UNIVERSITÀ DELLA CALABRIA | DIPARTIMENTO DI ECONOMIA, STATISTICA E FINANZA "GIOVANNI ANANIA"

Working Paper Series

WP n° 2, febbraio 2021

ON THE HETEROGENEITY IN THE JUDICIAL EFFICIENCY LITERATURE: A META-REGRESSION ANALYSIS

Francesco Aiello

University of Calabria, Department of Economics, Statistics and Finance "Giovanni Anania" – DESF

(e-mail: francesco.aiello@unical.it)

Graziella Bonanno

University of Salerno, Department of Economics

(e-mail: gbonanno@unisa.it)

Francesco Foglia

University Dante Alighieri, Reggio Calabria

(e-mail: francesco.foglia@unical.it)

Info
Via Pietro Bucci, Cubo **0/C**
87036 Rende (Cs) - Italia
tel. (+39) 0984 492434 / 492422

http://www.unical.it/desf

On the heterogeneity in the judicial efficiency literature: a meta-regression analysis[1]

Francesco Aiello[a], Graziella Bonanno[b], Francesco Foglia[c]

francesco.aiello@unical.it - gbonanno@unisa.it – francesco.foglia@unical.it

[a] Department of Economics, Statistics and Finance "Giovanni Anania"
University of Calabria - Italy

[b] Department of Economics and Statistics, University of Salerno - Italy

[c] University Dante Alighieri Reggio Calabria – Italy

[*This version: October 2021*]

Abstract. This study performs a meta-regression analysis (MRA) with the aim of explaining the differences in results obtained when estimating a frontier for judicial institutions. The metadata set comprises 264 efficiency scores retrieved from 36 papers published from 1992 to 2019. The empirical setting is based on the two-step approach proposed by Gallet and Doucouliagos (2014). Main results are threefold. Firstly, it is proven that primary papers using parametric methods yield higher efficiency scores than non-parametric studies. Secondly, efficiency scores from studies that analyse the first-instance courts are significantly higher than those obtained for the appeal courts. Finally, primary papers focusing on a specific court type (only tax, civil or criminal) yield efficiency scores that on average are higher than those obtained when the analysed courts are mixed.

JEL classification: C14, C80, D63, K40, P37

Keywords: district courts, efficiency, frontier models, judicial system, meta-analysis

1. Introduction

This paper analyses the literature focusing on judicial efficiency with the aim to explain why the results obtained by prior research are heterogeneous. It is motivated on two grounds.

Firstly, at a time of restrictive fiscal policies, the need for evidence-based policy decisions drives the demand for efficiency analyses on how public sectors work. In view of these concerns, it is of great interest to investigate whether courts act in a technically efficient way (i.e., reach specific economic scopes, such as cost minimisation or output maximisation).

Secondly, the judicial system is a sector producing a specific service - justice - and, accordingly, it can be studied by using the customary tools of production and efficiency theory (Falavigna et al., 2018; Marciano et al., 2019; Voigt, 2016)). Since the seminal work by Farrell (1957), it has been clearly explained whether a decision unit is efficient or not. However, from an empirical perspective there is much controversy among scholars because it is not certain that one method is superior than others (Berger and Humphrey, 1997; Coelli and Perelman, 1999; Fethi and Pasourias, 2010).

Here, it is also worth mentioning that if the analysis is policy oriented, such that results are used for decisions regarding judicial resource allocation, then the reliability of the estimated efficiency scores will become crucial. Indeed, it is expected that an efficiency study

[1] The authors thank Sergio Destefanis, Kristiaan H.J. Kerstens, Martin Paldam and Angelo Zago for valuable suggestions on an earlier version of the paper. The authors are grateful to the participants of the seminars at the Bank of Italy, Universities of Naples, Palermo, Reggio Calabria and Salerno for their helpful comments on preliminary versions of this work. Usual disclaimer applies. Early versions of the paper were presented at EWEPA 2019 (London, June 10-14, 2019), at 2019 MAER-Net Colloquium at Greenwich University London (October 11-12, 2019) and at the 60th Annual Scientific Meeting of the Italian Economic Society at the University of Palermo (October 24-26, 2019).

provides an explanation on the key determinants of the courts' performance, thereby allowing policymaking to adopt decisions to improve court efficiency (Silva 2018).

As in other sectors, the empirical literature on judicial efficiency argues that the results are sensitive to study design and model specifications (Marciano et al., 2019; Voigt, 2016). Marciano et al. (2019) address an intriguing issue in this field of research in regard to the difference between efficiency and efficacy of judicial institutions. These scholars propose theoretical and empirical evidence supporting the necessity of a clear distinction between the two dimensions of courts. The qualitative survey provided by Voigt (2016) is closer to the specific scope of our paper. While Voigt (2016) offers valuable arguments in terms of why efficiency differs when focusing on courts' delays, our study enriches this debate because it provides a quantitative evaluation on the variability of courts' efficiency determined by the study design of primary-papers. This is the novelty of this study, and at this end a meta-regression analysis (MRA) is performed.

MRA is a method used to collect and integrate the results from individual papers. It allows to evaluating whether, and to what extent, the features of each work (i.e. estimation method, year of publication, functional form, court-type, sample size, and dimension) affect the findings when performing an efficiency study (Stanley 2001). As Glass (1976) says, MRA 'connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempt to make sense of the rapidly expanding research literature' (Glass 1976, 3).

Given that the literature on judicial efficiency lends itself well to being summarized through an MRA, it is noteworthy that no exhaustive work has explored the 'court efficiency-study design' *nexus*. In attempting to fill this gap, we consider all the judicial studies published from 1992 to 2019 (which were available in August 2021). The final meta dataset comprises 264 observations from 36 papers. It comes from different sources with no restriction on the countries covered in the original studies, thereby assuring full coverage of the literature based on studies that estimate a frontier for judicial courts.

When specifying our meta regressions we follow two recent papers (Aiello and Bonanno, 2018; 2019) which - in explaining the variability of efficiency in banking and municipalities – firstly estimate a random effects model and then consider a Weighted Least Squares (WLS), as suggested by Gallet and Doucouliagos (2014).

Due to its main research focus, i.e., measuring the impact of potential sources of heterogeneity on judicial efficiency, this article contributes to the debate assessing the role of methodological choices made by researchers when specifying the frontier of courts. Therefore, the paper's contribution lies in the fact that it addresses a number of issues, such as for instance: whether the efficiency scores retrieved from non-parametric papers differ from those estimated by parametric papers; whether the sample size of courts covered by primary papers matters in influencing the results; and whether papers focusing on generic courts provide different efficiency scores than papers focusing on specialised courts. All this also helps to identify better the variability in judicial efficiency scores and thus addresses the so-called "*apples and oranges*" MRA issue due to the fact that the metadata set comes from studies which are different from one another (Glass et al., 1981).

The paper is organised as follows. Section 2 presents the method used to create the meta-dataset. Section 3 briefly describes the distribution of court efficiency scores. Section 4 presents the meta-regression and describes the variables used in the analysis. Section 5 discusses the results and Section 6 concludes.

2. Construction of the meta data-set on court efficiency scores

The number of papers on judicial efficiency is limited and thus this study covers all the related literature. However, in order to attain the full coverage of prior research, a search based on three different phases was performed.

Firstly, the main and most consulted scientific paper repositories were queried, and the research results for “judicial efficiency” and “court efficiency” were analysed. For the first term, 106,000 results were obtained in Google, and 12,100 in Google Scholar, while in the scientific databases the number was significantly lower: 120 in SSRN, 48 in Science Direct and 74 in WoS. With regard to the second search item, “court efficiency”, the results were 28,900 in Google, 2,270 in Google Scholar, 248 in Science Direct and 16 in WoS.

Secondly, the search was refined searching for “judicial”, “efficiency” and “frontier” in the titles, abstracts and keywords. The main journals on the fields were manually consulted, and papers were further selected with a focus on efficiency. The references from the qualitative surveys made by Marciano et al. (2019) and Voigt (2016) have been subsequently scanned. Before filtering this sample of works, we ensured that they (a) focused on the efficiency of the judicial system; (b) estimated a frontier using parametric or non-parametric methods (c) reported the results required to run an meta-regression (that is the average efficiency score and its standard deviation); (d) were published in English in a journal or as discussion paper. This process resulted in the selection of 44 papers.

Finally, the authors of selected papers were contacted for further information or clarifications, and eight papers were removed following this interaction with the authors. The search then yields a sample of 36 papers published from 1992 to 2019 (Figure 1).[2] Data extraction was concluded on 31st August 2021.

[2] Antonucci, L., et al. (2014); Castro, A. S. (2009); Elbialy, N., and García-Rubio, M. A. (2011); Espasa, M., and Esteller-Moré, A. (2015); Falavigna, G., et al. (2015); Falavigna, G., et al. (2018); Fauvrelle, T. A., and Almeida, A. T. C. (2018); Ferrandino, J. (2012); Ferro, G., et al.(2018); Finocchiaro Castro, M., and Guccio, C. (2014); Finocchiaro Castro, M., and Guccio, C. (2018); Fusco, E., et al.(2018); Gorman, M. F., and Ruggiero, J. (2009); Guzowska, M., and Strak, T. (2010); Ippoliti, R., et al.(2015a; 2015b); Ippoliti, R., and Ramello, G. B. (2018); Ivanova, N. B. (2014); Kittelsen, S. A. C., & Førsund, F. R. (1992); Major, W. (2015); Mattsson, P., and Tidanå, C. (2019); Melcarne, A., and Ramello, G. (2015); Nissi, E., et al. (2019); Nissi, E., & Rapposelli, A. (2010); Odhiambo, O. J. (2014); Pedraja-Chaparro, F., and Salinas-Jimenez, J. (1996); Peyrache, A., and Zago, A. (2016); Rushid, A. R. (2018); Santos, S. P., & Amado, C. A. F. (2014); Schneider, M. R. (2005); Silva, M. C. A. (2018); Sousa, M. de M., and Guimaraes, T. A. (2018); Tsai, C., and Tsai, J. (2010); Yeung, L. (2014); Yeung, L., and Azevedo, P. (2011a; 2011b).

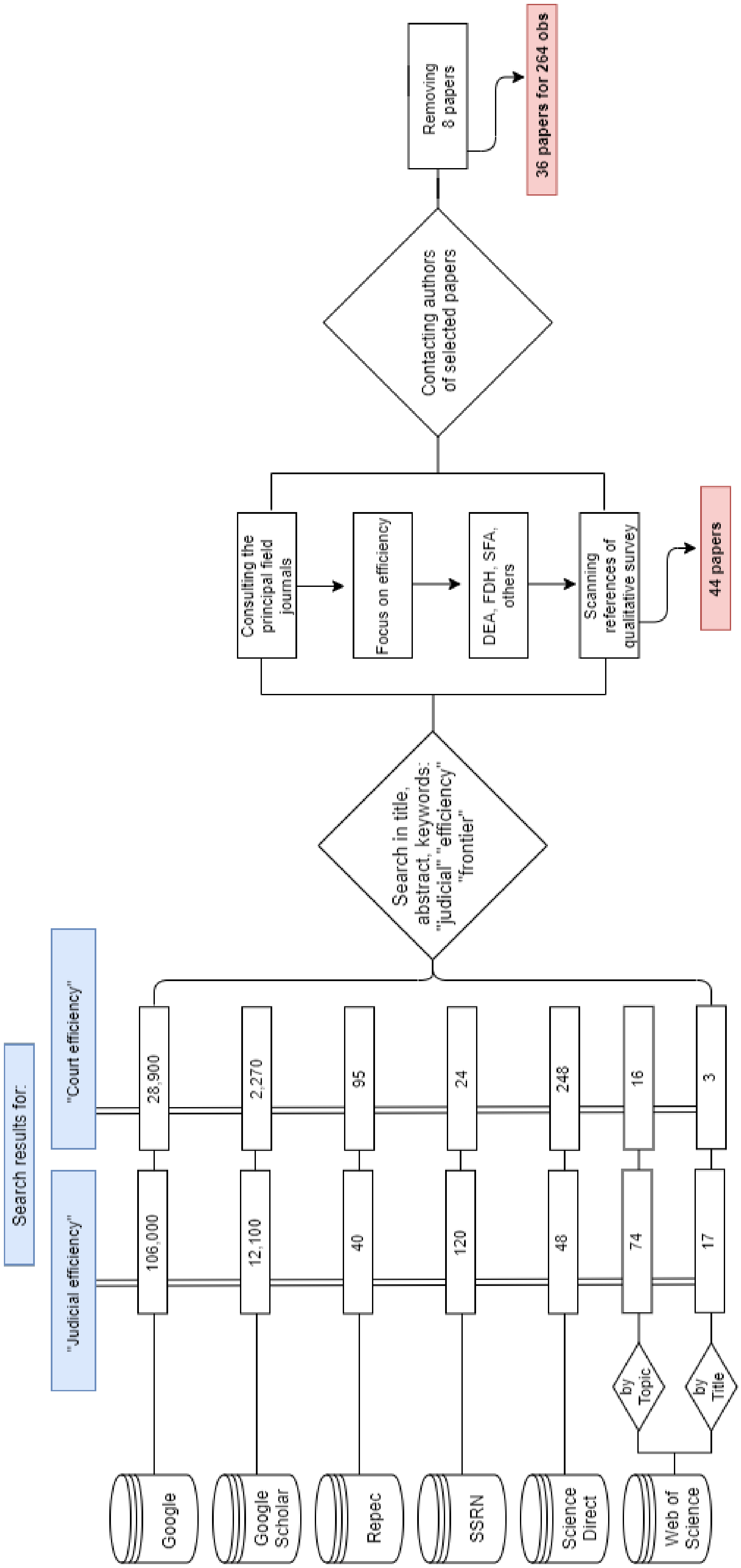


Figure 1. The dataset assembling process

3. The heterogeneity in the literature on courts' efficiency

Table 1 presents some descriptive statistics of the collected data by grouping efficiency scores according to (a) the approach used in the estimations (parametric or non-parametric), (b) the structure of the data (panel or cross-sectional), (c) the hypotheses regarding returns to scale (constant or variable), (d) the judicial degree (first instance or others), (e) the type of court (if specialised on a specific matter or generalist).

Overall, the sample of 264 observations yields an (unweighted) average efficiency of 0.752. It also emerges that the difference in mean (0.154 = 0.885 - 0.731) is statistically significant when comparing the 37 observations from parametric studies and the 227 observations from non-parametric papers. With regards to the data used in primary studies, 60% of the observations come from papers using cross-sectional data and 40% from panel data studies.

It emerges that the efficiency scores is 0.783 for the studies that use panel data, while is 0.732 for cross-sectional studies. A similar difference in mean is found between the mean efficiency reported in published papers and that of unpublished work (0.774 - 0.643 = 0.131). When considering the sub-sample of non-parametric frontiers, we find that papers assuming variable returns to scale (VRS) register a mean efficiency of 0.757, which is higher than 0.699, that is average retrieved from papers working under the hypothesis of constant returns to scale (CRS). Moreover, the difference between the average efficiency scores for estimations referring to European countries and those for estimations related to non-European countries is statistically significant (0.780 - 0.697 = 0.083).

As regards the structure of the degree of judgement, it should be noted that 225 out of 264 observations refer to the judgement of "first instance". The test to determine whether the difference in means group indicates that the papers on first-instance courts yields efficiency scores that do not differ from those obtained in papers focusing on appeal courts. Finally, there is a significant difference in average efficiency if we consider specialised courts in a specific subject (0.769) or generalist courts (0.726).

If on one hand, the mean of court efficiency differs by group (Table 1), it is interesting to consider the entire efficiency distribution. Figure 2 shows not only that there is much variability group-by-group, but also that the efficiency score distributions have different shapes and forms. All this indicates that any choice made by researchers in performing a court efficiency study significantly affects final estimations.

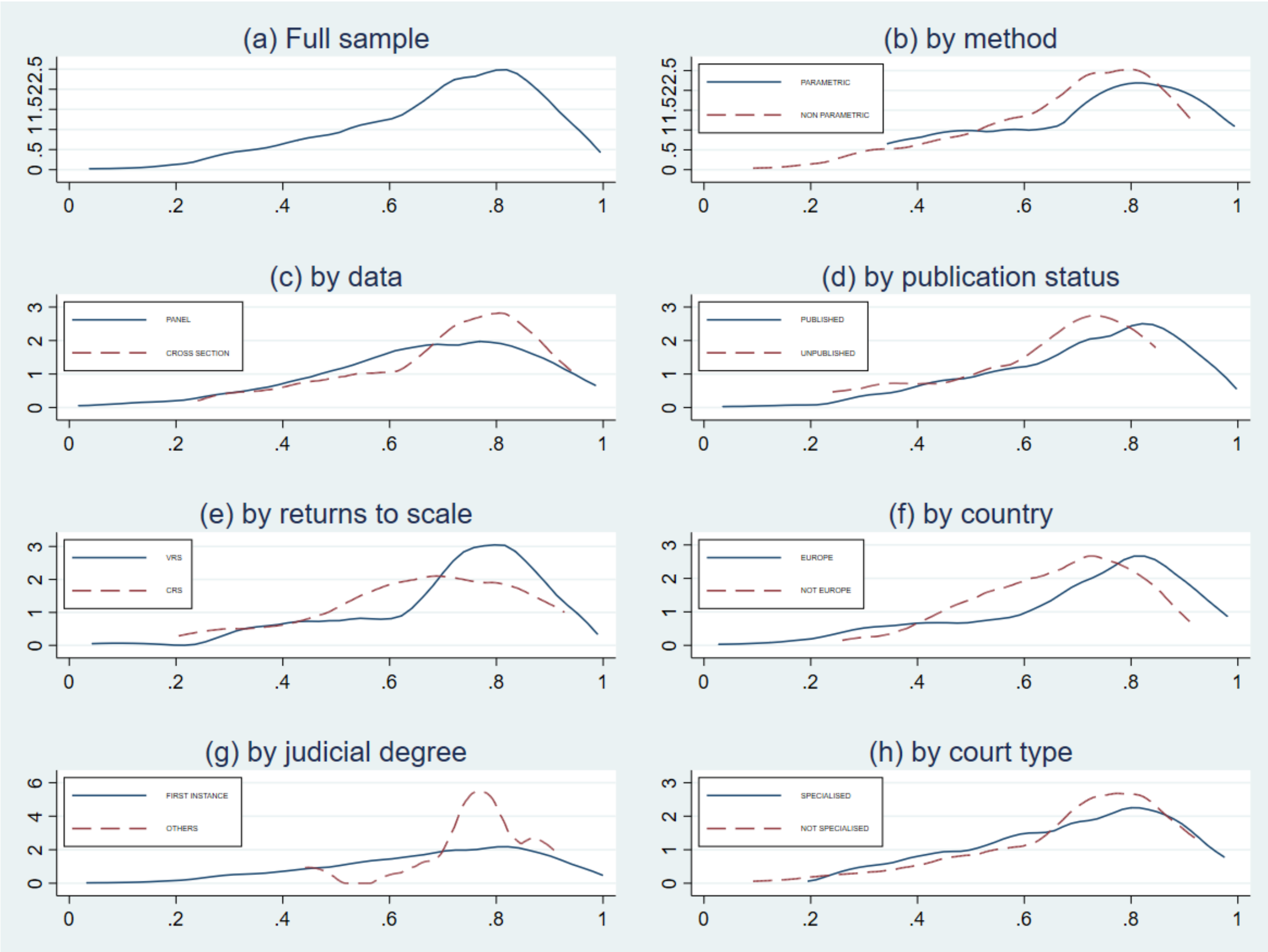
From this discussion, it emerges that the study design of primary papers plays an important role in determining differences, not only in the means but also in the distributions of judicial efficiency scores.

Table 1. Average, standard deviation and number of observations of efficiency scores, by group

	Mean	**SD**	**Obs**
Full sample	0.752	0.195	264
Estimation approach			
Non-parametric	0.731	0.193	227
Parametric	0.885	0.153	37
Test on difference in means (p-value)	**0.000**		
Data type			
Cross-section	0.732	0.181	159
Panel	0.783	0.212	105
Test: Difference in means	**0.039**		
Publication status			
Unpublished	0.643	0.171	44
Published	0.774	0.193	220
Test: Difference in means	**0.000**		
Returns to scale in non-parametric studies			
CRS	0.699	0.199	104
VRS	0.757	0.184	123
Test: Difference in means	**0.024**		
Country			
Non-Europe	0.697	0.169	88
Europe	0.780	0.202	176
Test: Difference in means	**0.001**		
Judicial degree			
Other instances	0.764	0.120	39
First degree	0.750	0.206	225
Test: Difference in means	0.692		
Type of courts			
Non-specialised	0.769	0.196	160
Specialised	0.726	0.193	104
Test: Difference in means	**0.083**		

Source: Authors' elaboration.
Note 1: means are unweighted
Note 2: t-test for difference in means: bold values of the p-value mean that the difference is statistically significant.

Figure 2. Heterogeneity in the judicial efficiency literature



Source: Authors' elaboration.

4. Meta-regressions for judicial efficiency literature

The literature on efficiency proves that the results are highly heterogeneous across studies, whatever the sector.[3] As we have shown before, this holds true for judicial literature, thereby motivating a meta-study aimed to detect the factors that explain such a variability. To this scope, let's consider the following model:

$$E_i = \beta_1 + \sum_j \beta_j X_{ij} + \varepsilon_i \quad [1]$$

where the dependent variable E_i is the i-*th* efficiency score from each primary paper. The matrix X_{ij} comprises the explanatory variables that summarise various model characteristics of every study. The error component ε_i is heteroscedastic because each estimated efficiency scores E_i retrieved by the primary papers has its own measure of variability, leading to the heteroscedasticity of the distribution of the error term ε_i. To address this issue, the weight we use is the standard deviation (S) of each observation of our meta-study. Thus, the estimated meta-regression is:

$$\frac{E_i}{S_i} = \beta_1 \frac{1}{S_i} + \sum_j \beta_j \frac{X_{ij}}{S_i} + e_i \quad [2]$$

$$E_i^* = \beta_1 S_i^* + \sum_j \beta_j X_{ij}^* + e_i \quad [3]$$

where the disturbance $e = \varepsilon/S$ is corrected for heteroscedasticity.

The method used in estimating eq. [3] may be a fixed effects or random-effects model. These methods differ in terms of treatment of heterogeneity. In detail, a fixed effects meta-regression assumes that all the heterogeneity is explained by the covariates and leads to excessive type I errors when there is residual, or unexplained, heterogeneity (Thompson and Sharp, 2012; Harbord and Higgins, 2008). On the contrary, a random-effects meta-regression allows for such residual heterogeneity (the between-study variance not explained by the covariates), thereby extending the fixed effects model. Formally, under the random-effects framework, the eq. [3] becomes:

$$E_i^* = \beta_1 S_i^* + \sum_j \beta_j X_{ij}^* + u_i + e_i \quad [4]$$

where $e_i \sim N(0, \sigma^2{}_i)$ is the within-study disturbance and $u_i \sim N(0, \tau^2)$ is the deviation due to the residual non-observable heterogeneity (between-study variance). The parameter τ^2 is a measure of between-study variability and is estimated as in Harbord and Higgins (2008).[4]

We adopt a two-step procedure as proposed in Gallet and Doucouliagos (2014) and applied in Aiello and Bonanno (2018; 2019). A Random Effect Maximum Likelihood (REML) regression is run in the first step, and, in the second step, we run a WLS regression in which the

[3] High heterogeneity in efficiency scores is found by Bravo-Ureta et al. (2007) farm performance. The same applies in Ogundari (2014), Thiam et al. (2001) and Djokoto and Gidiglo (2016) for agricultural efficiency, Iliyasu et al. (2014) for aquaculture, Fan et al. (2018) for water use efficiency, Brons et al. (2005) for urban transport, Odeck and Bråthen (2012) for seaports, Nguyen and Coelli (2009) and Kiadaliri et al. (2013) for hospital efficiency, Havránek and Iršová (2010) and Aiello and Bonanno (2018) for banking, Assaf and Josiassen (2016) for tourism, Aiello and Bonanno (2019) for municipalities, Tian et al. (2012) for productivity growth in China and finally in Simões and Marques (2012) for waste sector.

[4] Technically, the second step of estimations employs $1/(\sigma_i^2 + \tau^2)$ as weights, where σ_i^2 is the standard deviation of the estimated effect in study *i* and τ^2 is the between-study variability estimated in the first step.

weights also include the value of τ^2 retrieved from the first step. This ensures that the estimates will be robust to clustering at study level.

4.2 The explanatory variables in the meta-regression analysis

The explanatory variables that enter in the regression are identified by following the meta-studies that review the efficiency literature in specific sectors (see footnote 3). These variables can be classified in three groups, which are discussed in the following.

The variables included in the first group are retrieved from prior literature on meta-regression analysis. When the interest is in the approaches and methods used to estimate the frontier, a first control comes from distinguishing the primary papers applying a parametric approach and the non-parametric studies. To this end, the dummy variable used is Parametric, which is equal to unity for studies using parametric methods and 0 for the others. The dummy variable Panel is equal to unity if the original works used panel data, and 0 otherwise. Including the variable Year of publication of the primary paper provides information on how the estimated efficiency scores vary over time. The variable *Published* is a dummy equal to 1 if the paper is published in a peer-review journal, and 0 otherwise. Furthermore, the variable *Dimension* is the sum (in logs) of inputs and outputs used in the primary efficiency studies, while the variable *Sample Size* (in logs) is the number of observations, that is the number of courts analysed in each primary paper.[5]

The second group of regressors takes into consideration some specific factors related to the judicial literature. To this end we employ two dummies. The first is the dummy *Specialised courts*, which is equal to 1 if the primary paper focuses on a specific type of court (commercial, tax, civil, criminal) and 0 if estimations refer to more than one type of court or to a national judicial system as a whole. The second dummy is *First instance*, which is equal to 1 if the primary paper investigates the efficiency of courts of first instance and 0 when the efficiency is done for appeal courts.

Finally, the third group of regressors is to control for potential country-specific effects. They are the GDP per capita and the indicator *Legal enforcement of contracts*. The GDP per capita refers to the country analysed in the primary paper and is from the World Bank database (data are in PPP-constant 2011 dollars). The variable *Legal enforcement of contracts*, is used as a proxy for the quality of the legal system of each country, and it is from the Fraser Institute reports. It is measured through a scale—ranging from 0 to 10—for which higher values indicate better performance. Finally, the dummy *Europe* - which is equal to 1 if the study used data from a European country (the reference group comprises efficiency scores from papers focusing on the rest of the world) serves to control for unobservable territorial fixed-effects.

5. Estimated models

In presenting the results, we depart from Table 2 that reports the estimations of models including the variables related to the study design. We show different specifications gradually adding the regressors retrieved from the literature of meta-regression on efficiency scores.

[5] It is important to point out that the different ways of performing an efficiency study make conclusive expectations of the impact of each meta regressor difficult. In other words, despite the high degree of specialisation of efficiency scholars, the effect of some methodological choices remains uncertain. For instance, results in parametric papers may be higher or lower than that obtained in non-parametric papers, depending on the nature of disturbances from the frontier (Nguyen and Coelli, 2009). Again, panel data would yiled higher efficiency levels than cross-sectional data. Finally, efficiency would increase with the number of variables included in the frontier, while it would decrease with small sample sizes (Berger and Humphrey, 1997; Coelli, 1995; Fethi and Pasourias, 2010; Nguyen and Coelli, 2009). However, while theory predicts the likely impact of any choice, the actual measure of how sensitive the results are to the study design is an issue to be addressed empirically.

Table 3 displays the findings when using two sector-specific variables and two country observables.

As presented in Table 2, the proportion of the residual variance due to between-study heterogeneity is very high: it is 74.04%, while the proportion of between-variance explained by the covariates is 19% (the within-study sampling variability). As expected, when we include additional regressors, the between-study heterogeneity tends to decrease (in Model 4 of Table 2 it is equal to 27.40%), while the proportion of between-variance explained by the covariates tends to increase (in Model 4 it becomes 52.36%). In some specifications of Table 3, the residual variance due to the between-study heterogeneity falls to about 12%, thereby signalling that the models explain a very high proportion of variability of courts' efficiency. Finally, the statistic F-Fisher indicates that the joint significance of the explanatory variables is high in each regression.

To ensure clarity in the exposition, the discussion of results is divided into two sub-sections. Sub-section 5.1 is devoted to the baseline regressions, which include the explanatory variables related to the study design. Sub-section 5.2 focuses on specific variables for judicial efficiency and sub-section 5.3 shows results of the effects exerted by the country observables.

5.1 Estimation methods and other econometric issues

The estimated coefficient β_2 associated with *Year of publication* is always negative and significant (with the exception of Model 5 of Table 3). This indicates that the estimated efficiency scores of courts resulting from primary papers decrease over time. Estimates referring to published primary papers yield, on average, higher efficiency scores compared to estimations retrieved from working papers. The coefficient β_3 is always positive and significant.

Our regressions indicate that the method used to model court frontiers matters: the coefficient β_4 associated with the dummy *Parametric* is significant and positive in all specifications we estimate. In other words, parametric studies achieve a higher efficiency level than non-parametric estimations: on average, the differential impact of the parametric method is 0.1344 in column (2) and 0.1437 in column (3). The conclusion from this meta-study is that the method type matters in explaining the heterogeneity in courts' efficiency. This evidence is in line with other meta-studies, although the sign of the impact differs sector by sector. For instance, Nguyen and Coelli (2009) find a significant and negative coefficient for parametric studies in health care. Conversely, the effect is positive in Bravo-Ureta et al. (2007) for the agricultural efficiency in developed and developing economies, Odeck and Bråthen (2012) for efficiency in seaports, Aiello and Bonanno (2018) for banking and Aiello and Bonanno (2019) for local government. It is also worth pointing out that the parametric effect is neutral in Thiam et al. (2001) for agriculture in developing countries, Nguyen and Coelli (2009) for hospitals, Brons et al. (2005) for transport and in Oganduri and Brümmer (2011) for Nigerian agriculture.

We do not find conclusive results about the data type: the coefficient β_5 associated with the dummy *Panel* is not significant in Table 2, while it is significantly negative in some models of Table 3. Finally, there is no difference between the efficiency scores when distinguishing by geographical area (the coefficient of *Europe*, β_6, is always not significant).

As far as the role of *Dimension* is concerned, it emerges that $\hat{\beta}_7$ is negative, thereby indicating that an increase of inputs and outputs used in estimating the frontier is associated to a decrease, on average, of court efficiency.[6] Our estimates allow us to calculate the marginal effect on mean efficiency at different values of *Dimension*. Considering the use of logs, when

[6] This result on judicial efficiency contrasts with Aiello and Bonanno (2018; 2019), Nguyen and Coelli (2009), Ogundari and Brümmer (2011) and Thiam et al. (2001), who all found a positive impact of *Dimension* on the efficiency of the sector they meta-review.

Dimension is equal to the overall mean of the sample (1.680), the marginal effect is -0.0193 (given that $\beta_7 =$ -0.0325 in Model 3 of Table 2, the marginal effect is measured as the ratio between -0.0325 and 1.680). Again, when the number of inputs and outputs is 10, the marginal effect of *Dimension* on the variability of courts' efficiency is approximately 0 (Figure 3, Panel a).

The estimated coefficient of the relationship between courts' efficiency and the number of observations provides further interesting findings. Size may exert a non-linear effect and thus it enters into regressions in logs. The coefficient $\hat{\beta}_8$ is negative: the link between *Sample Size* and the efficiency scores of primary papers is, on average, negative. However, the magnitude of the marginal effect is very small, being -0.001 when the *Sample Size* is 40 and quickly tends to zero if courts are about 100 (Figure 3, Panel b).

In the last column of Table 2, we estimate Model (4) only for non-parametric primary estimations by adding the dummy D_{VRS}. The focus on this sub-sample corroborates all findings reported in column (4). Importantly, we find that assuming that the technology in producing judicial services exhibits constant or variable returns to scale does not help in explaining why efficiency scores differ paper by paper (the coefficient $\hat{\beta}_9$ associated with D_{VRS} is, indeed, not statistically significant).

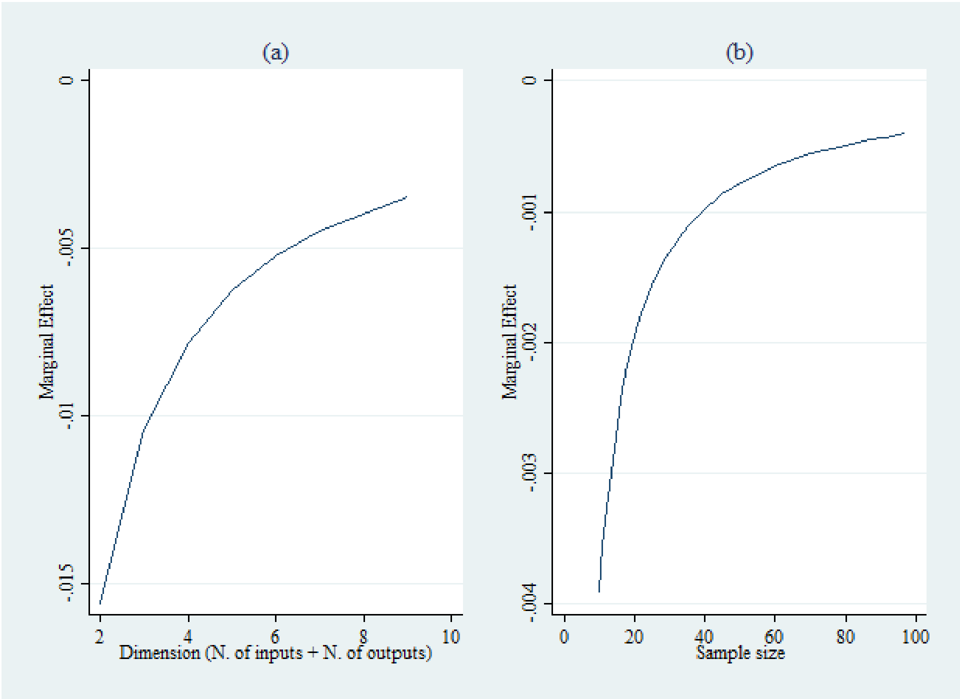
Table 2. Meta-regression analysis of judicial efficiency scores. The role of the study design[1]

Variables		Model 1		Model 2		Model 3		Model 4	
Constant	β_1	8.2169	***	13.9434	***	14.3295	***	15.456 2	***
Year of publication	β_2	-0.0037	***	-0.0066	***	-0.0067	***	-0.0072	***
Published	β_3	0.1838	***	0.1441	***	0.1644	***	0.1478	***
Parametric	β_4			0.1344	***	0.1437	***		
Panel	β_5			-0.0058		0.0001		-0.0219	
Europe	β_6			0.0066		0.0073		-0.0126	
Dimension	β_7					-0.0325	*	-0.0055	
Sample Size	β_8					-0.0430	***	-0.0406	***
D_{VRS}	β_9							-0.0109	
Observations		264		264		264		227	
tau² (between-study variance)		0.0078		0.0050		0.0038		0.0053	
% residual variation due to heterogeneity		74.04%		61.21%		53.79%		27.40%	
Adj R-squared		18.62%		47.95%		60.73%		52.36%	
F- Fisher		21.64		20.58		20.85		12.67	

Legend: * p < 0.2; ** p < 0.1; *** p < 0.05.
Note: The statistical significance of the REML results is robust to clustering at the study level, as in Gallet and Doucouliagos (2014). Model 5 refers only to non-parametric estimations from the primary papers.

(1) Estimates refer to eq. 4 with the standard deviation of each point-observation, that is the efficiency score, used as weights to control for heteroschedasticity.

Figure 3. Marginal effects of *Dimension* (inputs plus outputs) and *Sample Size* of primary papers



5.2 The role of specific factors in the judicial literature

Table 3 displays the results for specifications with the two sector-specific variables *First instance* and *Specialised courts*. Importantly, augmenting the meta-regressions with court-specific factors does not have an impact on the estimates of the explanatory variables so far discussed.

We firstly introduce one by one the variables *First instance* and *Specialised courts*, and then they are in the full Model 3 of Table 3. The evidence is robust and indicates a positive impact for both *First instance* and *Specialised courts*.

The ev9idence is that papers focusing on courts in the first instance provide higher efficiency scores than papers analysing the efficiency of appeal courts. This probably is due to the incentive for this type of tribunal to be more efficient because it has to timely treat a massive number of disputes.

Furthermore, the primary studies on specialised courts yield a higher level of efficiency than the papers referring to more than one type of court or to a national judicial system as a whole. These results might depend on the homogeneity of the sample of courts used in estimating the frontier: it is expected that the efficiency scores from homogenous courts are higher than those obtained when considering heterogeneous samples (i.e., data of courts dealing with different judicial topics and different judgement levels). Differently said, similar courts perform similarly with the result that they are more clustered around a frontier than different courts with divergent aims.

5.3 The impact of country observables

Final considerations regard the country observables included in the meta-regression. As far as GDP per capita is concerned, the expected results are uncertain. Indeed, courts' efficiency would be high in developed countries, as higher-income residents pay greater taxes and have more requirements on services and facilities. On the other hand, in high-income countries courts are expected to manage higher financial resources, thereby implying that the efforts to be efficient might be reduced. The positive and significant coefficient (around 0.04 in Model (4) of Table 3) indicates that the efficiency scores of courts in rich countries are, on average, higher than the studies focusing on less rich countries. This is in line with the argument that people ling in developed nations expect to receive good services from public institutions because they pay more taxes.

As far as the specific theme of the meta-study is concerned, valuable evidence is obtained from Model 5 of Table 3. Indeed, the estimated coefficient of *Legal enforcement of contracts* is positive (0.0093) and highly significant, confirming the role of the institutional context in explaining how the courts effectively work when offering judicial services. It is proven that primary papers focusing on countries with high-quality legal systems yield, on average, higher efficiency scores than studies focusing on countries with weak legal systems. Differently phrased, in countries with high levels of enforcement of contracts the disputes are easy to be fulfilled and, thus, courts work in a timely and efficient manner. The opposite holds in countries with low enforcement of contracts. In such a case, disputes are complex and difficult to be fulfilled, thereby having an impact on courts' efficiency. In a nutshell, institutions differ country-by-country and thus it is expected that these differences translate into heterogeneity in efficiency scores of courts when performing a frontier analysis.

Table 3. Meta-regression analysis of judicial efficiency score
The role of sector-specific explanatory variables and country observables[1]

Variables		Model 1		Model 2		Model 3		Model 4		Model 5	
Constant	β_1	17.3385	***	15.7791	***	19.0603	***	13.3042	***	-2.5195	
Year of publication	β_2	-0.0082	***	-0.0074	***	-0.0091	***	-0.0065	***	0.0015	
Published	β_3	0.1504	***	0.1750	***	0.1604	***	0.1708	***	0.1655	***
Parametric	β_4	0.1533	***	0.1394	***	0.1494	***	0.1126	***	0.1033	***
Panel	β_5	-0.0291	*	-0.0165		-0.0490	***	-0.0402	***	-0.0536	***
Europe	β_6	-0.0061		-0.0154		-0.0313	*	-0.0002		0.0352	
Dimension	β_7	-0.0246		-0.0536	***	-0.0453	***	-0.0405	***	-0.0509	***
Sample Size	β_8	-0.0438	***	-0.0332	***	-0.0333	***	-0.0207	***	-0.0209	***
Specialised court	β_9	0.0524	***			0.0558	***	0.0543	***	0.0696	***
First instance	β_{10}			0.0822	***	0.0870	***	0.1298	***	0.0893	**
Log (GDP per capita)	β_{11}							0.0401	***	0.0179	
Legal enforcement of contracts	β_{12}									0.0093	**
Observations		264		264		264		253		218	
tau^2 (between-study variance)		0.0035		0.0032		0.0029		0.0001		0.0001	
% residual variation due to heterogeneity		50.23%		50.20%		48.39%		12.23%		23.55%	
Adj R-squared		64.15%		66.84%		69.44%		98.51%		98.93%	
F- Fisher		19.64		20.34		19.47		43.20		38.04	

Legend: * p < 0.2; ** p < 0.1; *** p < 0.05.
Note: The statistical significance of the REML results is robust to clustering at the study level, as in Gallet and Doucouliagos (2014).
(1) Estimates refer to eq. 4 with the standard deviation of each point-observation, that is the efficiency score, used as weights to control for heteroschedasticity.

6. Conclusions and future work

This paper presents a meta-study on the judicial efficiency literature with the aim of detecting the heterogeneity in prior results and explaining the role of the methodological choices on efficiency scores. The study motivations are twofold.

The first refers to the huge debate on the reliability of the findings obtained from efficiency studies and, in particular, how the methods and frontier specifications influence the estimates of court efficiency scores. This is a line of research widely explored in the scientific community of efficiency analysts, with the result that a number of meta-reviews have been recently performed by focusing on specific sectors.

The second motivation lies in our interest in understanding whether the courts optimise the use of resources in offering judicial services. As in many other public sectors, the use of frontier methods in studying the courts' efficiency allows us to identify the best practices and helps policymakers to implement proper responses to increase the sectoral efficiency. In this regard, the more accurate the efficiency scores, the easier and more effective the reforms of judicial institutions.

Here it is also worth mentioning that solving the problem associated with the measurement and assessment of court efficiency is necessary for management because of the relatively high amount of public expenditure on justice, in conjunction with the time which courts need for issuing judgements in cases (Major, 2015).

In explaining the huge variability in courts' efficiency scores, this meta-study uses 264 observations of courts' efficiency from 36 primary studies published from 1992 to 2019.

Compared to other qualitative surveys (Marciano et al., 2019; Voigt, 2016), the paper contributes by identifying the direction and quantifying the size of the effects that methodological choices exert on the mean efficiency scores of courts. We prove that much of the variability of the judicial efficiency scores is due to the study design of each primary paper.

For instance, it is found that primary papers using parametric methods yield higher efficiency scores than non-parametric studies. There is also significant evidence to support the hypothesis that efficiency scores decrease as the number of observations increases. Because of non-linearity, the effect of size is more pronounced when the sample size is small and its marginal effect quickly tends to zero when sample size in primary papers comprises about 100 courts. Similarly, courts' efficiency scores fall as the number of inputs and outputs increases. In such a case, the marginal effect of dimension is very small (although it remains negative) when the sum of inputs and outputs is about 10. Furthermore, meta-regressions suggest the efficiency estimates from published primary papers are higher than those reported in working papers. The effects of other modelling choices, such as the assumption on the return to scale, are not found to be statistically significant at any level; moreover, the effect of data (panel vs cross-section) is not always robust but depends on the specification of the meta-regressions.

Heterogeneity in results of primary papers also depends on the type of courts that researchers include in their efficiency study. This meta-analysis demonstrates that efficiency scores from specialised papers, that is, papers that focus on one court type (commercial, tax, administrative, criminal), are significantly higher than results from unspecialised papers. A similar evidence emerges when considering the level of judgement: the efficiency of papers focusing on first-instance courts is statistically higher than that provided by papers analysing appeal courts or mixed samples (first-instance and appeal courts). This might be driven by the fact that the efficiency analyses focusing on a specific sample of courts (first-instance or sector-specialised courts) benefit from data homogeneity compared to primary samples of mixed courts. With homogenous data, it is expected that the clustering around the frontier of courts pursuing the same scope is pronounced, thereby translating into high-efficiency estimates.

While the key results of this meta-study are robust to different samples of observations, the analysis has some limitations depending on data quality. Differently said, a number of areas remains for future work.

The first limitation is due to the circumstance that scholars do not report details of the empirical setting. This hidden information in primary papers impedes replicability. Therefore, providing full explanations would be a good practice not only to let be readers informed on each study, but also to understand better the role of crucial choices that researchers made when estimating a frontier. For instance, it would be valuable for practitioners and researchers know if heterogeneity in judicial efficiency might be explained by orientation in technology (input- vs output-oriented models). Similarly, the meta-dataset signals that there has been a predominance of non-parametric techniques (DEA in particular), whereas the parametric approach is disregarded by scholars. This is not neutral. Indeed, taking into consideration evidence proven in this meta-study, i.e. that parametric studies yield on average higher efficiency scores than non-parametric papers, the massive use of DEA in the judicial literature leads courts to be perceived as less efficient than they effectively are.

An additional need regards the geographical coverage of efficiency studies. The literature counts only 36 studies. This does not reflect the relevance of the judicial efficiency over the world. Therefore, this MRA recommends that future work would focus more on the courts efficiency of the countries that have so far disregarded (i.e., Japan, the US, Germany, the UK).

References

Aiello, F., and Bonanno, G. (2018). On the sources of heterogeneity in banking efficiency literature. *Journal of Economic Surveys*, *32*, 194–225.

Aiello, F., and Bonanno, G. (2019). Explaining Differences in Efficiency: A Meta-Study on Local Government Literature. *Journal of Economic Surveys*, *33*, 999–1027.

Antonucci, L., Crocetta, C., and d'Ovidio, F. D. (2014). Evaluation of Italian Judicial System. *Procedia Economics and Finance*, *17*, 121–130.

Assaf, A. G., and Josiassen, A. (2014). Frontier Analysis: A State-of-the-Art Review and Meta-Analysis. *Journal of Travel Research*, *55*, 612–627.

Berger, A. N., and Humphrey, D. B. (1997). Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research*, *98*, 175–212.

Bravo-Ureta, B. E., Solís, D., Moreira López, V. H., Maripani, J. F., Thiam, A., and Rivas, T. (2007). Technical efficiency in farming: A meta-regression analysis. *Journal of Productivity Analysis*, *27*, 57–72.

Brons, M., Nijkamp, P., Pels, E., and Rietveld, P. (2005). Efficiency of urban public transit: A meta-analysis. *Transportation*, *32*, 1–21.

Castro, A. S. (2009). *Court Performance in Brazil: Evidence from Judicature-Level Data* (No. ID 2612941). Rochester, NY: Social Science Research Network.

Coelli, T. J. (1995). Recent Developments in Frontier Modelling and Efficiency Measurement. *Australian Journal of Agricultural Economics*, *39*, 219–245.

Coelli, T., and Perelman, S. (1999). Comparison of parametric and non-parametric distance functions: With application to European railways. *European Journal of Operational Research*, *117*, 326–339.

Djokoto, J. G., and Gidiglo, K. F. (2016). Technical Efficiency in Agribusiness: A Meta-Analysis on Ghana. *Agribusiness*, *32*, 397–415.

Elbialy, N., and García-Rubio, M. A. (2011). Assessing Judicial Efficiency of Egyptian First Instance Courts: A DEA Analysis. In *MAGKS Papers on Economics* (No. 201119). Philipps-

Universität Marburg, Faculty of Business Administration and Economics, Department of Economics (Volkswirtschaftliche Abteilung).
Espasa, M., and Esteller-Moré, A. (2015). Analyzing judicial courts' performance: Inefficiency vs. Congestion. *Revista de Economia Aplicada*, *23*, 61–82.
Falavigna, G., Ippoliti, R., Manello, A., and Ramello, G. B. (2015). Judicial productivity, delay and efficiency: A Directional Distance Function (DDF) approach. *European Journal of Operational Research*, *240*, 592–601.
Falavigna, G., Ippoliti, R., and Ramello, G. B. (2018). DEA-based Malmquist productivity indexes for understanding courts reform. *Socio-Economic Planning Sciences*, *62*, 31–43.
Fan, Y., Wang, C., and Nan, Z. (2018). Determining water use efficiency of wheat and cotton: A meta-regression analysis. *Agricultural Water Management*, *199*, 48–60.
Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, *120*, 253–281.
Fauvrelle, T. A., and Almeida, A. T. C. (2018). Determinants of Judicial Efficiency Change: Evidence from Brazil. *Review of Law & Economics*, *14*.
Ferrandino, J. (2012). The Impact of Revision 7 on the Technical Efficiency of Florida's Circuit Courts. *Justice System Journal*, *33*, 22–46.
Ferro, G., Romero, C. A., and Romero-Gómez, E. (2018). Efficient courts? A frontier performance assessment. *Benchmarking: An International Journal*, *25*, 3443–3458.
Fethi, M. D., and Pasiouras, F. (2010). Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *European Journal of Operational Research*, *204*, 189–198.
Finocchiaro Castro, M., and Guccio, C. (2014). Searching for the source of technical inefficiency in Italian judicial districts: An empirical investigation. *European Journal of Law and Economics*, *38*, 369–391.
Finocchiaro Castro, M., and Guccio, C. (2018). Measuring Potential Efficiency Gains from Mergers of Italian First Instance Courts through Nonparametric Model. *Public Finance Review*, *46*, 83–116.
Fusco, E., Laurenzi, M., and Maggi, B. (2018). A data envelopment analysis of the Italian judicial efficiency. In *DSS Empirical Economics and Econometrics Working Papers Series* (No. 2018/2). Centre for Empirical Economics and Econometrics, Department of Statistics, 'Sapienza' University of Rome.
Gallet, C. A., and Doucouliagos, H. (2014). The income elasticity of air travel: A meta-analysis. *Annals of Tourism Research*, *49*, 141–155.
Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*, 3–8.
Glass, G. V., McGaw, B., and Smith, M. L. (1981). Meta-analysis in Social Research. *Sage Publications*, *Beverly Hills*.
Gorman, M. F., and Ruggiero, J. (2009). Evaluating U.S. judicial district prosecutor performance using DEA: are disadvantaged counties more inefficient? *European Journal of Law and Economics*, *27*, 275–283.
Guzowska, M., and Strak, T. (2010). An Examination of the efficiency of Polish public sector entities based on public prosecutor offices. *Operations Research and Decisions*, *2*, 41–57.
Harbord, R. M., and Higgins, J. P. T. (2008). Meta-regression in Stata. *Stata Journal*, *8*, 493–519.
Havranek, T., Irsova, Z., and Janda, K. (2012). Demand for gasoline is more price-inelastic than commonly thought. *Energy Economics*, *34*, 201–207.
Iliyasu, A., Mohamed, Z. A., Ismail, M. M., and Abdullah, A. M. (2014). A Meta-Analysis of Technical Efficiency in Aquaculture. *Journal of Applied Aquaculture*, *26*, 329–339.

Ippoliti, R., Melcarne, A., and Ramello, G. B. (2015a). Judicial efficiency and entrepreneurs' expectations on the reliability of European legal systems. *European Journal of Law and Economics*, *40*, 75–94.
Ippoliti, R., Melcarne, A., and Ramello, G. B. (2015b). The Impact of Judicial Efficiency on Entrepreneurial Action: A European Perspective. *Economic Notes*, *44*, 57–74.
Ippoliti, R., and Ramello, G. B. (2018). Governance of tax courts. *Economics of Governance*, *19*, 317–338.
Ivanova, N. B. (2014). Judicial reform and its effect on the economy: The case of Bulgaria. *Master Thesis*.
Kiadaliri, A. A., Jafari, M., and Gerdtham, U.-G. (2013). Frontier-based techniques in measuring hospital efficiency in Iran: A systematic review and meta-regression analysis. *BMC Health Services Research*, *13*.
Kittelsen, S. A. C., and Førsund, F. R. (1992). Efficiency analysis of Norwegian district courts. *Journal of Productivity Analysis*, *3*, 277–306.
Major, W. (2015). Data Envelopment Analysis as an instrument for measuring the efficiency of courts. *Operations Research and Decisions*, *4*, 19–34.
Marciano, A., Melcarne, A., and Ramello, G. B. (2019). The economic importance of judicial institutions, their performance and the proper way to measure them. *Journal of Institutional Economics*, *15*, 81–98.
Mattsson, P., and Tidanå, C. (2019). Potential efficiency effects of merging the Swedish district courts. *Socio-Economic Planning Sciences*, *67*, 58–68.
Melcarne, A., and Ramello, G. (2015). Judicial Independence, Judges' Incentives and Efficiency. *Review of Law & Economics*, *11*, 149-169.
Nguyen, K. H., and Coelli, T. (2009). Quantifying the Effects of Modelling Choices on Hospital Efficiency Measures: A Meta-Regression Analysis. *Centre for Efficiency and Productivity Analysis*.
Nissi, E., Giacalone, M., and Cusatelli, C. (2019). The Efficiency of the Italian Judicial System: A Two Stage Data Envelopment Analysis Approach. *Social Indicators Research*, *146*, 395–407.
Nissi, E., and Rapposelli, A. (2010). A Data Envelopment Analysis of Italian courts efficiency. *Statistica Applicata Italian Journal of Applied Statistics, 22(2)*, 199–210.
Odeck, J., and Bråthen, S. (2012). A meta-analysis of DEA and SFA studies of the technical efficiency of seaports: A comparison of fixed and random-effects regression models. *Transportation Research Part A: Policy and Practice*, *46*, 1574–1585.
Odhiambo, O. J. (2014). Technical Efficiency of the Kenyan Judiciary: A case of the magistrates' courts. *Master Thesis*.
Ogundari, K. (2014). The Paradigm of Agricultural Efficiency and its Implication on Food Security in Africa: What does meta-analysis Reveal? *World Development*, *64*, 690–702.
Pedraja-Chaparro, F., and Salinas-Jimenez, J. (1996). An assessment of the efficiency of Spanish Courts using DEA. *Applied Economics*, *28*, 1391–1403.
Peyrache, A., and Zago, A. (2016). Large courts, small justice! The inefficiency and the optimal structure of the Italian justice sector. *Omega*, *64*, 42–56.
Rushid, A. R. (2018). Technical efficiency of Swedish district courts. *Master Thesis*.
Santos, S. P., and Amado, C. A. F. (2014). On the need for reform of the Portuguese judicial system – Does Data Envelopment Analysis assessment support it? *Omega*, *47*, 1–16.
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*, 127–144.
Silva, M. C. A. (2018). Output-specific inputs in DEA: An application to courts of justice in Portugal. *Omega*, *79*, 43–53.

Simões, P., and Marques, R. C. (2012). Influence of regulation on the productivity of waste utilities. What can we learn with the Portuguese experience? *Waste Management*, *32*, 1266–1275.
Sousa, M. de M., Guimaraes, T. A., Sousa, M. de M., and Guimaraes, T. A. (2018). Resources, innovation and performance in labor courts in Brazil. *Revista de Administração Pública*, *52*, 486–506.
Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *Journal of Economic Perspectives*, *15*, 131–150.
Thiam, A., Bravo-Ureta, B. E., and Rivas, T. E. (2001). Technical efficiency in developing country agriculture: A meta-analysis. *Agricultural Economics*, *25*, 235–243.
Thompson, S. G., and Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine*, *18*, 2693–2708.
Tian, X., and Yu, X. (2012). The Enigmas of TFP in China: A meta-analysis. *China Economic Review*, *23*, 396–414.
Tsai, C., and Tsai, J. (2010). Performance Evaluation of the Judicial System in Taiwan Using Data Envelopment Analysis and Decision Trees. *2010 Second International Conference on Computer Engineering and Applications*, *2*, 290–294.
Voigt, S. (2016). Determinants of judicial efficiency: A survey. *European Journal of Law and Economics*, *42*, 183–208.
Yeung, L. (2014). Measuring Efficiency of Courts: An Assessment of Brazilian Courts Productivity. In A. Emrouznejad and E. Cabanda (Eds.), *Managing Service Productivity: Using Frontier Efficiency Methodologies and Multicriteria Decision Making for Improving Service Performance* (pp. 155–165). Berlin, Heidelberg: Springer.
Yeung, L., and Azevedo, P. (2011a). *Measuring the Efficiency of Brazilian Courts from 2006 to 2008: What Do the Numbers Tell Us?.* Insper Working Paper.
Yeung, L. L., and Azevedo, P. F. (2011b). Measuring efficiency of Brazilian courts with data envelopment analysis (DEA). *IMA Journal of Management Mathematics*, *22*, 343–356.