scores used in the two model definitions (highlighted in red). We can thus

[5] Table S.1 in the Appendix B gives a more detailed picture of the TE scores based on the model definition, presenting descriptive statistics at the second instance level.

identify the *false negatives*, which are DMUs with a TE score above the national average in the DDF model but below the national average if we consider the financial negative externalities created by the production process (i.e., DDF model with bad outputs). These observations are courts with good internal organisation, able to control negative financial externalities, which means that they can produce justice with lower additional costs (type II error). Even more important are the *false positives*, since they could represent a benchmark for the policy reform. These *false positives* are DMUs with a TE score below the national average in the DDF model but above the national average if the financial externalities are also taken into account (i.e., DDF model with bad outputs). These courts are very efficient in producing justice, but they are unable to control the additional financial resources used by judges and staff in the production process. Without appropriate model definition, policy makers could mistakenly regard these courts as an organisational benchmark, even though they are not (type I error). Our results show that type I error is equal to 10.37%, while type II error is equal to 3.66%.

Figure 3. DDF TE scores, with and without bad outputs (Italy, 2010)



Adopted reference value to guide the policy makers: national average (highlighted in red)

Finally, in order to collect even more robust results, we tested the differences between the two DDF models by means of several Wilcoxon rank-sum tests (Wilcoxon, 1945; Mann and Whitney, 1947), which are a special case used for two samples of the Kruskal-Wallis statistic (Kruskal and Wallis, 1952). The Wilcoxon rank-sum tests perform well with the results of non-parametric techniques, verifying the null hypothesis that the estimated values are, on average, from the same population. Thanks to these tests, we can reject this null hypothesis, confirming that different model definitions can produce different populations of TE scores that are statistically significant, potentially misleading policy makers into identifying incorrect benchmarks (see Table S.2 in the Appendix B).

5. Conclusions

The impact of judicial efficiency on market dynamics is relevant (Chemin, 2009; Wang et al., 2014; Chakraborty, 2016; Shah et al., 2017; Moro et al., 2018), as it can support firm dynamics on the market (e.g., entrepreneurship, investments) as well as bank dynamics (i.e., access to the financial market, size of loans, interest rates). Consequently, there is a great need to improve the supply of justice, by correctly assessing the performance of the courts and of the judicial system as a whole. Moreover, in the current age of austerity, the limited public resources available have to be used in the best possible way, so as to reduce the financial negative externalities generated by this production process.

This work proposes a specific OR technique (i.e., DDF with bad outputs) that is able to estimate a judicial TE score for every court, benchmarking DMUs. This approach can effectively support the policy makers in reforming the national justice system and in controlling public expenditure. Furthermore, it represents a valid scientific basis allowing for the interests of the stakeholders to converge, and possibly generating wide consensus on the proposed policy reform.

Obviously, further improvements may be proposed based on fresh data. New information such as, for example, micro data on judges and staff (e.g., education, professional curriculum) may strengthen our current knowledge of the determinants of judicial inefficiency, but we nevertheless believe that this work represents a first step in this complex but much needed reform process. The die is cast, and the policy maker has to play.

6. References

Anderson, G., Crawford, I., & Leicester, A. (2011). Welfare rankings from multivariate data, a nonparametric approach. *Journal of Public Economics*, *95*(3-4), pp. 247-252.

Arai, R., Naito, K., & Ono, T. (2018). Intergenerational policies, public debt, and economic growth: A politico-economic analysis. *Journal of Public Economics*, 166, pp. 39-52.

Banker, R., Charnes, A., Cooper, W.W., Swarts, J., & Thomas, D.A. (1989). An introduction to data envelopment analysis with some of its models and uses. *Research in Governmental and Nonprofit Accounting*, 5, pp. 125-165.

Banker, R., Emrouznejad, A., Lopes, A. L. M., & de Almeida, M. R. (2012). *Data Envelopment Analysis: theory and Applications* (Proceedings of the 10th International Conference on DEA August 2012), Natal, Brazil, 350pp. ISBN:978 185449 437 5.

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), pp. 1078-1092.

Barseghyan, L., & Battaglini, M. (2016). Political economy of debt and growth. *Journal of Monetary Economics*, 82, pp. 36-51.

Beenstock, M., & Haitovsky, Y. (2004). Does the appointment of judges increase the output of the judiciary?. *International Review of Law and Economics*, *24*(3), pp. 351-369.

Berg S., (2010). *Water Utility Benchmarking: Measurement, Methodology, and Performance Incentives*. London, UK: IWA Publishing. ISBN: 9781843392729.

Bjurek, H., & Hjalmarsson, L. (1995). Productivity in multiple output public service: A quadratic frontier function and Malmquist index approach. *Journal of Public Economics*, *56*(3), pp. 447-460.

Boussofiane, A., Dyson, R.G., & Thanassoulis, E. (1991). Applied data envelopment analysis. *European Journal of Operational Research*, 52, pp. 1-15.

Buscaglia, E., & Ulen, T. (1997). A quantitative assessment of the efficiency of the judicial sector in Latin America. *International Review of Law and Economics*, *17*(2), pp. 275-291.

Chakraborty, P. (2016). Judicial quality and regional firm performance: the case of Indian states. *Journal of Comparative Economics*, *44*(4), pp. 902-918.

Chambers, R. G., Chung, Y., & Färe, R. (1996). Benefit and distance functions. *Journal of economic theory*, *70*(2), pp. 407-419.

Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of optimization theory and applications*, *98*(2), pp. 351-364.

Charnes, A., W. W. Cooper, & E. Rhodes. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research* 2, pp. 429-444.

Checherita-Westphal, C., & Rother, P. (2012). The impact of high government debt on economic growth and its channels: An empirical investigation for the euro area. *European economic review*, *56*(7), pp. 1392-1405.

Chemin, M. (2009). The impact of the judiciary on entrepreneurship: Evaluation of Pakistan's "Access to Justice Programme". *Journal of Public Economics*, *93*(1-2), pp. 114-125.

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), pp. 30-37.

Chudik, A., Mohaddes, K., Pesaran, M. H., & Raissi, M. (2017). Is there a debt-threshold effect on output growth?. *Review of Economics and Statistics*, *99*(1), 135-150.

Chung, Y.H., Färe, R., & Grosskopf, S. (1997). Productivity and undesirable outputs: a directional distance function approach. *Journal of Environmental Management*, *51*(3), 229-240.

Cooper, W.W., Li, S., Seiford, L.M., & Zhu, J. (2004). Sensitivity analysis in DEA. In W.W. Cooper, L.M. Seiford, & J. Zhu (Eds.), *Handbook on data envelopment analysis* (pp. 75–98). New York: Kluwer Academic Publishers, Inc. Chapter 3.

Cooper, W. W., Seiford, L. M., & Tone, K. (2000). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software*. Boston, MA: 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. Yale Human Rights and Development Law Journal*, *2*(1), pp. 87-142.

Dakpo, K.H., Jeanneaux, P., & Latruffe, L. (2016). Modelling pollution-generating technologies in performance benchmarking: Recent developments, limits and future prospects in the nonparametric framework. *European Journal of Operational Research*, *250*(2), pp. 347-359.

Daouia, A., Simar, L., & Wilson, P.W. (2017). Measuring firm performance using non-parametric quantile-type distances. *Econometric Reviews*, 36, pp. 156-181.

Daraio, C., & Simar, L. (2016). Efficiency and benchmarking with directional distances: a data-driven approach. *Journal of the Operational Research Society*, *67*(7), pp. 928-944.

De Grauwe, P. (1996). Monetary union and convergence economics. *European Economic Review*, *40*(3-5), pp. 1091-1101.

Deprins, D., L. Simar, & H. Tulkens. (1984). Measuring Labor Inefficiency in Post Offices. In M. Marchand, P. Pestieau, & H. Tulkens eds. *The Performance of Public Enterprises: Concepts and Measurements.* Amsterdam: North-Holland, pp. 243-267.

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), pp. 477-493.

Dyson, R.G., Allen, R., Camanho, A. S., Podinovski, V.V., Sarrico, C.S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132, pp. 245-259.

Falavigna, G., Ippoliti, R., & Manello, A. (2019). Judicial efficiency and immigrant entrepreneurs. *Journal of Small Business Management*, *57*(2), pp. 421-449.

Falavigna, G., Ippoliti, R., & Ramello, G.B. (2018). DEA-based Malmquist productivity indexes for understanding courts reform. *Socio-Economic Planning Sciences*, 62, pp. 31-43.

Falavigna, G., Ippoliti, R., Manello, A., & Ramello, G. B. (2015). Judicial productivity, delay and efficiency: A directional distance function (DDF) approach. *European Journal of Operational Research*, *240*(2), pp. 592-601.

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*, https://doi.org/10.1016/j.jpolmod.2021.06.001 (*In Press, Corrected proof*).

Färe, R., Grosskopf, S., & Whittaker, G. (2004). Distance Functions, Chapter 5. In W.W. Cooper, L.M. Seiford, & J. Zhu eds. *Handbook on Data Envelopment Analysis*. Norwell, Mass.: Kluwer Academic Publishers.

Färe, R., & Grosskopf, S. (2000). Theory and application of directional distance functions. *Journal of productivity analysis*, *13*(2), pp. 93-103.

Färe, R., Grosskopf, S., & Pasurka Jr, C.A. (2007). Environmental production functions and environmental directional distance functions. *Energy*, *32*(7), pp. 1055-1066.

Felli, E.L., Londoñ-Bedoya, D.A., Solferino, N., & Tria, G. (2007). The demand for justice: civil litigation and the judicial system. An Italian case study. In R. Ricciuti, & F. Padovano. Italian

Institution Reform: a public choice perspective. New York, NY: Springer. https://doi.org/10.1007/978-0-387-72141-5

Ferrandino, J. (2012). The Impact of Revision 7 on the Technical Efficiency of Florida's Circuit Courts. *Justice system journal*, *33*(1), pp. 22-46.

Finocchiaro Castro, M., & Guccio, C. (2014). Searching for the source of technical inefficiency in Italian judicial districts: an empirical investigation. *European Journal of Law and Economics*, 38(3), 369-391.

Finocchiaro Castro, M., & Guccio, C. (2015). Bottlenecks or Inefficiency? An assessment of first instance Italian courts' performance. *Review of Law & Economics*, *11*(2), pp. 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), pp. 83-116.

Fukuyama, H., & Weber, W. L. (2004). Economic inefficiency measurement of input spending when decision-making units face different input prices. *Journal of the Operational Research Society*, *55*(10), pp. 1102-1110.

García-Posada, M., & Mora-Sanguinetti, J. S. (2015). Does (average) size matter? Court enforcement, business demography and firm growth. *Small Business Economics*, *44*(3), pp. 639-669.

Giacomelli, S., & Menon, C. (2017). Does weak contract enforcement affect firm size? Evidence from the neighbour's court. *Journal of Economic Geography*, *17*(6), pp. 1251-1282.

Homburg, C. (2001). Using data envelopment analysis to benchmark activities. *International journal of production economics*, *73*(1), pp. 51-58.

Ippoliti R., (2015). The reform of judicial geography: technical efficiency and demand of justice. *Italian Journal of Public Economics*, Vol. 2 (2015), pp. 91-124. DOI: 10.3280/EP2015-002003.

Ippoliti, R., Melcarne, A., & Ramello, G. B. (2015a). Judicial efficiency and entrepreneurs' expectations on the reliability of European legal systems. *European Journal of Law and Economics*, *40*(1), pp. 75-94.

Ippoliti, R., Melcarne, A., & Ramello, G. B. (2015b). The Impact of judicial efficiency on entrepreneurial action: A European perspective. *Economic Notes: Review of Banking, Finance and Monetary Economics*, *44*(1), pp. 57-74.

Kittelsen, S. A., & Førsund, F. R. (1992). Efficiency analysis of Norwegian district courts. *Journal of Productivity Analysis*, *3*(3), pp. 277-306.

Kneip, A., Simar, L., & Wilson, P. W. (2016). Testing hypotheses in nonparametric models of production. *Journal of Business and Economic Statistics*, 34, pp. 435-456.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), pp. 583-621.

Kumar, N., & Singh, A. (2014). Efficiency analysis of banks using DEA: A review. *International Journal of Advance Research and Innovation*, 1, pp. 120-126.

Kuosmanen, T., & Kortelainen, M. (2005). Measuring eco-efficiency of production with data envelopment analysis. *Journal of Industrial Ecology*, 9(4), pp. 59-72.

Lewin, A. Y., Morey, R. C., & Cook, T. J. (1982). Evaluating the administrative efficiency of courts. *Omega*, *10*(4), pp. 401-411.

Lim, B., Lee, K., & Lee, C. (2016). Free Disposal Hull (FDH) analysis for efficiency measurement: an update to DEA. *The Stata Journal*, pp. 1-8.

Lovell, C., S. Grosskopf, E. Ley, J. Pastor, D. Prior, & P. Eeckaut. (1994). Linear programming approaches to the measurement and analysis of productive efficiency. *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research 2*(2), pp. 175-248.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pp. 50-60.

Mattsson, P., & Tidanå, C. (2019). Potential efficiency effects of merging the Swedish district courts. *Socio-Economic Planning Sciences*, 67, pp. 58-68.

Mattsson, P., Månsson, J., Andersson, C., & Bonander, F. (2018). A bootstrapped Malmquist index applied to Swedish district courts. *European Journal of Law and Economics*, 46, pp. 109-139.

Melcarne, A., & Ramello, G. B. (2015). Judicial independence, judges' incentives and efficiency. *Review of Law & Economics*, *11*(2), pp. 149-169.

Mitsopoulos, M., & Pelagidis, T. (2007). Does staffing affect the time to dispose cases in Greek courts?. *International Review of Law and Economics*, *27*(2), pp. 219-244.

Moro, A., Maresch, D., & Ferrando, A. (2018). Creditor protection, judicial enforcement and credit access. *The European Journal of Finance*, *24*(3), pp. 250-281.

Müller, A., Storesletten, K., & Zilibotti, F. (2016). The political color of fiscal responsibility. *Journal of the European Economic Association*, *14*(1), pp. 252-302.

Paradi, J. C., Vela, S., & Yang, Z. (2004). Assessing bank and bank branch performance: Modelling considerations and approaches. In W.W. Cooper, L.M. Seiford, & J. Zhu (Eds.), *Handbook on data envelopment analysis* (pp. 349-400). New York: Kluwer Academic Publishers, Inc. Chapter 13.

Park, B.U., Jeong, S.-O., & Simar, L. (2010). Asymptotic distribution of conical-hull estimators of directional edges. *Annals of Statistics*, 38, pp. 1320-1340.

Park, B.U., Simar, L., & Weiner, C. (2000). FDH efficiency scores from a stochastic point of view. *Econometric Theory*, 16, pp. 855-877.

Pedraja-Chaparro, F., & Salinas-Jimenez, J. (1996). An assessment of the efficiency of Spanish Courts using DEA. *Applied economics*, *28*(11), pp. 1391-1403.

Peyrache, A., & Zago, A. (2016). Large courts, small justice! The inefficiency and the optimal structure of the Italian justice sector. *Omega*, *64*, pp. 42-56.

Raab, R., & Lichty, R. (2002). Identifying sub-areas that comprise a greater metropolitan area: The criterion of country relative efficiency. *Journal of Regional Science*, 42, pp. 579-594.

Ramseyer, J.M. (2012). Talent matters: Judicial productivity and speed in Japan. *International Review of Law and Economics*, *32*(1), pp. 38-48.

Riccardi, R., Oggioni, G., & Toninelli, R. (2012). Efficiency analysis of world cement industry in presence of undesirable output: application of data envelopment analysis and directional distance function. *Energy Policy*, 44, pp. 140-152.

Santos, S. P., & Amado, C. A. (2014). On the need for reform of the Portuguese judicial system–Does Data Envelopment Analysis assessment support it?. *Omega*, 47, pp. 1-16.

Schneider, M.R. (2005). Judicial career incentives and court performance: an empirical study of the German labour courts of appeal. *European Journal of Law and Economics*, *20*(2), pp. 127-144.

Shah, A., Shah, H.A., Smith, J.M., & Labianca, G.J. (2017). Judicial efficiency and capital structure: An international study. *Journal of Corporate Finance*, 44, pp. 255-274.

Silva, M. C. A. (2018). Output-specific inputs in DEA: An application to courts of justice in Portugal. *Omega*, 79, pp. 43-53.

Simar, L., & Wilson, P. W. (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research*, *139*(1), pp. 115-132.

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, *136*(1), pp. 31-64.

Simar, L., & Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: recent developments and perspectives. Cap. 4th. In H.O. Fried, C.A. Knox Lovell, & S.S. Schmidt (eds). *The measurement of productive efficiency and productivity growth*. Oxford Press.

Tulkens, H. (2006). On FDH efficiency analysis: some methodological issues and applications to retail banking, courts and urban transit. In *Public goods, environmental externalities and fiscal competition* (pp. 311-342). Boston, MA: Springer.

Valdmanis, V. (1992). Sensitivity analysis for DEA models: An empirical example using public vs. NFP hospitals. *Journal of Public Economics*, *48*(2), pp. 185-205.

Wang, Y., Wang, Y., & Li, K. (2014). Judicial quality, contract intensity and exports: Firm-level evidence. *China Economic Review*, 31, pp. 32-42.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, 1, pp. 80-83.

Wilson, P. W. (2018). Dimension reduction in nonparametric models of production. *European Journal of Operational Research*, 267(1), pp. 349-367.

Woo, J., & Kumar, M. S. (2015). Public debt and growth. *Economica*, *82*(328), pp. 705-739.

Yeung, L. L., & Azevedo, P. F. (2010). Measuring efficiency of Brazilian courts with data envelopment analysis (DEA). *IMA Journal of Management Mathematics*, *22*(4), pp. 343-356.

Zhou, P., Ang, B. W., & Poh, K. L. (2008). A survey of data envelopment analysis in energy and environmental studies. *European journal of operational research*, *189*(1), pp. 1-18.

Zofío, J. L., & Prieto, A. M. (2001). Environmental efficiency and regulatory standards: the case of CO2 emissions from OECD industries. *Resource and Energy Economics*, *23*(1), pp. 63-83.

Zofio, J. L., Pastor, J. T., & Aparicio, J. (2013). The directional profit efficiency measure: on why profit inefficiency is either technical or allocative. *Journal of Productivity Analysis*, *40*(3), pp. 257-266

7. Appendix A

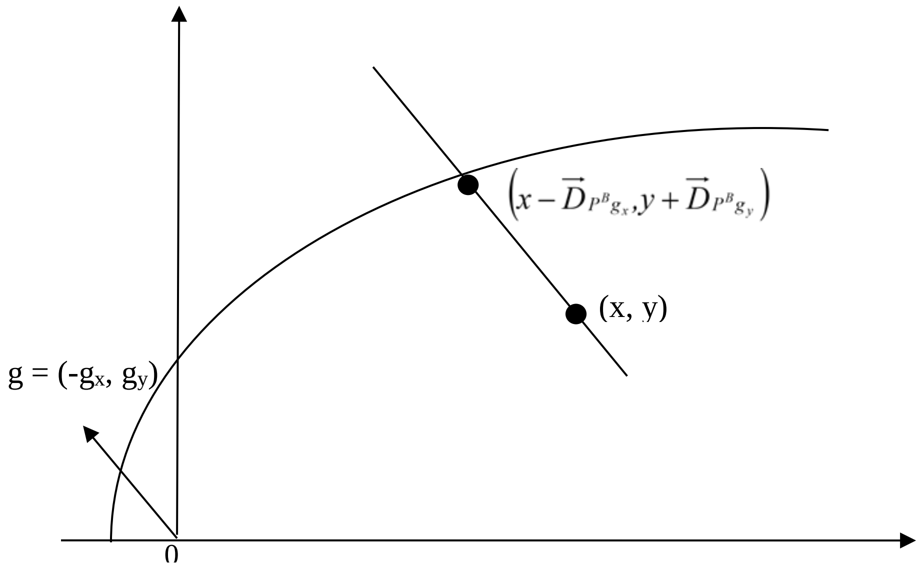
As suggested by Falavigna and Ippoliti (2021), the DDF models proposed allow to to obtain an efficiency measure projecting the observation (x, y) in a pre-assigned direction g = ($-g_x$, g_y) ≠ 0 in proportion to β. In general, through the DDF model we can obtain the same results given by the DEA approach when the directional vector is set equal to the inputs and outputs. However, following Chambers et al. (1998) and Färe and Grosskopf (2000), the directional function is set as $\vec{D}\left(\boldsymbol{x},\boldsymbol{y};-\boldsymbol{g_x},\boldsymbol{g_y}\right) = sup\{\beta:(\boldsymbol{x}-\beta\boldsymbol{g_x},\boldsymbol{y}+\beta\boldsymbol{g_y}) \in P\}$, and in this manner, it is able to simultaneously maximize outputs and minimize inputs. Considering a single input and a single output case, Figure S.1 represents the directional vector ($-g_x$, g_y), in which the vector of input and output (x, y) is projected onto P^B (i.e., the efficient frontier) at $\left(\boldsymbol{x}-\vec{D}_{P^B}\boldsymbol{g_x},\boldsymbol{y}+\vec{D}_{P^B}\boldsymbol{g_y}\right)$ and $\vec{D}_{P^B}$ = $\vec{D}_{P^B}\left(\boldsymbol{x},\boldsymbol{y};-\boldsymbol{g_x},\boldsymbol{g_y}\right)$.

The linear programming for DDF with Constant Returns to Scale (CRS) assumption is as follows (Falavigna and Ippoliti, 2021):

$$\max_{\beta,\lambda} \quad \beta$$
subject to
$$X\lambda \le \mathbf{x_o} - \beta\mathbf{g_x}$$
$$Y\lambda \ge \mathbf{y_o} + \beta\mathbf{g_y}$$
$$\lambda \ge \mathbf{0} \qquad (1)$$

the observation is efficient when β^*_{CRS} = 0, (i.e., a benchmark); if the score increases, inefficiency also increases. This model represents the benchmark for policy makers in order to define the correct policy reforms, considering the number of judges and the incoming and pending cases (i.e., the demand for justice) as inputs and the number of settled cases as output. The increases the TE score values (i.e., β^*_{CRS} > 0) mean that the level of inefficiency in supplying justice increases, considering the current demand for justice and the available resources.

Figure S.1. Directional Distance Function (DDF)



Under some specific assumptions, as previously suggested, the DDF can consider two different types of outputs (i.e., good or desirable, and bad or undesirable). Taking our specific case study into account, policy makers should consider the unavoidable additional expenditure of the supply of justice as undesirable or bad (*b*) output. This specific assumption requires a reformulation of the production possibility set, that now is defined as follows:

$$P = \{(\boldsymbol{x}, \boldsymbol{y^d}, \boldsymbol{y^u}) | \boldsymbol{x} \geq X\boldsymbol{\lambda}, \boldsymbol{y^d} \leq Y\boldsymbol{\lambda}, \boldsymbol{y^u} = Y\boldsymbol{\lambda}, \boldsymbol{\lambda} \geq 0\}$$

where undesirable outputs are weakly disposable (Cooper et al., 2007). The TE scores represent the ability of the observation to reduce the bad output increasing the good one and taking inputs equal.

The linear program [1] has to be rewritten as follows (Falavigna and Ippoliti, 2021):

$$\max_{\beta,\lambda} \quad \beta$$

subject to

$$\begin{aligned} & X\lambda \leq \mathbf{x}_{\mathbf{ox}} \\ & Y^d\lambda \geq \mathbf{y}_{\mathbf{o}}^{\mathbf{d}} + \beta\mathbf{y}_{\mathbf{o}}^{\mathbf{d}} \\ & Y^u\lambda \leq \mathbf{y}_{\mathbf{o}}^{\mathbf{u}} + \beta\mathbf{y}_{\mathbf{o}}^{\mathbf{u}} \\ & \max\{\mathbf{y}_{\mathbf{i}}^{\mathbf{u}}\} \geq \mathbf{y}_{\mathbf{o}}^{\mathbf{u}} + \beta\mathbf{y}_{\mathbf{o}}^{\mathbf{u}} \\ & \lambda \geq \mathbf{0} \end{aligned} \qquad 2$$

As in previous case, the optimal solution is again β^*_{CRS} = 0, otherwise the observation is inefficient (i.e., β^*_{CRS} > 0)[6]. In addition, Färe et al. (2007) suggests that the technical frontier has to satisfy some standard axioms (inactivity, compactness of production possibility set, free disposability of inputs, weak disposability of outputs, null-jointness, free disposability of good outputs) that in the case of the justice production process are satisfied.

The comparison of two presented models (without and with bad outputs) can verify if the model definition can lead the policy maker consideration and consequently hinder the necessary reforms. Considering the weak disposability axiom, the additional costs can be reduced only if also the production of good output decreases. In addition, because of null-jointness assumption, additional costs are necessary to settle judicial cases and then it is not possible to supply justice without increasing justice expenditure.

Finally, from the technical point of view, the size of input-output space has been defined according with Wilson (2018) that suggests a simple formula in order to guarantee robust results. In details, considering the DDF with the bad output, the 4 variables introduced in the model allow to obtain convincing results as an OLS with a sample equal to 165 observations. In the formulation without the undesirable output, the input-output space considers 3 items allowing to obtain convincing results as an OLS with a sample equal to 905 observations.

[6] Notice that, in order to compare results, all the efficiency scores range between 0 and 1, where scores equal to 0 identify efficient observations, otherwise the level of inefficiency increases.

8. Appendix B

Figure S.2: Italian judicial geography, first and second instance districts (Italy, 2005-2011)

First instance districts (i.e., Circondari Giudiziari) | *Second instance districts (i.e., Distretti di Corte di Appello)*

Table S.1. Average Technical Efficiency (TE) scores at the second instance level (Italy, 2005-2010)

District of Second Instance	Number of Courts	DDF TE Scores (without Bad Outputs)	DDF TE Score (with Bad Outputs)
Ancona	7	0.3261	0.4469
Bari	4	0.5412	0.6322
Bologna	9	0.3505	0.4644
Brescia	5	0.2702	0.3771
Cagliari	6	0.2195	0.3728
Caltanissetta	4	0.0887	0.1904
Campobasso	3	0.2713	0.4055
Catania	5	0.2445	0.4185
Catanzaro	8	0.2779	0.4586
Firenze	10	0.3027	0.4394
Genova	7	0.2409	0.3786
L'Aquila	8	0.3540	0.4853
Lecce	3	0.3566	0.5407
Messina	4	0.2479	0.4186
Milano	11	0.2000	0.3349
Napoli	8	0.3566	0.4194
Palermo	6	0.1401	0.2842
Perugia	4	0.2691	0.3706
Potenza	4	0.2545	0.4651
Reggio Calabria	3	0.1809	0.3596
Roma	9	0.3667	0.5118
Salerno	4	0.3104	0.4374
Torino	17	0.1946	0.3076
Trento	3	0.1857	0.3148
Trieste	5	0.1992	0.3090
Venezia	8	0.3049	0.4203
Italy	**165**	**0.2720**	**0.4032**

Table S.2: Wilcoxon rank-sum tests (Italy, 2005-2010)

Year = 2005	North West	North East	Centre	South	Islands
H0: DDF=DDF no Bad H1: DDF ≠ DDF no Bad	0.000	0.000	0.000	0.000	0.001
H0: DDF=DDF no Bad H1: DDF>DDF no Bad	0.000	0.000	0.000	0.000	0.001
H0: DDF=DDF no Bad H1: DDF<DDF no Bad	0.999	0.999	0.999	0.999	0.999
Freq.	40	25	30	45	25

Year = 2006	North West	North East	Centre	South	Islands
H0: DDF=DDF no Bad H1: DDF ≠ DDF no Bad	0.001	0.015	0.008	0.000	0.011
H0: DDF=DDF no Bad H1: DDF>DDF no Bad	0.000	0.007	0.004	0.000	0.005
H0: DDF=DDF no Bad H1: DDF<DDF no Bad	0.999	0.992	0.996	0.999	0.995
Freq.	40	25	30	45	25

Year = 2007	North West	North East	Centre	South	Islands
H0: DDF=DDF no Bad H1: DDF ≠ DDF no Bad	0.000	0.000	0.000	0.000	0.001
H0: DDF=DDF no Bad H1: DDF>DDF no Bad	0.000	0.000	0.000	0.000	0.000
H0: DDF=DDF no Bad H1: DDF<DDF no Bad	1.000	0.999	0.999	0.999	0.999
Freq.	40	25	30	45	25

Year = 2008	North West	North East	Centre	South	Islands
H0: DDF=DDF no Bad H1: DDF ≠ DDF no Bad	0.000	0.000	0.000	0.000	0.000
H0: DDF=DDF no Bad H1: DDF>DDF no Bad	0.000	0.000	0.000	0.000	0.000
H0: DDF=DDF no Bad H1: DDF<DDF no Bad	0.999	0.999	0.999	1.000	0.999
Freq.	40	25	30	45	24

Year = 2009	North West	North East	Centre	South	Islands
H0: DDF=DDF no Bad H1: DDF ≠ DDF no Bad	0.000	0.004	0.002	0.000	0.003
H0: DDF=DDF no Bad H1: DDF>DDF no Bad	0.000	0.002	0.001	0.000	0.002
H0: DDF=DDF no Bad H1: DDF<DDF no Bad	0.999	0.998	0.999	0.999	0.998
Freq.	40	25	30	45	25

Year = 2010	North West	North East	Centre	South	Islands
H0: DDF=DDF no Bad H1: DDF ≠ DDF no Bad	0.000	0.001	0.000	0.000	0.000
H0: DDF=DDF no Bad H1: DDF>DDF no Bad	0.000	0.000	0.000	0.000	0.000
H0: DDF=DDF no Bad H1: DDF<DDF no Bad	1.000	0.999	0.999	1.000	0.999
Freq.	40	25	30	45	24

ISSN (online) 2421-7158

CNR-IRCrES Working Papers

2021

N. 5/2021 Different waves and different policy interventions in 2020 Covid-19 in Italy: did they bring different results?. Mario Nosvelli.
N. 4/2001 On Search of a General Model of Technology Innovation. Angelo Bonomi.
N. 3/2021 Design and implementation of a web survey on the effects of evaluation on academic research. Andrea Orazio Spinello, Emanuela Reale, Antonio Zinilli.
N. 2/2021 An online survey on the effects of agile working in Italian Public Research Organisations. Serena Fabrizio, Valentina Lamonica, Andrea Orazio Spinello.
N. 1/2021 Technology Transfer Activities in Universities and Public Research Organizations: A Literature Overview. Ugo Finardi, Rolfo Secondo, Isabella Bianco.

2020

N. 12/2020 Unexpected loss multiperiodale e pricing del rischio di credito. Franco Varetto.
N. 11/2020 La ricerca in Nanotecnologie e Nanoscienze in Italia: spesa del settore pubblico e aree tematiche prevalenti. Ugo Finardi, Andrea Orazio Spinello.
N. 10/2020 Persistent fast growth and profitability. Lucio Morettini, Bianca Potì, Roberto Gabriele.
N. 9/2020 Binomio *Burnout* e *Mindfulness* nelle organizzazioni. Alcuni studi e scenari di applicazione. Oriana Ippoliti, Riccardo Briotti, Bianca Crocamo, Antonio Minopoli.
N. 8/2020 Innovation and communication of companies on Twitter before and during COVID-19 crisis. José N. Franco-Riquelme, Antonio Zinilli, Joaquín B. Ordieres-Meré and Emanuela Reale.
N. 7/2020 The proposal of a new hybrid methodology for the impact assessment of energy efficiency interventions. An exploratory study. Monica Cariola, Greta Falavigna.
N. 6/2020 The technology innovative system of the Silicon Valley. Angelo Bonomi.
N. 5/2020 Storia dell'industria delle macchine utensili in Piemonte dalle origini alla seconda guerra mondiale. Secondo Rolfo.
N. 4/2020 Blockchain e Internet of Things per la logistica Un caso di collaborazione tra ricerca e impresa. Edoardo Lorenzetti, Lucio Morettini, Franco Mazzenga, Alessandro Vizzarri, Romeo Giuliano, Paolo Peruzzi, Cristiano Di Giovanni.
N. 3/2020 L'impatto economico e fiscale di un evento culturale: misure e scala territoriale. Giovanna Segre, Andrea Morelli.
N. 2/2020 Mapping the tangible and intangible elements of the historical buildings and spaces. Edoardo Lorenzetti, Nicola Maiellaro.
N. 1/2020 Il lavoro agile negli enti pubblici di ricerca. Emanuela Reale, Serena Fabrizio, Andrea Orazio Spinello.

2019

N. 6/2019 Women's candidatures in local elections: does the context matter? Empirical evidence from Italian municipalities. Igor Benati, Greta Falavigna, Lisa Sella.
N. 5/2019 Research activities in Nanotechnologies and Nanosciences: an analysis of Piedmont's nanotech research system. Ugo Finardi.
N. 4/2019 Xylella fastidiosa: patogenesi, danni economici e lotta al disseccamento rapido dell'olivo. Maurizio Conti.
N. 3/2019 Flussi di traffico attraverso il tunnel automobilistico del Frejus: un semplice esercizio di *forecasting* e alcune considerazioni a margine. Ugo Finardi.
N. 2/2019 The Start-up Venture Capital Innovation System Comparison with industrially financed R&D projects system. Angelo Bonomi.
N. 1/2019 Complessità delle organizzazioni, complessità della formazione. Report di studio qualitativo ed analisi ermeneutica del Modello TRASE – IRCRES/CNR-IMO. Anna Chiara Scardicchio.

Numeri precedenti/Previous issues

This manuscript focuses on the Italian judicial system and on how to shape a policy reform aimed at increasing court efficiency, taking the financial negative externalities generated by this production process into account. On the one hand, the authors identify the benchmarks and main drivers of judicial inefficiency, while, on the other hand, they show how incorrect model definition may mislead policy makers tackling this reform process, based on an analysis of the Directional Distance Function with and without bad outputs. According to the results, incorrect model definition causes a type I error equal to 10.37% and a type II error equal to 3.66%. Policy implications concern the opportunity to adopt the proposed model and the collected benchmarks to reform the judicial system, improving its technical efficiency and maintaining the public budget under control.

CNR - Consiglio Nazionale delle Ricerche

IRCrES - Istituto di Ricerca sulla Crescita Economica Sostenibile