CATÓLICA
CEGE · CENTRO DE ESTUDOS EM GESTÃO E ECONOMIA
PORTO

Courts performance: The case of Portugal

Nuno Q. M. Lopes
Maria C. A. Silva

CEGE WORKING PAPER SERIES WP23.04
December 2023

fct Fundação para a Ciência e a Tecnologia

This work was supported by Fundação para a Ciência e Tecnologia (through project UIDB/00731/2020).

Courts performance: The case of Portugal

Nuno Q.M. Lopes[1], Maria C. A. Silva [1]

December 13, 2023

1- Universidade Católica Portuguesa, Católica Porto Business School, CEGE, Porto, Portugal
csilva@ucp.pt

Abstract

For some time now, the judicial systems of countries have been intrinsically connected to their development. This analysis uses data from 2015 to 2021 on every bench in the Portuguese first instance courts, consisting of 3,249 observations and 12 different variables. To analyze the efficiency of benches, Data Envelopment Analysis was used and scale and scope economies were investigated. The results showed that the efficiency of benches was very balanced throughout the country and over the years. Given the clear difference between specialized and non-specialized benches, we assessed benches within and between groups and computed the group frontier gap. This allowed us to draw conclusions regarding the economies of scope of specialized vs generic benches. The results uncovered diseconomies of scope (generic benches are less efficient than specialised ones), and diseconomies of scale with some specialised benches experiencing decreasing returns to scale (criminal, enforcement and commerce benches) while others experienced increasing returns (Civil and generic benches).

Data envelopment analysis, Benchmarking, Courts Efficiency, Economies of Scope.

1 Introduction

Courts and countries' judicial systems are intrinsically connected to the economic development and growth of nations (Ramello and Voigt, 2012). The importance of judicial efficiency is on the rise, with major institutions worldwide issuing reports and recommendations on the status of several countriesâ judicial systems (Dakolias, 1999; WorldBank, 2018; EC, 2022). This awareness is driven by a variety of factors, including the need to ensure judicial independence, increase public confidence, and boost countriesâ economic interests (Feld and Voigt, 2003; Voigt et al., 2015). Given their pivotal function in society, the efficiency of judicial systems is also a concern on the rise. As a result, an ever-increasing amount of research and studies are being developed with information regarding techniques/tools to measure courts optimal organization (scale), judicial systems' performance, and staff (judges and others) deficits/surplus in terms of number and incentives.

Courts' efficiency assessment is generally made through KPIs that follow international guidelines. The European Commission for the Efficiency of Justice (CEPEJ) has put forth a number of KPIs based on which courts should be analysed (e.g. clearance rates, case turnover, backlog, case per judge, etc.) including quality indicators (e.g. distinction between processes finishing by merit decision and complete appreciation of the case, from the remaining cases). In addition, in order to foster the quality and efficiency of judicial practices within Europe, the Council of Europe and the European Commission organise since 2005 a competition 'The Crystal Scales of Justice Prize' to reward innovative and efficient practices in European courts. This means that the efficiency and quality of justice is a supranational concern. The efficiency of courts and the judiciary is mainly related to the demand side of justice (influenced by the regulation of lawyers, costs incurred by judicial parties, propensity to litigate, court delay). In Voigt (2016) another source of judicial efficiency is mentioned - that related with the supply side of justice (e.g. quality of the law, judicial organization, judges' individual incentives, etc.).

In this paper we address judicial efficiency of Portuguese first instance courts by taking this demand side perspective. The main contributions we bring to the literature is on adding to previous work on Portuguese courts by Santos and Amado (2014) and Silva (2018) taking a broader perspective than these two studies. While Santos and Amado (2014) focused on first instance courts or judicial counties (223 in total), Silva (2018) focused on benches (267 in total) of generic competence that operate within courts. In this paper we focus on benches, which are the most disaggregate unit of analysis of courts. Indeed, first instance judicial courts are organized within judicial districts, which in themselves are organized in judicial counties (comarcas), usually referred as 'courts'. Within courts (or judicial counties) judges are usually organized in benches (constituted in many cases by a single judge and some administrative staff). Benches can be of various types: generic benches or specialised (e.g. Criminal benches, civil benches, commerce benches, labour benches, etc.). The inclusion of both types of benches in the same efficiency analysis as we do in this paper, allowed us to investigate economies of scope in the Portuguese judiciary (i.e. the extent to which the joint 'production' is cheaper than the separate production of goods or services). This is an important contribution of this paper

to the literature. Indeed, we draw conclusions regarding the existence of economies of scale and scope in the Portuguese courts, namely if smaller benches experience increasing returns to scale, and whether specialized benches perform better than non-specialized (generic). To the authors' knowledge, only the study of Mattsson and Tidanå (2019) focused on economies of scope applied to courts.

The reform undertaken in Portugal in 2014 (Law 62/2013 regulated by Decree law, 49/2014) which implemented specialized jurisdictions at the national level and resulted in the disappearance of most generic courts and on the centralization of work on specialized courts, was based on an implicit, but not tested, assumption that there were diseconomies of scope. This paper provides confirming evidence that indeed that is so.

Methodologically, we employ the non-parametric technique of Data Envelopment Analysis (DEA). DEA has been the prevalent technique used to measure the efficiency of courts, since the precursor study by Lewin et al. (1982) that analysed the efficiency of 100 criminal courts in the US state of North Carolina.

The present paper is organized as follows: Section 2 describes and explores the literature available regarding efficiency analysis, and how it was applied in the judiciary analysis as well as economies of scope, and outputs and inputs used. Section 3 focuses on the DEA and economies of scope techniques and how they are to be applied. Section 4 describes the Portuguese Judicial system structure and the data to be analysed. Section 5 analyses the data and presents the main results, while Section 6 discuss those results, and finishes with a conclusion.

2 Previous Literature

The performance or efficiency of courts is measured many times by single KPIs, one of which is judges' productivity. Studies focusing on judges' productivity (e.g. Choi et al. (2010); Ramseyer (2012); Schneider (2005)) tend to analyse the impact of the educational background of judges on the time needed to reach a decision, as well as their incentives (promotions and monetary), and also reversals from courts of appeal (Schneider, 2005).

Judge's productivity is very much associated with one of the main issues on the judicial efficiency which is pending cases or backlog. Indeed, some research (e.g. Dakolias, 1999) suggests that judges' productivity is boosted when caseload increases, meaning that the more cases (pending and new) the judge has to handle the higher his/her productivity tend to be. However this boost in productivity may have impact on quality. This has been suggested by Kim and Min (2017) who found a negative correlation between caseload and quality.

CEPEJ has in its website a Dynamic database of European judicial systems where two main indicators are used to measure the efficiency of the European judiciary: The clearance rate (the number of cases solved divided by the number of incoming cases), and the disposition time (the pending cases divided by the resolved cases times 365) (Magalhães and Garoupa, 2020). In spite of the widespread use of these two KPIs, they have some problems in measuring the performance of courts. Take for example the fact that"a high clearance rate in a particular year is consistent with both a small backlog and a large backlog, situations very different when it

comes to quality of the judicial system" (Magalhães and Garoupa, 2020, p.1756). In addition, disposition time is obviously a dimension of the efficiency of judicial systems as one envisages that an efficient system is also a quick one. However, it is important to consider more dimensions to efficiency than just the time, as the quality of the decisions is also important and other factors may contribute to higher justice efficiency (e.g. more resources and more investments in IT). Therefore performance measures that can take into account more than one dimension of the performance of courts are important for accurately measuring courts performance. Frontier methods, such as Data Envelopment Analysis (DEA) allow for this.

Since the precursor study of Lewin et al. (1982), addressing efficiency using multiple inputs and outputs through DEA, the literature on the topic has grown significantly. DEA was first introduced by Charnes et al. (1978) and it allows the accounting of several indicators at the same time, contextualizing the outputs produced by the inputs consumed. Regarding courts' efficiency, Voigt (2016) provides a survey of 34 studies that relied on parametric and non-parametric techniques to measure the efficiency of courts. From most of the studies in Voigt (2016), within countries efficiency analysis are performed and only six consider more than two countries. Since the 2016 review there have been some further studies from which we highlight the ones of Peyrache and Zago (2016), Falavigna et al. (2018), Giacalone et al. (2020), Fusco et al. (2021), Falavigna and Ippoliti (2021), and Falavigna and Ippoliti (2022) on Italian courts, Mattsson et al. (2018), Mattsson and Tidanå (2019), Agrell et al. (2020), and Månsson et al. (2022) on Swedish courts, Bełdowski et al. (2020) on Polish courts, Bajrić and Kadrić (2020) on Bosnia-herzgovina courts,Silva (2018) on Portuguese courts, and Gupta and Bolia (2020) on Indian courts. Therefore, within Europe the main studied countries have been Italy and Sweeden.

2.1 Frontier methods used

While traditional DEA models (see Charnes et al. (1978); Banker et al. (1984)) are the most widely used models to measure courts' efficiency, other DEA models have also been applied to the context of courts. For example, Falavigna et al. (2015) and Falavigna and Ippoliti (2021) used Directional Distance Functions (DDF), while (Tulkens, 1993; De Sousa et al., 2005) used Free Disposable Hull (FDH). Besides these techniques some authors performed additional analysis to the computation of DEA efficiency scores through DEA. For example, regression analysis has been undertaken by Beenstock and Haitovsky (2004); Dimitrova-Grajzl et al. (2012); Bhattacharya and Smyth (2001), and Malmquist indexes have been employed by Mattsson et al. (2018), and both the techniques have been used by Falavigna et al. (2018); Giacalone et al. (2020). Malmquist Index is a non-parametric method introduced by Caves et al. (1982) that allows the analysis of efficiency changes over-time, while regression tries to understand the impact of environmental or contextual variables on the efficiency or on the productivity of a court[1]

[1]Deyneli (2011) considered computerization and judges' salaries (e.g.)

2.2 Outputs and Inputs

The use of DEA models implies the division of used variables into inputs and outputs. Inputs are usually understood as the resources used in the production of the outputs. Courts outputs are mainly "cases resolved" or finished. This has been the output that all articles on efficiency analysis agree as the main output of a court. Differences happen in the way cases resolved are aggregated or disagregated. While some authors aggregated the cases resolved as a single output (Hagstedt and Proos, 2008; Elbialy and García-Rubio, 2011), others specified the cases resolved taking into consideration their category (e.g., criminal, civil, labour) (De Sousa et al., 2005; Kittelsen, 1992; Silva, 2018; Gupta and Bolia, 2020), and some went further by subdividing the outputs considering some type of weight that accounted for the complexity/heterogeneity of each case category (Santos and Amado, 2014).Agrell et al. (2020) used three types of cases resolved (Civil, Criminal and petitionary matters) but used Self-reported time consumption for aggregation of cases resolved, in order to handle the different complexities. The issue of case mix and case complexity is however, not a closed issue in judicial efficiency.

To the author's knowledge only Falavigna et al. (2015) and Falavigna and Ippoliti (2021) considered "bad outputs" in their DDF efficiency scores. In Falavigna et al. (2015) the bad output was "Time needed to solve a tax controversy", and in Falavigna and Ippoliti (2021) the bad output was related to additional costs. Other type of outputs are very rare, and an example can be found in Lewin et al. (1982) that used cases pending for less than 90 days (along number of dispositions).

Regarding inputs, there is more divergence in the literature, mainly as far as caseload is considered. Indeed, there is no agreement on whether it should be considered as aggregate caseload or as both new cases and pending cases, or just one of the latter. Several authors, Deyneli (2011); Kittelsen (1992); Tulkens (1993); Finocchiaro Castro and Guccio (2014); Agrell et al. (2020), opted to use solely courts staff as input (judges or/and others). Schneider (2005) pointed the importance of having a measure of workload on the analysis, since it is the measure of demand of a court's services, being directly linked with the output cases resolved (without caseload no cases could be resolved!). Therefore, several studies included pending and/or incoming cases as measure of workload (Schneider, 2005; De Sousa et al., 2005; Peyrache and Zago, 2016; Silva, 2018; Nissi et al., 2019; Giacalone et al., 2020). Even so, the consideration or not of pending cases raises some concerns in the literature, mainly because the demand of courts, as measured through new cases, or its past performance, as measured through pending cases, are not controllable factors. Gupta and Bolia (2020) tested various models in accordance with different input specifications: 1 - just human resources (HR),2 - HR plus new cases, and 3 - HR plus caseload. The authors concluded that including new cases or caseload implied big changes in efficiency and as a result these variables should be included. By comparing the 3 models the authors could reach conclusions regarding whether courts were able to manage annual incoming cases (when they are efficient under model 2 specification) but unable to manage the existing backlog (when they loose the efficiency status in specification 3). Note that the consideration of pending cases as an input will always imply that courts with higher number of pending cases

are more inefficient all else the same - this is the same as penalising a court for past performance irrespective of the current one. That is obviously debatable.

Regarding the use of other sorts of input variables its use is not very prevalent. For example, capital inputs are often not considered, exception made to Elbialy and García-Rubio (2011), that used number of computers as a proxy for capital[2] and Kerstens et al. (2022); Agrell et al. (2020); Mattsson and Tidanå (2019) that considered office space for the same purpose. Falavigna et al. (2015, 2018) used the geographic distribution of courts and the time required to perform certain steps in the workflow.

2.3 Main policy implications from the literature

Most courts' studies addressed questions with political impact. Some of these questions are listed below:

- "***Does size matters?***"
- "***Are reforms effective?***"
- "***How homogeneous is the efficiency within the same country?***"
- "***Are specialized courts better than non-specialized***"
- "***What are the factors that affect courts' efficiency ?***"
- "***How can judiciary efficiency be improved?***"

For the first question on whether the size of a court is an important matter, the analysis of economies of scale is generally the means to reach this conclusion. Some authors, like Kittelsen (1992), Santos and Amado (2014) or De Sousa et al. (2005) found scale inefficiencies mainly associated with small courts. Kerstens et al. (2022) approach is different, but found that in mergers in Swedish courts some evidence of efficiency improvement is found. Besides, they also found that horizontal mergers improve plant capacity utilisation. So these studies corroborate the hypothesis that smaller courts tend to be more efficient than larger courts.

On the contrary, Peyrache and Zago (2016) state that 35% of total inefficiency of the judiciary is due to size and together with reallocation of inputs, splitting large courts into smaller ones leads to higher efficiency scores. Tulkens (1993) says that most inefficient units are middle size courts as measured by clerical staff.

Regarding the effectiveness of reforms, Hagstedt and Proos (2008) concluded that in Sweden, the reform program (reducing the number of existing courts) has improved the relative efficiency of most district courts. Agrell et al. (2020) also concluded that the merger wave within Swedish district courts in 2000 - 2009, resulted in the desired increased efficiency.

The issue of homogeneity of efficiency within country, is related with regional differences that have been mostly investigated in Italy. In this country Falavigna et al. (2015) suggest that Italy North-West is the most efficient geographical macro-area, Peyrache and Zago (2016) state

[2]Deyneli (2011) also used level of computerization but in a Tobit Model (2nd stage)

that North Italy is more efficient than south, and Nissi et al. (2019) show a decline in average level of services as you move from the North of Italy to the South and islands.

The efficiency of specialised versus non-specialised courts is related to the issue of scope economies, which we also address in this paper. Early evidence on this issue by Elbialy and García-Rubio (2011) suggests that criminal district courts are more efficient (mean efficiency 68%) than their corresponding civil district courts (mean efficiency 64%), and that higher shares of criminal case load tend to reduce court inefficiency, while more civil and tutelary caseloads do not influence court inefficiency. On the opposite side, Pereira and Wemans (2017) underline the positive effect on judges' productivity when cases are judged in courts where the vast majority of cases are in the civil area. Mattsson and Tidanå (2019) although not discriminating specifically the specialized vs non-specialized courts, argue that quality may be enhanced with a higher degree of specialization, and concluded that an efficiency improvement between 10,1%-19% may be achieved for 17,2% of the Swedish courts if those courts are merged based on economies of scope. The evidence on economies of scope is therefore still scarce in the judicial literature and that is a gap that the current paper fills.

Regarding the factors that impact courts' efficiency, this is related with second stage analysis where environmental variables are used to explain courts' efficiency scores. In this vein, the work of Melcarne and Ramello (2015) reveals that a greater judicial independence positively affects the aggregate performance of the judiciary. On a similar note, Feld and Voigt (2003) further revisited in Voigt et al. (2015), highlight that *de facto* judicial independence increases the economic growth of a country, underlining the importance of the judicial in the economy, as stated in the beginning of this work. Giacalone et al. (2020) analysed the impact of geographical location and litigation rate on productivity measures of courts in Italy, concluding for the least efficiency of south Italian courts and for the positive impact of litigation rate and negative impact of pending cases on all the components of the productivity change Malmquist index. Falavigna and Ippoliti (2021) conclude that support staff in courts have a positive impact on efficiency, while the population of the district of the court has a negative impact on efficiency. Falavigna and Ippoliti (2022) analysed the impact of case matter on efficiency and concluded that the composition of workload can directly affect the performance of judicial courts. That is, a superior performance may depend not on managerial practices but the existence of simpler cases.

Understanding the factors that impact efficiency, is a means to understand and answer the important question on how judiciary efficiency can be improved. There is a considerable number of possible answers to this question. Reducing backlog may be a way of improving efficiency. Tulkens (1993) state that 70% of the backlog cannot be reduced with current inputs thus personnel increases seem to be justified; García-Rubio and Rosales (2010) found that efficiency improvements of Andalusian Civil First Instance Courts may help reduce the pending cases at most 10% (9.38%), which means that efficiency improvements will not entirely fix the problem of backlogs for Andalusian Civil Jurisdictions. Espasa and Esteller-More (2015) argue that efficiency tends to increase over time, but that high efficiency is not necessarily good if courts' congestion is not controlled (relative efficiencies); Efficiency is also negatively affected the

higher the percentage of work days of temporary workers. Bielen et al. (2018) say that reducing backlogs may negatively impact litigation rates in highly litigious independent judiciaries (hence hiring more judges might not suffice), whereas in regions with lower litigation rates, increasing judges' number (to reduce backlog) may be a good approach.

Another way of improving courts performance may be an increase in support staff. Indeed, Santos and Amado (2014) show that courts with a higher proportion of support staff perform better than courts with a higher proportion of judges. Pereira and Wemans (2017) results provided a positive effect on productivity with the increase of the number of employees per judge and Mattsson et al. (2018) suggest that flexibility, especially of smaller courts, might be achieved with the development of a back-up labour force, allowing adjustments to demand fluctuations (see also Falavigna and Ippoliti (2021)). In addition the reduction of judge turnover can contribute to reduce delays in courts as outlined by Guerra and Tagliapietra (2017) regarding the Italian case.

Another measure for improving efficiency may be the reduction of opportunistic behaviour. This suggestion is brought by Finocchiaro Castro and Guccio (2014), showing that opportunistic behaviour from both claimants and lawyers negatively affects technical efficiency in Italian judicial districts. In the case of higher numbers of lawyers in a court of justice, the length of civil proceedings and case resolution may be negatively affected.

Increasing judges' salaries and quality is considered by Deyneli (2011) (the salaries increase) as significant for the efficiency of courts, but certainly not the sole solution for justice services efficiency. Although having PhD does not guarantee *per se*, that a judge is better than the other, Schneider (2005) concludes that the more PhD holders judges, the better the efficiency of those particular courts.

Legal system simplification is a criteria explored by Di Vita (2010) who state that reduction in legal complexity, (e.g., 10% decrease in existing laws) is expected to drop the average duration of civil proceedings in a potentially greater deal than increasing expenditures in social security, judges' productivity and/or reduction of the pending suits stock.

3 Methodology

3.1 *Data Envelopment Analysis*

First introduced by Charnes et al. (1978), DEA allows the computation of technical efficiency of Decision Making Units (DMUs). DEA allows the construction of a deterministic, non-parametric production frontier, which is used to compare the technical efficiency of the DMUs, based on the radial distance of each of those DMUs to the frontier. Cost or revenue efficiency can also be computed following Farrell (1957), but in the court's setting economic efficiency is rarely found. Indeed only the study of Månsson et al. (2022) analysed cost efficiency of Sweedish courts. The remaining existing studies all focused on technical efficiency and most of them used an output oriented perspective - since changing court's inputs is frequently not possible, as most of the time, those inputs consist on demand of the judicial services from the general public, and the number of personnel.

DEA, as conceived by Charnes et al. (1978) adopts Constant Returns to Scale (CRS), which assumes that an increase in inputs will generate a proportional increase in the outputs. Banker et al. (1984) proposed a model that adopts Variable Returns to Scale (VRS), which allows for constant increasing and decreasing returns. The issue of returns to scale will be investigated in this paper and the model that best fits the data will be adopted. Model (1) shows the VRS model, where m and s is the number of inputs and outputs, respectively. The subscript n relates to the DMUs and y_{ro} and x_{io} are respectively the amount of output r generated by unit o and the amount of input i used by unit o under assessment. λ_j are intensity variables showing the weight being attached to each DMU j to form the efficient benchmark for the DMU o under analysis.

$$\begin{aligned} \max \quad & \beta \\ s.t. \quad & \textstyle\sum_{j=1}^{n} x_{ij}\lambda_j \leq x_{io} \qquad (i,\ldots,m) \\ & \textstyle\sum_{j=1}^{n} y_{rj}\lambda_j \geq \beta y_{ro} \qquad (r,\ldots,s) \\ & \textstyle\sum_{j=1}^{n} \lambda_j = 1 \\ & \lambda_j \geq 0 \qquad (j,\ldots,n) \end{aligned} \tag{1}$$

When the convexity constraint ($\sum_{j=1}^{n} \lambda_j = 1$) is dropped (1) becomes a CRS model. If the convexity constraint is replaced by ($\sum_{j=1}^{n} \lambda_j \leq 1$) then model (1) becomes a Non-Increasing returns to scale (NIRS). The optimal solution from model (1) returns a value β^*, which is the expansion needed in outputs of DMU o that led this unit to the frontier without decreasing inputs. The efficiency score of DMU o is given by $\theta^* = 1/\beta^*$. When efficiency is equal to 1 the DMU is located at the technical frontier, while if θ^* is smaller than 1, the DMU is inefficient when compared with other DMUs.

Model (1) can be solved in relation to a pooled or meta-technology when $j = 1, ..., n$ represent all units observed over time. If units are somehow grouped by specific criteria (e.g. in our empirical application we have different types of benches that are also compared amongst themselves) we may consider a sub-sample of j, i.e. just those units belonging to a specific group. In the first case we obtain a meta-efficiency score (θ^M), while in the second case we obtain a group specific efficiency score (θ^G). Clearly the ratio between the two scores ($\frac{\theta^M}{\theta^G}$) yields a gap between the two frontiers that we named frontier gap.

3.2 *Economies of Scale*

There are various methods in the literature to assess economies of scale. In this paper we followed the method of Färe et al. (1985), which requires three efficiency estimates in relation to three technological RTS specifications: CRS, VRS, and NIRS. From the efficiency measures obtained from each of these models, conclusions can be reached concerning returns to scale: (1) If the CRS, VRS and NIRS models yield exactly the same efficiency measure, then the unit lies, or is projected, on a boundary region exhibiting local CRS; (2) If the CRS and NIRS efficiency measures are both equal and lower than the VRS efficiency measure, then the unit lies, or is projected, on an IRS region of the boundary; (3) If VRS and NIRS efficiency measures are both equal and higher than the CRS efficiency measure, then the unit lies, or is projected, on a DRS

region of the boundary. The Färe et al. (1985) method has the advantage of being unaffected by the existence of multiple optimal solutions. Its main disadvantage seems to be the need to solving three DEA problems (Seiford and Zhu, 1999).

3.3 *Economies of Scope*

Economies of scope were first described by Panzar and Willig (1981), as an intuitive production property based on cost savings resultant from the scope (instead of the scale) of a company. One can say there are economies of scope when producing two or more products in one firm is less costly than producing them separately. If the multi-product is more costly to produce in one firm rather than in separate firms, then there are diseconomies of scope. Panzar and Willig (1981) developed the theoretical concept behind economies of scope, and Morita (2003) was among the first author to operationalise the concept of economies of scope in the non-parametric setting of DEA where no cost information was required. Morita (2003) expressed economies of scope as $C(u_1, 0) + C(0, u_2) > C(u_1, u_2)$ where $C(u_1, u_2)$ represents the cost of producing u_1 and u_2 units of product A and B, respectively. Then the author defends that when no information on input costs exists, one can ascertain the existence of economies of scope through efficiency improvements instead of cost savings. The main idea is to compare the DEA frontiers of joint productions and separate production. If a joint production increases the efficiency over an initial separate production, there are economies of scope. If the value remains the same then there are no economies of scope and if the efficiency decreases we are facing diseconomies of scope.

In spite of its importance, the issue of scope economies has not been much addressed in the DEA and efficiency literature. In particular, regarding courts applications, there is only one article, to the author knowledge (Mattsson and Tidanå, 2019) that address this issue. In Mattsson and Tidanå (2019), the authors analysed the potential efficiency effects of Swedish courts mergers. To do so, they decomposed the results into three estimates namely, learning effect, harmony effect and scale effect. The harmony effect is the one related with scope economies. Authors state that economies of scope, more specifically the harmony effect, are "achieved by either a more productive mix of inputs, or a more easily produced output mix". The authors look for how much more average outputs can be produced with a certain average of inputs. The usage of averages is necessary to nullify the effect of size. Though, as the authors note, that is only relevant when units are closely the same size. When they are too different in size, it might be difficult to distinguish between harmony and size effect. They concluded that efficiency gains from economies of scope ranged between 0%-20%, with 17,2% of the courts being placed in between the efficiency improvement of 10,1%-19%.

In the present application to Portuguese courts, the joint production is represented by the generic benches, which deal with all types of cases, whereas the separate production are the courts that deal with only two or less than the 4 types of cases considered (the structure of Portuguese judicial system is presented in the next section).

In our empirical application neither Morita (2003) nor Mattsson and Tidanå (2019) approaches to economies of scale could be employed. The reason is related to the fact that many

inputs and outputs will have zero values, when we consider the full set of variables for all types of courts (for example criminal courts do not handle labour cases whose value is zero). This would result in a big number of undefined efficiencies, that lead to no conclusions. To tackle this situation, economies of scope will be analysed resorting to the concept of frontier gap. This means that, when this ratio is closer to 1, the within-group frontier is closer to the meta-frontier. This approach will allow an analysis that could potentially reveal the true efficiency of a group by comparing its within-group frontier (not affected by others) with the meta-frontier (composed by every bench and affected by every bench as well), and a differentiation between generic (non-specialized) and the remaining bench types (specialized).

4 The Portuguese System and Data

4.1 Portuguese judicial system structure

Portuguese Judiciary system is administrated by the Ministry of Justice (government department) but unlike other countries, the head of that department does not exercise any authority over the Public Ministry nor heads the public prosecutions[3]. Courts themselves may be divided into 4(5) different larger categories (orders), Constitutional, Judicial, Administrative, and Auditing (the fifth, Military, was extinguished in 2003, but may be re-established in times of war[4]) and 3 minor jurisdictions, Peace Courts, Courts of Arbitration and Ecclesiastical Court. This paper handles only the Judicial courts which are hierarchically divided into, The Supreme Court of Justice (as its decisions are final in terms of Law, it is only possible to appeal to the Constitutional Court), Courts of Appeal (which handle the appeals from Courts of First Instance), and Courts of First Instance. The latter will be the sole focus of this analysis. As stated before, Santos and Amado (2014) uncovered some issues regarding the scale of courts which were considered in the reform of 2014 where some courts (below 250 cases yearly) were extinguished resulting in 23 general jurisdiction constituencies (*comarcas*) that deal with generic or specific competence cases. Those constituencies are yet divided into Municipalities which subdivide into benches (*Juizos*). The latter comprise the division between specialized competence, generic competence, and proximity. Specialized competence benches deal with cases according to their proximity, namely central civil and criminal, local civil and criminal, local minor crimes. The remaining benches are criminal prosecution, labour, tutelary, commerce, and enforcement. In 2016, some of those changes were undone with Decree-Law 86/2016, more precisely the reactivation of the extinguished courts, based on the assumption that some population access to justice was compromised. According to EC (2022), Portugal has been improving in the past few years, nevertheless, concerns about the adequate number of human resources and transparency remain.

[3]Public prosecution is assured by a government-independent body of magistrates headed by the Attorney General's Office.

[4]Their military judges were incorporated in Judicial Courts and together with civil judges, they handle military crimes within "Collective Courts".

4.2 Data

As previously mentioned, the data used for this study consists of all Portuguese first instance benches from 2015-2021. We opted to leave military benches and Criminal instruction out, as the former are only used in some special occasions, and the latter are responsible for conducting the preliminary investigation to decide whether a case should go to trial or not (and do not perform trials *per se*). As a result, these two types of benches are not comparable with the remaining.

Benches analysed are included in one of 10 types: generic, central civil, central civil and criminal, central criminal, civil, commerce, criminal, enforcement, tutelary, minor crimes, and labour. The 10 types of benches were aggregated, for purposes of this study, into a smaller number of groups, based on the type of case they handled: Agg_Civil (which contemplates central civil and civil), Agg_Criminal (which contemplate all criminal types of benches), generic, labour, enforcement, commerce, and tutelary. This results in 7 different groups of benches, that deal with different mixes of the 4 types of cases that our dataset contains (civil, criminal, labour and tutelary). Although enforcement and commerce (e.g.) deal with civil cases, just like the benches in Agg_Civil, it is clear that the procedures are different in each situation, therefore this separation is kept.

After the data were cleaned of some errors such as absolute zeros in judges / staff and cases, 3 365 observations remained (R software was used for both this purpose and data manipulation). The data were provided by the Portuguese Directorate-General for Justice Policy' statistics[5]. For the study at hand no benches were considered as outliers, therefore, there are benches with less than 100 cases incoming or solved whereas bigger benches have over 200,000 cases.

The inputs considered for the study at hand, consist of case load defined as incoming plus pending cases of the 4 different types (civil, criminal, labour, and tutelary) and personnel (Judges and Other Staff). As outputs, cases resolved of civil, criminal, labour and tutelary types were considered.

A descriptive summary of the variables used may be found in Table 1[6].

Table 1: Descriptive summary of variables (Inputs/Outputs) from year 2015 until 2021

Years 2015 - 2021							
		Average per bench	Median	Standard Deviation	Min	Max	Average per year
Inputs	Judges	2.6	2	2.65	1	33	1205
	Other Staff	8.71	6	7.53	1	80	4042
	Civil Case load	2,460.65	509	9,905.94	5	184,937	1,142,092
	Criminal Case load	256.40	50	594.27	0	8,473	119,006
	Labour Case load	166.58	0	631.21	0	6,929	77,315
	Tutelary Case load	184.70	0	644.94	0	7,911	85,728
Outputs	Civil Finished	964.97	278	2,819.96	0	46,383	447,885
	Criminal Finished	175.90	38	383.15	0	4,461	81,644
	Labour Finished	108.33	0	425.21	0	5,016	50,278
	Tutelary Finished	130.49	0	442.69	0	4,462	60,566

[5] https://estatisticas.justica.gov.pt/sites/siej/pt-pt

[6] As this analysis is being done at the most disaggregated level, the zeros found in medians show that the majority of benches do not deal with some type of cases

On average benches are relatively small having 2.6 judges and 8.71 other staff. The largest case load is found on civil cases.

In Tables 2 and 3 we consider data disaggregated per year, and show the totals observed in each year. On average the total case load over the period of analysis is 1,424,141 cases, which are mainly constituted by civil cases. There is a decreasing trend of case load over time being the max value found in 2015 (1,915,924) and minimum value found in 2021 (1,006,046). This decrease also happens for the total number of judges and other staff, where the averages are 1,205 and 4,042, respectively. Regarding cases finished (outputs) in Table 3 the numbers differ. The average of cases finished is 640,375, but although year 2015 is still the year with most cases finished (745,777) the year with the least number is year 2020 (466,901), probably as a result of the pandemic.

Table 2: Year-to-year descriptive summary of inputs

Year	Cload Civil	Cload Criminal	Cload Labour	Cload Tutelary	Total Cload cases	Judges	OStaff
2015	1,575,684	152,083	77,623	110,534	1,915,924	1,271	4,164
2016	1,358,296	144,558	81,792	98,025	1,682,671	1,230	4,066
2017	1,201,026	123,990	85,436	87,576	1,498,028	1,226	4,111
2018	1,065,281	119,455	77,528	79,038	1,341,302	1,188	4,117
2019	1,113,232	116,931	77,640	81,986	1,389,795	1,170	4,042
2020	892,895	102,852	67,791	71,691	1,135,229	1,176	3,941
2021	788,230	73,175	73,395	71,246	1,006,046	1,177	3,854
Average	1,142,092	119,006	77,315	85,728	1,424,141	1,205	4,042

Table 3: Year-to-year descriptive summary of outputs

Years	Fin Civil	Fin Criminal	Fin Labour	Fin Tutelary	Total Fin Cases
2015	517,406	98,487	52,304	77,580	745,777
2016	517,036	96,143	53,112	72,149	738,440
2017	472,411	89,538	57,137	65,993	685,079
2018	435,557	79,140	54,981	56,422	626,100
2019	546,530	77,621	52,076	56,973	733,200
2020	327,878	58,523	34,243	46,257	466,901
2021	318,379	72,056	48,099	48,593	487,127
Average	447,885	81,644	50,279	60,567	640,375

The most common KPIs to assess courts' efficiency are Clearance Rates (CR) (Cases finished / cases entered) and Disposition time (DT) ((pending cases / finished cases) * 365). In Table 4 we show these two indicators for the period considered disaggregated by type of case (as available in Portuguese Directorate-General for Justice Policy's statistics website).

Table 4: Descriptive summary of Clearance Rate and Disposition Time from year 2015 to 2021

	Civil		Criminal		Labour		Tutelary	
	CR	DT	CR	DT	CR	DT	CR	DT
2015	127%	824	98.39%	226	118.50%	234	121.52%	220
2016	146.44%	709	113.91%	171	104.75%	214	116.86%	184
2017	141.00%	669	114.99%	153	105.14%	181	112.46%	160
2018	137.69%	622	103.88%	171	105.52%	169	105.89%	167
2019	121.06%	421	103.40%	154	99.79%	179	100.50%	163
2020	124.17%	650	93.88%	234	89.59%	314	100.18%	200
2021	122.56%	597	101.10%	188	104.80%	206	103.31%	178
Average	131.49%	642	104.22%	185	104.01%	214	108.67%	182

Looking at Table 4, it is possible to see that on average, civil cases take longer to solve than all the remaining case types together. If one disregards year 2020 (COVID-19 pandemic), it is possible to see that DT was decreasing over the years for most of the case types, and it is believed that year 2021, although having lower values than 2020, is still a repercussion of what happened in the previous year. As for CR, it reflects the ability of courts to deal with caseload and therefore its values should aim to be higher than 100%. Values over 100% mean that courts are dealing with backlog of cases, which is one of the main goals. Although the DT apparently shows an increasingly efficiency of Portuguese courts (as they are dealing with cases faster), CR reveals lower values in the last years when compared with the beginning of the period of analysis. Therefore, it appears that courts are dealing with cases faster, but they are finishing less cases than the ones entering/pending. Nevertheless, while the values remain above 100%, this decrease should not be seen as critical.

5 Efficiency Results

For obtaining efficiency scores we solved model (1) under various types of returns to scale and for the meta and group specific technologies. We shall report first the results obtained from the RTS analysis and then we will report the results from the scope economies analysis.

5.1 Returns to scale in Portuguese Courts

The investigation of returns to scale used all data pooled together and model (1) was solved under constant, variable and non-increasing returns to scale. We will call the efficiency scores computed in relation to the pooled frontier the 'Meta-efficiency' scores.

The use of pooled data ignores possible changes of the efficient frontier over time. This was considered reasonable in face of our context. For example, the VRS efficiency of courts did not suffer great changes over the period of analysis as seen in Table 5. It is possible to see that year 2020 has the lowest value, but since that year is the year when COVID-19 pandemic reached its peak, it is expected that it affected efficiency (but not technology), just as in 2021 it is expected that an extra effort was made by the professionals to deal with cases pending from the previous year. In Table 5, one can see that (excluding year 2020) the efficiency varies in a range close to 6%, from 68,64% (year 2018) to 74,19% (year 2021)

Table 5: Efficiency over years

Years	Count of benches	Average of VRS Meta-Efficiency
2015	453	72.31%
2016	452	72.62%
2017	459	71.99%
2018	459	68.64%
2019	473	71.02%
2020	476	60.68%
2021	477	74.19%

Regarding returns to scale, Table 6 shows the number of benches that were identified (through the procedure of Färe et al. (1985)) in any type of returns to scale (increasing (IRS), decreasing (DRS) or constant.

Table 6: RTS over years

Years	Count of CRS	Count of IRS	Count of DRS
2015	33	220	200
2016	40	238	174
2017	33	236	190
2018	36	267	156
2019	37	278	158
2020	32	287	157
2021	56	273	148
Total	267	1799	1183

That is, about 55% of the benches analysed experience over the whole period of analysis Increasing returns to scale, while about 36% experience decreasing returns to scale. Just a minor percentage of around 9% experience constant returns to scale. We can notice in Table 6 that there are not many variations over time. Most variations in RTS are possibly by type of bench, and that analysis is shown in Table 7.

Table 7: RTS by Specialization

Specialization	Count of CRS	Count of IRS	Count of DRS
Agg-Civil	38	551	145
Agg-Criminal	57	284	472
Generic	29	543	156
Commerce	7	51	91
Enforcement	52	52	67
Tutelary	46	157	141
Labour	38	161	111

It is interesting to notice from Table 7, that the benches that deal mainly with civil cases (Agg.Civil) face mainly IRS (around 75%), while the ones dealing mainly with criminal cases face DRS (58%). Generic benches also face mostly IRS (75%). This implies that in most cases the size of the benches is not optimal with some benches being too small (the ones experiencing increasing returns) while others are too large (the ones experiencing DRS).

Figure 1 shows the percentage of benches identified under the three types of RTS per size of the bench (proxied by the number of judges). All benches with more than 9 judges experience DRS. The highest percentage of courts exhibiting CRS happens for benches with 2 judges.

The same picture can be produced per type of bench as shown in Figure 2.

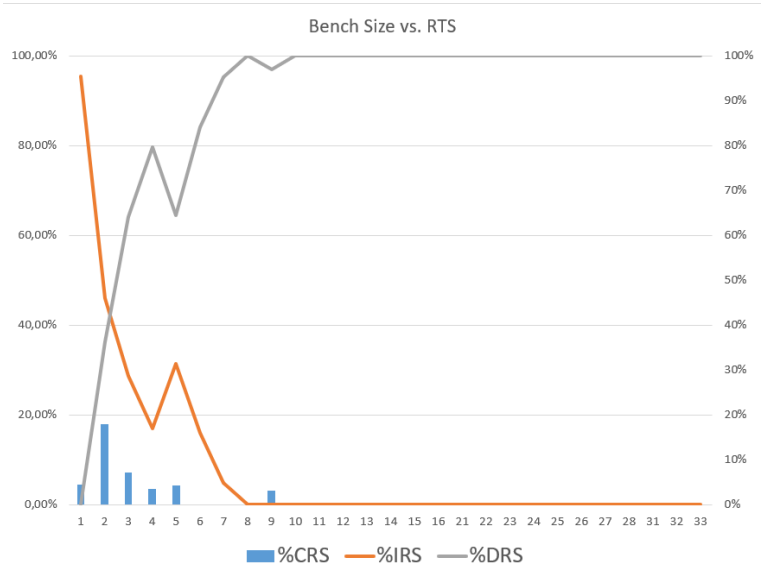

Figure 1: Bench Size vs. RTS

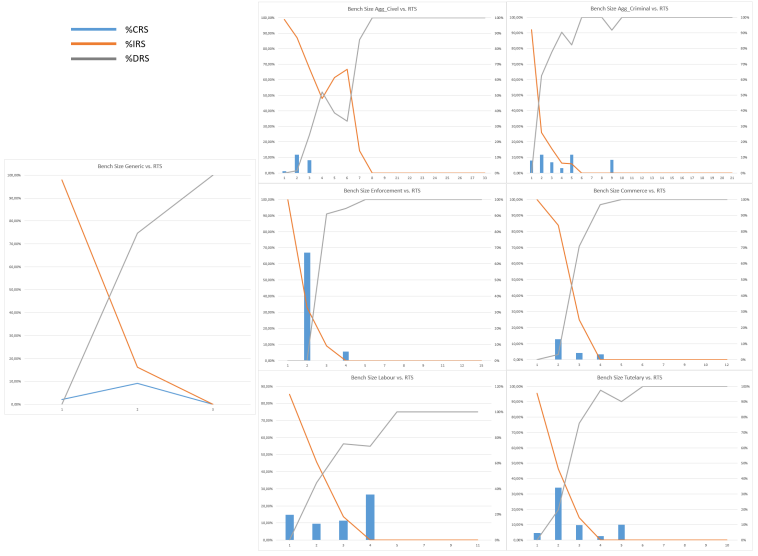

Figure 2: Bench Size for each type vs. RTS

This figure puts in evidence that optimal scale differs depending on the type of bench. For example generic benches seem to operate optimally with 2 judges, since with 3 judges all generic benches experience decreasing returns to scale. Benches dealing only with civil cases should not be larger than 8 judges, since after this value all benches experience DRS, while benches dealing only with criminal cases should be no larger than 6 judges since after that no court experience IRS.

Given that most courts experience decreasing or increasing returns to scale a VRS specification seems the most appropriate. Therefore we report some details of this specification for the meta-efficiency analysis. The national map with the VRS meta-efficiency scores aggregated by district is shown in Figure 3.

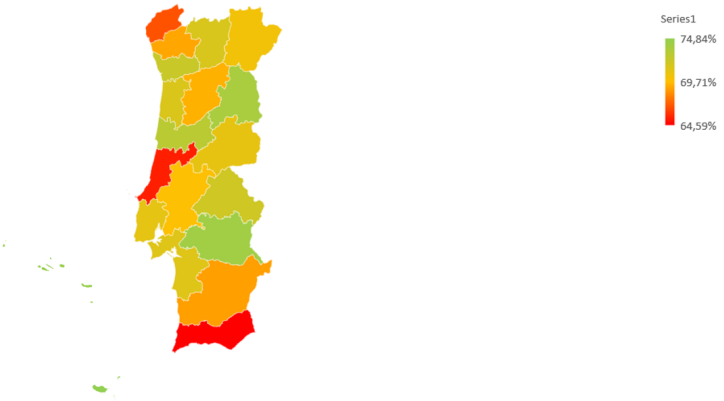

Figure 3: Efficiency Map by district

That is, the worst performing districts in the country are Algarve in the south, Leiria in the centre and Viana do Castelo in the north. So there is not a geographical patterns in the efficiency of the Portuguese benches as best and worst performances are spread all over the country.

In order to understand what is the general profile of the best performing benches we divided them into 3 categories of efficiency, namely P1 if they had efficiency above 66%, P2 if the efficiency was between 33% and 66%, and P3 if their efficiency was below than 33%. The average of inputs and outputs for these 3 groups of benches can be seen in Figure 4 (data normalised using means).

From Figure 4 it is possible to see that most inefficient benches (P3) are the ones that have the highest number of civil case load as well as the ones that resolve more civil cases. P3 courts also have the larger number of judges. It is also possible to see that courts in P2, although not the most efficient, tend to be very balanced in all aspects. The main difference between the 3 groups regards the mix of cases, as indeed P1 benches do receive/finish, on average, more cases than the remaining courts in tutelary, labour and criminal classifications, while receiving the least number of civil cases. This shows the importance of the case mix, and the fact that part of the inefficiencies identified in the meta-analysis may be related to the type of bench and the mix of cases handled rather than with technical inefficiency.

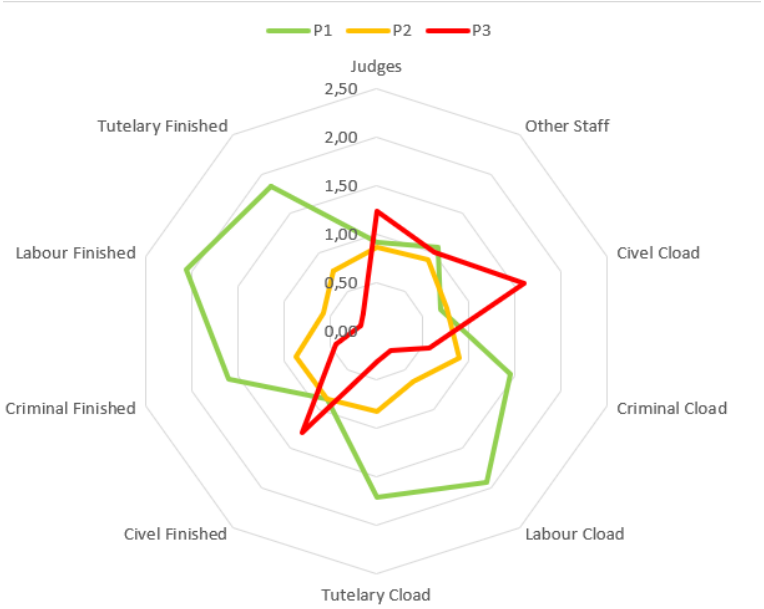

Figure 4: Distribution of inputs/outputs per Category

5.2 Economies of Scope in Portuguese Courts

In the analysis of scope economies we used the VRS model and computed efficiency scores within group and against the meta-frontier. Regarding the meta-frontier results the values obtained for each type of bench can be seen in Table 8. It is possible to see that enforcement benches have the worst efficiency, while labour and tutelary have higher values of efficiency. One of the main reasons for this is the sense of urge in finishing cases regarding labour and tutelary. The former regards with the economic safety of the population as well as the economic health of the country itself. The latter deals with divorces (e.g.), which often involve children, therefore it is of utter importance to deal with these cases as fast as possible. Enforcement very often deal with companies and these processes tend to be stretched out through time by both lawyers and stakeholders to avoid potential fines, increase gains, among other interests. We also show in Table 8 the number of judges and Other Staff (OS) of each type of bench, since we have seen before that size matters and is different between types of benches with generic benches (the non-specialised ones) being in fact much smaller than the rest.

Table 8: Efficiency by Specializations

		N. observations	N. Efficient	Avg Judges	Avg OS	Avg OS per judge	Avg Meta Efficiency
Specialization	Enforcement	171	12	3,06	12,04	3,93	58.44%
	Agg_Civil	734	11	3.12	7.56	2.43	60.99%
	Commerce	149	5	3.53	15.25	4.32	69.65%
	Generic	728	26	1.30	5.14	3.97	70.32%
	Agg_Criminal	813	63	3.29	10.68	3.25	72.66%
	Tutelary	344	30	2.45	11.28	4.60	79.26%
	Labour	310	50	2.07	6.82	3.29	81.70%

The results of the frontier gap (ratio) and meta efficiency and within-group efficiency, may be found in Table 9.

Specialization	Count	Avg Meta efficiency	Avg Within-group efficiency	Avg Ratio(gap)
Enforcement	171	58.44%	61.30%	95.72%
Agg_Civil	734	60.99%	66.73%	92.08%
Commerce	149	69.65%	78.07%	89.39%
Generic	728	70.32%	84.54%	83.23%
Agg_Criminal	813	72.66%	74.89%	97.24%
Tutelary	344	79.26%	81.18%	97.70%
Labour	310	81.70%	87.15%	93.80%

Table 9: Meta Efficiency, Within-group efficiency and Gap per specialization

Looking at the results, one can notice that generic benches, which were above the average in meta efficiency, and had the second highest within-group efficiency, have the worst gap to the meta-frontier, which is 83,23%. This means that the frontier of generic benches are the farthest from the meta-frontier composed by all other benches. Looking closely, it is also possible to notice specialized benches, such as enforcement, labour, tutelary jumping into values above 90%. Agg_Civil, which, *per se*, is a mix of civil and criminal cases, is also one of the worst. Agg_Criminal, although being an aggregation of several benches, all of them deal with criminal, so it can be considered a specialized type of bench as well. Therefore, it is possible to conclude that the more specialized, the closest the efficient frontier is from the overarching frontier of benches, and so, there are reasons to believe that diseconomies of scope apply in the Portuguese judiciary.

The above results may somehow be influenced by scale efficiencies given the fact that generic benches tend to have the smallest number of judges. As a result we repeated the calculation of within group and meta efficiency scores for a subsample of benches - those with a dimension between 2 and 6 judges in order to guarantee comparability in terms of size between the generic benches and the specialized ones (note that all generic benches with a single judge were excluded). The results from this analysis are shown in Table 10.

Specialization	Count	Avg Meta efficiency	Avg Within-group efficiency	Avg Ratio(gap)
Enforcement	73	0.524	0.541	0.969
Agg-Civil	278	0.540	0.640	0.849
Commerce	48	0.655	0.712	0.924
Generic	126	0.671	0.859	0.783
Agg-Criminal	220	0.724	0.752	0.961
Tutelary	114	0.749	0.776	0.966
Labour	104	0.828	0.897	0.922

Table 10: Meta Efficiency, Within-group efficiency and Gap per specialization only for benches with 2 to 6 judges

The analysis of the sub-sample of benches reveal the same conclusions. Generic benches have the lowest average frontier gap implying that this type of bench is indeed the least efficient of all. The specialised benches dealing mostly with civil cases (and as we know also have criminal cases and therefore are not completely specialised) experience the second lowest frontier gap, implying that more specialization indeed generates more efficiency. Notice that parametric and non-parametric tests revealed that differences between groups of benches are statistically significant.

6 Conclusion

This paper analysed the efficiency of Portuguese first instance benches using DEA for the years of 2015-2021, and uncovering whether there were or not economies of scale and scope in the Portuguese judicial structure. Conclusions point to stable efficiency scores over time except in the year 2020, which we believe was a result of the pandemic. Overall a very small percentage of benches experience constant returns to scale (8.2%) and 55% experience IRS implying that they are smaller than optimum size. Generic benches dealing with civil cases are the ones where IRS are more prevalent, while in criminal benches DRS are more prevalent.

Overall our results suggest a huge impact of the case mix on the efficiency of benches, since those benches dealing with more civil cases tended to have lower efficiencies. Considering the question of whether Portuguese judicial system had or not economies of scope, we found diseconomies of scope in the sample analysed. This means that generic benches tend to be further away from the meta frontier than the specialized benches. This result is consistent when we considered the full sample of benches and also a subsample of benches with comparable sizes (between 2 and 6 judges). Our results go in line with some previous results such as Elbialy and García-Rubio (2011) who found that benches dealing with criminal cases are more efficient than the ones tackling civil cases, but when benches deal with all cases together (generic benches) their efficiency lowers (they get further away from the meta frontier). These results also go hand-in-hand with Mattsson and Tidanå (2019) results, that provided similar results for Swedish courts.

7 Acknowledgments

Financial support from the FundaÃ§Ã£o para a CiÃªncia e Tecnologia (through project UIDB/00731/2020) is gratefully acknowledged.

References

Agrell, P. J., Mattsson, P., and Månsson, J. (2020). Impacts on efficiency of merging the swedish district courts. *Annals of Operations Research*, 288(2):653–679.

Bajrić, H. and Kadrić, E. (2020). Efficiency assessment of municipal courts in federation of bosnia and herzegovina using data envelopment analysis. *International Journal for Quality Research*, 14(2):347.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30:1078–1092.

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

Bełdowski, J., Dabroś, Ł., and Wojciechowski, W. (2020). Judges and court performance: a

case study of district commercial courts in poland. *European Journal of Law and Economics*, 50:171–201.

Bhattacharya, M. and Smyth, R. (2001). Aging and productivity among judges: Some empirical evidence from the high court of australia. *Australian Economic Papers*, 40(2):199–212.

Bielen, S., Peeters, L., Marneffe, W., and Vereeck, L. (2018). Backlogs and litigation rates: Testing congestion equilibrium across european judiciaries. *International Review of Law and Economics*, 53:9–22.

Caves, D., Christensen, L. R., and Diewert, W. (1982). Multilateral comparisons of output, input and productivity using superlative index numbers. *Economic Journal*, 92:73–86.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring efficiency of decision making units. *European Journal of Operational Research*, 2:429–444.

Choi, S. J., Gulati, M., and Posner, E. A. (2010). What do federal district judges want? an analysis of publications, citations, and reversals. *University of Chicago Law & Economics, Olin Working Paper*, ""(508):10–06.

Dakolias, M. (1999). *Court performance around the world: a comparative perspective*, volume 23. World Bank Publications.

De Sousa, M., Schwengber, S., et al. (2005). Efficiency estimates for judicial services in brazil: Nonparametric fdh (free disposal hull) and the expected order-m efficiency scores for rio grande do sul courts. *Encontro da anpec, 33th*.

Deyneli, F. (2011). Analysis of relationship between efficiency of justice services and salaries of judges with two stage DEA method. *European Journal of Law and Economics*, 34(3):477–493.

Di Vita, G. (2010). Production of laws and delays in court decisions. *International Review of Law and Economics*, 30:276–281.

Dimitrova-Grajzl, V., Grajzl, P., Sustersic, J., and Zajc, K. (2012). Court output, judicial staffing, and the demand for court services: Evidence from slovenian courts of first instance. *International review of law and economics*, 32(1):19–29.

EC (2022). Justice and consumers - european commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1658828718680uri=CELEX%3A52022DC0500. Accessed: 2022-09-25.

Elbialy, N. and García-Rubio, M. A. (2011). Assessing judicial efficiency of egyptian first instance courts: A dea analysis. Technical report, MAGKS joint discussion paper series in economics.

Espasa, M. and Esteller-More, A. (2015). Analyzing judicial courts' performance: Inefficiency vs. congestion. *Revista de Economía Aplicada*, 23(69):61–82.

Falavigna, G. and Ippoliti, R. (2021). Reform policy to increase the judicial efficiency in italy: The opportunity offered by eu post-covid funds. *Journal of Policy Modeling*, 43(5):923–943.

Falavigna, G. and Ippoliti, R. (2022). Model definitions to identify appropriate benchmarks in judiciary. *Journal of Applied Economics*, 25(1):339–360.

Falavigna, G., Ippoliti, R., Manello, A., and Ramello., G. (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.

Färe, R., Grosskopf, S., and Lovell, C. A. K. (1985). *The Measurement of Efficiency of Production*. Kluwer-Nijhoff Publishing, Boston.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, Series A, general 120(Part 3):253–281.

Feld, L. P. and Voigt, S. (2003). Economic growth and judicial independence: cross-country evidence using a new set of indicators. *European Journal of Political Economy*, 19(3):497–527.

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(3):369–391.

Fusco, E., Laurenzi, M., and Maggi, B. (2021). Length of trials in the italian judicial system: An efficiency analysis by macro-area. *Justice System Journal*, 42(1):78–105.

García-Rubio, M. A. and Rosales, V. (2010). Justice & economy: Evaluating judicial efficiency in andalusia. *InDret*, 4.

Giacalone, M., Nissi, E., and Cusatelli, C. (2020). Dynamic efficiency evaluation of italian judicial system using dea based malmquist productivity indexes. *Socio-Economic Planning Sciences*, 72:100952.

Guerra, A. and Tagliapietra, C. (2017). Does judge turnover affect judicial performance? evidence from italian court records. *Justice System Journal*, 38(1):52–77.

Gupta, M. and Bolia, N. B. (2020). Efficiency measurement of indian high courts using dea: a policy perspective. *Journal of Policy Modeling*, 42(6):1372–1393.

Hagstedt, K. and Proos, J. (2008). Has the recent restructuring of the swedish district courts improved efficiency?: A dea analysis. Technical report, Nationalekonomiska institutionen.

Kerstens, K., Xiaoqing, C., et al. (2022). Evaluating horizontal mergers in swedish district courts using plant capacity concepts: With a focus on nonconvexity. Technical report, IESEG.

Kim, D. and Min, H. (2017). Appeal rate and caseload: Evidence from civil litigation in korea. *European Journal of Law and Economics*, 44(2):339–360.

Kittelsen, S. A. C.and Forsund, F. R. (1992). Efficiency analysis of Norwegian district courts. *Journal of Productivity Analysis*, 3(3):277 – 306.

Lewin, A. Y., Morey, R. C., and Cook, T. J. (1982). Evaluating the administrative efficiency of courts. *Omega, The International Journal of Management Science*, 10(4):401 – 411.

Magalhães, P. C. and Garoupa, N. (2020). Judicial performance and trust in legal systems: Findings from a decade of surveys in over 20 european countries. *Social Science Quarterly (Wiley-Blackwell)*, 101(5).

Månsson, J., Andersson, C., and Bonander, F. (2022). What lessons can be learned from cost efficiency? the case of swedish district courts. *European Journal of Law and Economics*, 54(3):431–451.

Mattsson, P., Månsson, J., Andersson, C., and Bonander, F. (2018). A bootstrapped malmquist index applied to swedish district courts. *European Journal of Law and Economics*, 46(1):109–139.

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. B. (2015). Judicial independence, judgesâ incentives and efficiency. *Review of Law & Economics*, 11(2):149–169.

Morita, H. (2003). Analysis of economies of scope by data envelopment analysis: comparison of efficient frontiers. *International Transactions in Operational Research*, 10(4):393–402.

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(1):395–407.

Panzar, J. C. and Willig, R. D. (1981). Economies of scope. *The American Economic Review*, 71(2):268–272.

Pereira, M. C. and Wemans, L. (2017). Productivity in civil justice in portugal: A crucial issue in a congested system. *Banco de Portugal Economic Studies*, 3(1):1–29.

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.

Ramello, G. B. and Voigt, S. (2012). The economics of efficiency and the judicial system. *International Review of Law & Economics*, 1(32):1–2.

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

Santos, S. and Amado, C. (2014). On the need for reform of the portuguese judicial system - does data envelopment analysis assessment support it? *Omega, The International Journal of Management Science*, 47:1 – 16.

Schneider, M. (2005). Judicial career incentives and court performance: An empirical study of the german labour courts of appeal. *European Journal of Law and Economics*, 20(2):127–144.

Seiford, L. M. and Zhu, J. (1999). An investigation of returns to scale in data envelopment analysis. *Omega, The International Journal of Management Science*, 27:1–11.

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

Tulkens, H. (1993). On FDH efficiency analysis: Some methodological issues and applications to retail banking, courts and urban transit. *Journal of Productivity Analysis*, 4:183–210.

Voigt, S. (2016). Determinants of judicial efficiency: a survey. *European Journal of Law and Economics*, 42:183–208.

Voigt, S., Gutmann, J., and Feld, L. P. (2015). Economic growth and judicial independence, a dozen years on: Cross-country evidence using an updated set of indicators. *European Journal of Political Economy*, 38:197–211.

WorldBank (2018). Doing business - reforming to create jobs. *World Bank Group, Washington.*