What lessons can be learned from cost efficiency? The case of Swedish district courts

Jonas Månsson1 · Christian Andersson1,2 · Fredrik Bonander2

Accepted: 5 April 2022 / Published online: 3 May 2022
© The Author(s) 2022

Abstract
Many studies have reported findings regarding technical efficiency (i.e., resource use) for district courts around the world. However, our review of previous research did not find any studies targeting cost efficiency. To fill this gap, this study investigates the cost efficiency of Swedish district courts. We apply the Farrell framework and decompose cost efficiency into allocative and technical efficiency. The study finds substantial cost inefficiency for district courts in Sweden, which is mainly due to allocative inefficiency (i.e., having the wrong input mix or paying too much for inputs). This result is not surprising since, for example, there is regional heterogeneity in input prices (i.e., wages and rents). However, it also reveals the potential of, for example, moving district courts out of city centres and high-priced areas. On the other hand, the location and staffing of district courts serve other competing policy concerns, such as the proximity of citizens to district courts. The cost-efficiency analysis gives information about the cost of these competing policy concerns.

Keywords Cost efficiency · Decomposition · Allocative efficiency · Technical efficiency · District courts

JEL Classification D24 · H40 · L80

1 Introduction

Efficiency and productivity in the public sector are major concerns for most parliaments and governments. In Sweden, for example, it is explicitly stated in the Budget Act (SFS 2011:203, 2011) that all state services should be provided with a high level of efficiency. However, what is meant by efficiency is far from unambiguous since
the efficiency concept can relate to different aspects of production. Studies of technical efficiency for district courts have been carried out in several countries; however, to our knowledge, no studies have investigated cost efficiency. This study targets the cost efficiency of Swedish district courts.\(^1\)

When studying technical efficiency, efficiency is expressed in real terms (i.e., minimising input use at a constant level of production or maximising production at a given level of resources). Cost efficiency expands this concept and requires more conditions to be fulfilled. These conditions are using the minimum amount of inputs (technical efficiency), paying the correct (lowest) prices, and having an optimal mix of inputs (allocative efficiency). Even though a number of studies have applied the concept of cost efficiency to other entities and investigated district courts in terms of technical efficiency, the novelty of using cost efficiency in this study is that it provides new insights compared with previous research. Further, technical efficiency is included as one component.

There are regulations and policy concerns surrounding the Swedish courts that make an analysis of cost efficiency interesting. First, district courts are often located close to the city centre. According to the Swedish National Court Administration (SNCA), this decision is mainly based on public access to district court services. However, central locations also come with higher prices for office space. The second feature is that district courts are geographically distributed to ensure fair proximity to district courts for all citizens. This means that some district courts are located in areas where the demand for their services is lower than would be justifiable from a production point of view. The third feature is that the minimum staff requirement is regulated, so there will always be a minimum number of judges, law clerks, and support staff in a district court. All these policy concerns are summarised in the cost-efficiency measure. Studying cost efficiency thus makes it possible to estimate a monetary value of these other policy concerns and regulations.

The outline of the paper is as follows. Section 2 briefly presents the legal system in Sweden and the role of the district courts. Section 3 provides an overview of previous research, and Sect. 4 presents the theoretical framework. Our point of departure is Farrell (1957), who showed that cost efficiency can be decomposed into two components: allocative and technical efficiency. Section 5 presents the data, data limitations, and empirical considerations. In Sect. 6, our results are presented, and, finally, conclusions and concluding remarks are provided in Sect. 7.

2 The Swedish legal justice system: a brief description

The Ministry of Justice is responsible for the judicial system in Sweden, including, for example, legislation on the fields of civil law and criminal law. The Swedish court system consists of about 80 different authorities and committees, of which 48 are district courts, which are an important part of the justice system. For the

---

\(^1\) The concept of cost, or overall, efficiency was introduced by Farrell (1957) and extended by Eichhorn (1978), Färe et al. (1985), Kopp and Diewert (1982), and Zieschang (1983).
Swedish legal justice system to provide fair trials, it is necessary for the courts to be independent and autonomous in relation to the parliament, the central government, and other authorities. Furthermore, the Swedish courts aim to deal with cases and matters in an efficient manner.

Three different types of courts build up the Swedish court system: general courts, administrative courts, and special tribunals. The general courts consist of district courts, the Courts of Appeal, and the Supreme Court. Each of these provide possibilities to appeal to achieve a fair trial, which is a fundamental right in any legal justice system. The Supreme Court is the last instance and has the main mission of providing district courts with legal practice to enhance the uniformity of actions in legal decisions, securing a fair trial regardless of which district court makes the decision. All the courts report to the SNCA.

The district courts serve as the first instance in this legal system. Each district court mainly handles cases related to its specific geographical area. Having district courts spread around Sweden ensures that district courts are within a reasonable distance of every citizen in Sweden.\(^2\)

The organisation within different courts is very similar. Each court has chief judges, senior judges, and judges, who are considered as permanent judges. The former act as the head of the court. Judges are appointed by the government. The second category is law clerks. This group consists of non-permanent judges, including both recent law graduates following the training programme to become a permanent judge and regularly employed law clerks who are not participating in the judge training. Law clerks normally prepare cases but can also decide in simple cases, acting as non-permanent judges. Finally, lay judges have special competence in a specific area outside law, for example environmental matters, but they do not necessarily have an education in law. Lay judges are selected by the municipality of the district court and are appointed for a 4-year fixed period.\(^3\)

3 Literature review

The first study of district court efficiency was conducted by Lewin et al. (1982). Since then, several studies have been published on different aspects of efficiency for courts, and these were surveyed by Voigt (2016). Even since that survey, numerous studies have been undertaken. Focusing on those studies that were not covered in the survey by Voigt (2016), we identified the following studies, sorted by studied country:

---

\(^2\) Five courts have a special assignment concerning land and environment cases. These courts deal, for example, with environmental and water issues, property registration, and building matters in addition to their regular duties.

\(^3\) See, for example, Mattsson (2018) for a more comprehensive presentation of the Swedish legal justice and district court system.
• **Sweden** Agrell et al. (2020), Mattsson and Tidanå (2019), and Mattsson et al. (2018).
• **Poland** Major (2015).
• **Portugal** Silva (2018).
• **Italy** Falavigna et al. (2015, 2018), Finocchiaro Castro and Guccio (2014, 2018), Giacalone et al. (2020), Ippoliti and Tria (2020), Ippoliti et al. (2015a, b), Nissi et al. (2018), and Peyrache and Zago (2016).
• **Bosnia and Herzegovina** Bajrić and Kadrić (2019).
• **Argentina** Ferro et al. (2018, 2020).
• **India** Gupta and Bolia (2020).
• **USA** Ferrandino (2014).
• **Europe** Deyneli (2012).

Overall, the papers that have studied the technical efficiency of courts have typically reported inefficiency measures ranging from 10 to 40% of the average technical inefficiency.

In all studies of efficiency, there are four main choices that the researcher must make and motivate: method, selection of input and output, orientation (input or output based), and scale assumption. From a review of the studies above, it is clear that the predominant method used in previous studies is the traditional DEA framework. All the studies listed above except Falavigna et al. (2015) and Peyrache and Zago (2016) used the DEA framework to compute efficiency. Both these studies adopted a directional distance function approach.

The choices of inputs and outputs are, of course, a critical issue, as pointed out by, for example, Ippoliti and Tria (2020). Even though there is variation in the choices of inputs and outputs in previous research, most studies used variations of court staff as one or several input variables. For example, Ferrandino (2014) only used judges as an input. However, the majority of studies listed above divided staff into different categories, for example judges and other types of court staff.

Pending cases and/or incoming cases, and in some cases a measure of workload, were used by Falavigna et al. (2015, 2018, 2019), Peyrache and Zago (2016) and Ferro et al. (2018) as input variables. This group of variables can in some respect be viewed as the demand for court production, and it can be questioned whether the demand should enter the production model.

In classical production theory, models often involve a choice between labour and capital that allows substitution. However, surprisingly few previous studies of court efficiency have used some measure of capital, probably mainly because of a lack of data. This is potentially a problem, especially if production is studied over time, since the model will not capture technological progress. In the studies by Agrell et al. (2020), Mattsson and Tidanå (2019), and Mattsson et al. (2018), capital was approximated by used office space.

On the output side, the majority of studies have used some measure of settled cases to gauge the production of a court. It has also been quite common to divide settled cases into sub-groups of, for example, criminal cases and civil cases (see, e.g., Mattsson et al., 2018; Falavigna et al., 2019). Falavigna et al.’s (2015) study is the only one to have used delay as an output dimension, considering it as a bad
output in the model. A weakness in many other studies is that they have not considered that aggregating very disparate cases into a few sub-groups (or no sub-groups at all) is highly likely to introduce aggregation errors. This potential aggregation error has already been pointed out by Kittelsen and Førsund (1992). However, there are some exceptions, for example Mattsson et al. (2018), who used hearing times, that is, the time spent in the courtroom, to adjust the output variables based on the complexity of cases and matters. Agrell et al. (2020) and Mattsson and Tidanå (2019) used information from time reports compiled by the SNCA.

The predominant orientation used is an output-based model. However, since our objective is to study cost efficiency, the dual of cost is an input-based model. Further, even though we have argued for an output-based model in previous research, it makes little sense to compute revenue maximisation for courts, and it would be impossible since there are no output prices.

Previous research has used a variety of scale assumptions, mostly without any economic concerns. Studies using both constant and variable returns to scale have often also computed scale efficiency.

Finally, we can conclude that all the mentioned studies above have targeted technical efficiency and, to our knowledge, no studies have analysed cost efficiency. The main reason for the lack of research in this area is probably that it is difficult to find data on input prices.

4 Theoretical framework

The concept of overall efficiency, or cost efficiency, was introduced by Farrell (1957) and extended by Eichhorn (1978). Let \( y \) denote a vector of outputs, \( x \) a vector of inputs, and \( w \) a vector of input prices. The starting point for the analysis is to define the technology. Assuming that courts minimise costs given production, the input requirement set \( T \) can be defined as:

\[
T = \{ (x, y) : x \text{ can produce } y \}. \tag{1}
\]

Cost efficiency can then be defined in relation to the input requirement set. Let cost efficiency, or overall efficiency, be denoted as \( O(y, x, w) \). Cost efficiency is defined as:

\[
O(y, x, w) = \min \{ wx : x \in T \} / w_{\text{obs}}x_{\text{obs}}, \tag{2}
\]

where \( w_{\text{obs}} \) and \( x_{\text{obs}} \) are observed processes and inputs. \( w_{\text{obs}}x_{\text{obs}} \) is thus the observed cost.

Technical efficiency is related to input use and output produced and is denoted as \( TE(x, y) \). Technical efficiency is defined as:

\[
TE(x, y) = \{ (x, y) : \lambda x \in T \forall \lambda \in (0, 1) \}. \tag{3}
\]

Finally, allocative efficiency relates to inputs, outputs, and input prices. Allocative efficiency is denoted as \( AE(y, x, w) \). The relationship between the three efficiency aspects is:
This means that one of the efficiency components can be derived from two of the others. In our analysis, we computed cost efficiency and technical efficiency. Allocative efficiency is then derived as:

\[ O(y, x, w) = TE(x, y) \cdot AE(y, x, w). \] (4)

This means that one of the efficiency components can be derived from two of the others. In our analysis, we computed cost efficiency and technical efficiency. Allocative efficiency is then derived as:

\[ AE(y, x, w) = \frac{O(y, x, w)}{TE(x, y)} \] (5)

Figure 1 illustrates an input-based framework in which \( \tilde{y} \) is the fixed amount of output produced (i.e., the isoquant). The lines \( B^1, B^2, \) and \( B^3 \) correspond to the different budget levels needed to produce the fixed amount of output. In Fig. 1, court A is evaluated. Since A is located away from the isoquant, it will be possible to reduce the inputs to produce the given level of outputs. The amount of inputs that can be reduced is illustrated by moving from A to A*. The first part of the cost efficiency concept is thus the existence of technical inefficiency. However, to produce this amount, the budget needs, at given input prices, to correspond to the budget line \( B^2 \). From Fig. 1, it is clear that there is an even cheaper way to produce the given output, which is illustrated by A***. The budget line \( B^3 \) is below \( B^2 \), which indicates a lower total cost. The amount that can be saved by changing the mix of inputs is measured by \( \|A **, A *\| \). This is referred to as allocative inefficiency. The reason for the label allocative efficiency is that the measure relates to the input prices, the mix of inputs, or both. One example of the former could be that the courts are in areas with high rents. There are also regional variations in wages, and, if a court is...
in a region with high wages, it will become allocatively inefficient. An example of a non-optimal input mix would be to hire overqualified staff by demanding formal legal competence for positions that do not truly require it. To compute the efficiency measures stated above, we make use of the non-parametric data envelopment analysis (DEA) framework originally proposed by Farrell (1957) and further developed by Charnes et al. (1978).

5 Data

To gain an overview of the organisation of the district courts’ activities and the availability of data, interviews were conducted with representatives of three district courts and the SNCA. The purpose was to obtain a better understanding of which variables should be used in an efficiency evaluation of Swedish district courts. The data were then extracted from the SNCA’s register, providing access to good-quality first-hand disaggregated data.4

5.1 Inputs and input prices

Labour is the single biggest cost for district courts in Sweden. During the period 2013–2015, about 70% of the total costs corresponded to labour. The second-largest cost is rent, which amounted to about 13% of the total costs in the same period. Other costs, including stationery, computers, copy machines, and so on, make up the remaining 17%.

As seen from previous research, some type of staff measure is used in all studies. We divide staff into three categories based on their role in the district court. The first group contains judges. Judges are the category of staff who decide the most district court cases. Some judges are employed on a regular basis, while others are employed on a temporary basis. There are also junior judges, who are employed temporarily by the court during their legal training.5 All of the staff who are judges are merged into one variable. The average hourly wage cost, computed as the sum of the wage cost divided by the hours worked, is used as the input price.

The second input variable consists of law clerks, both those who are regularly employed as law clerks by the district courts and those who are in the law clerk training programme. Law clerks and those who are participating in the law clerk training programme are allowed to decide in simpler court matters.6 The average hourly

---

4 The data were extracted with high quality and precision by SNCA controller Claes Tidanå, whom we cannot thank enough.
5 Furthermore, since 2012, the court administration has employed a group of judges to work as a reinforcement crew in district courts with temporary needs. Finally, technical experts are hired by the district courts that decide land and environmental cases.
6 The law clerk training programme is a two-year trainee programme for people who have university law degrees. Law clerks have completed the programme and have the right to decide in simpler cases. Furthermore, they prepare and write proposals for decisions by the judges.
wage cost, computed as the sum of the wage cost divided by the hours worked, is used as the input price.

The third input variable is other staff employed by the district court on both a regular and a temporary basis. An important sub-group is court secretaries, who are usually not legally trained but nevertheless have an important role in preparing drafts of court opinions. Other sub-groups within this input variable are IT staff, janitors, and cleaners. The average hourly wage cost, computed as the sum of the wage cost divided by the hours worked, is used as the input price.

The fourth variable approximates the capital used by district courts. Rent for the court buildings is the largest capital cost for district courts. Office space is used as an approximation for physical capital. The rent per square meter is used as the price of this input. In the short run all prices are given. However, a court can choose between different locations as well as different characteristics of the office space rented. Choices that will implicate different prices per square meter of office space rented. Court facilities can be owned by either private landlords or the state. However, regardless of ownership, all courts pay market rent for their premises.

Finally, we use a variable that collects all other expenses, labelled ‘other costs’. This includes stationery, computers, printers, office equipment, and so on. We assume that all courts face the same prices. The argument is that it would be easy for a court, for example, to look on the Internet and find price information. In addition, if large quantities are bought, it is mandatory in Sweden for such purchases to be procured in competition. Since the DEA method is invariant of scaling, an arbitrary price can be chosen if all courts face the same input prices. Thus, the final model contains the following four input variables and input prices:

| Input | Input price |
|-------|-------------|
| Number of hours worked by regular and irregular judges of any kind | Hourly wage |
| Number of hours worked by law clerks and law clerk trainees | Hourly wage |
| Number of hours worked by other staff | Hourly wage |
| Office space used by the district court | Rent/m² |
| Other costs | Fixed = 1 |

5.2 Outputs

The main task of district courts is to handle and decide on cases and court matters. There are strict eligibility requirements for the professional categories that determine the different categories of cases and matters. Cases are decided by judges. Matters, however, can often be delegated by the chief judge to law clerks or those taking part in the law clerk training programme. Court matters usually require a smaller outlay of resources to decide than cases. Consequently, they are included in a separate output variable in this study.

District court cases can mainly be divided into two categories: criminal and civil. Civil cases are primarily disputes that are amenable to out-of-court settlement and family cases, which, unlike disputes, are not amenable to out-of-court settlement.
Disputes may, for example, involve demands for money, interpretations of contracts, or any other financial obligation. Family cases may consist of divorces or disputes over children’s residence, right of access, and child support maintenance. During the interviews, it was mentioned that negotiations of out-of-court settlements between parties are an important factor for district courts to save resources. If a district court manages to encourage settlements out of court for more cases, fewer resources have to be spent on these cases and thus the court efficiency is expected to be higher. Therefore, civil cases constitute the second output variable.

A crime occurs when someone commits an act that is punishable by law. When the crime is brought to court, it is called a criminal case. The number of criminal cases is used as the third output dimension.

There are also other types of cases that are neither criminal nor civil. These include, for example, property and environmental cases and are determined by district courts that have an extended mission to handle these cases. In Sweden, there are five Land and Environment Courts. To manage the extra workload, these courts are compensated with extra resources. These cases are aggregated in the fourth output variable.

The four output variables included in the final specification are:

- the number of determined court matters per year
- the number of determined civil cases per year
- the number of determined criminal cases per year
- the number of other cases per year.

In the model used, output is not registered as output before a final district court decision has been made. This means that, if a case/matter starts in year \( t \) and is decided in year \( t+1 \), the output is registered in year \( t+1 \).

Major differences exist in the various types of cases. Criminal cases can, for example, require more or fewer resources depending on their complexity. This may be reflected in how many people are being prosecuted, how many crimes are being prosecuted, and how complicated the investigation is.

As mentioned in Sect. 3, much of the research on court efficiency has used a straightforward aggregation of all court cases and matters into one or a few output variables without any weighting. This assumes that all cases require the same level of resources to be decided, at least on average, an assumption that is highly questionable. Court cases and matters in the registers of the SNCA are divided into 292 subgroups. To collect information on the four output dimensions presented above, we use the same approach as Mattsson et al. (2018). This means that, from the reported time for each case and matter, we compute weights related to a specific type of case or matter and then aggregate the data into the above four dimensions using these

---

7 These are located in Nacka, Vänersborg, Växjö, Umeå, and Östersund.
8 See, for example Kittelsen and Førsund (1992), Pedraja-Chaparro and Salinas-Jiménez (1996), Azevedo and Yeung (2011), and Santos and Amado (2014).
In Table 1, descriptive statistics of the inputs, outputs, and input prices are presented.

Table 1 shows only small differences in the output dimensions between the studied years. However, there is an increase in the number of hours worked between 2012 and 2014, which may reflect either that more staff have been hired or that the existing staff are working more hours. The hours worked dropped in 2015. Since we are using nominal prices, the increase in hourly wages corresponds to a yearly adjustment of the wages. The rent per square meter varies, and the lowest price per square meter is found for the last year, 2015. This reflects that either some courts having moved to areas with lower rent, were able to negotiate the rent or have downsized.

In Sweden, cases and matters are divided into sub-categories that are aggregated from a certain category. For example, both some traffic offences and murders are sub-categories of criminal cases. For every case and matter that is handled by the court, the time in the courtroom, the hearing time, is recorded. To take this heterogeneity into consideration, the average hearing time spent within each sub-category is computed. The relative time consumption in each sub-category of cases and matters is then used to aggregate, for example, the category of criminal cases. To illustrate how the weights are constructed: (1) the mean hearing time of economic crimes was 215 min during the time period 2007–2015; and (2) the mean hearing time for drug offences was 57 min during the same time period. Thus, in workload one case in economic crime corresponds to 3.8 drug cases. If a district court has one economic crime and one drug offence, the reported output in the group of criminal cases would be 3.8 and 1, respectively, equal to a total output of 4.8 criminal cases.
In this section, the results for long-run cost efficiency (constant returns to scale) are presented first, followed by the results for allocative and technical efficiency. The results for the studied period 2013–2015 are reported, but extra attention is paid to the results for 2015.

### 6.1 Cost efficiency

Table 2 presents the average cost efficiency in the period studied.

| Year | Mean (%) | SD  | Min. (%) | No. of cost-efficient courts |
|------|----------|-----|----------|-----------------------------|
| 2012 | 86.7     | 10.2| 57.9     | 6                           |
| 2013 | 88.2     | 8.8 | 67.5     | 7                           |
| 2014 | 91.6     | 7.9 | 69.4     | 11                          |
| 2015 | 88.1     | 8.6 | 64.5     | 8                           |
| Average 2012–2015 | 88.6 | 9.0 | 57.9     | 32                          |

A district court that is efficient will have a cost-efficiency score of one. If the efficiency score is less than one, it reveals the potential to produce the observed amount of output but at a lower cost. In Table 2, the average cost efficiency for 2012 is 86.7%. This means that it would, on average, be possible to produce the observed output with 86.7% of the observed cost. Put differently, there is 13.3% inefficiency. As shown in Table 2, the cost efficiency varies to some degree over the studied period. One exception is 2014, when the cost efficiency increased by more than 3 percentage points. For the whole period, the average cost efficiency was around 89%. Looking at extreme values, the minimum cost efficiency scores are between around 58% and 69%. This means that these courts, which are the worst-performing courts, should be able to produce the same amount of output using only 58–69% of the budget that they were allocated. Put differently, the maximum inefficiency is between 42 and 31% over the studied period.

Since the method evaluates the efficiency of each district court, it is also possible to point out efficient scores for individual district courts. Figure 2 illustrates the situation in 2015.

---

10 Following Simar and Wilson (2002, 2011) we have tested for returns to scale and could not reject CRS = VRS. However, the main argument for using CRS is that it corresponds to long run efficiency and therefore not allow for inappropriate size of the court.

11 The Benchmarking package in R (see Bogetoft & Otto, 2020) and the software OnFront (Färe & Grosskopf, 1998) are used for computing cost efficiency. To test the scale assumption, the package rDEA is used (Simm & Besstremyannaya, 2020).

12 The results for all the years are presented in Fig. 6 in the Appendix.
In Fig. 2, the amount of cost efficiency is reported, ranked from lowest to highest, on the vertical axis, expressed as a percentage. The horizontal axis presents the courts in the data. Cost inefficiency is viewed as deviation from 100%. For example, the district court located in Sundsvall (left) has cost efficiency of around 65%, which means that it has cost inefficiency of 35%. The average cost efficiency in 2015 was 88.1%. Since the average observed cost per court in 2015 was around SEK 34 million the result indicates a cost saving potential equal to around SEK 4 million per court on average. In 2015, 8 out of 48 courts were cost efficient (i.e., they had a cost-efficiency score equal to 100%). However, since the inefficiency scores are unequally distributed, there is great potential for some district courts, while the potential for other courts is limited. There are three district courts, located on the left-hand side of Fig. 2, which all have the potential to reduce their costs at the given level of production by more than 25%. There could be several explanations for this rather high level of potential. One such factor is the minimum capacity requirement. A district court must have a minimum capacity in terms of staff. For example, the court in Lycksele is a small court and should, in the absence of this regulation, be able to reduce its inputs. However, due to this minimum staff requirement and the relatively few court cases in the area, the court in Lycksele emerges as inefficient.

In the next two sections, we turn our attention to other, internal, sources of cost inefficiency – specifically having the wrong input mix, i.e., choosing an input that is priced above the minimum price available on the market (allocative inefficiency), or overuse of inputs (technical inefficiency).
Allocative inefficiency can be due to either a non-optimal input mix or that a court choose one or several inputs that are priced above the observed minimum market price. For example, by choosing a city centre location or an office space of very high standard the court would most likely have to allocate a larger proportion of the budget to rent compared to a non-city centre location with lower standard. The point is that the choice of different characteristics of for example the office space chosen by a district court will influence the allocative efficiency of that specific court. The average allocative efficiency is presented in Table 3.

The average allocative efficiency for the Swedish district courts is 92.6% for the studied period. On a yearly basis, the variation is small. The highest allocative efficiency of 94.2% is found for 2014 and the lowest of 91.7% for 2015 (i.e., a variation of only 3 percentage points). Courts with the highest potential—that is, those with the lowest allocative efficiency score—are between 71 and 79%. This means that the district courts with the highest saving potential could produce the same amount of output with only 70–79% of their observed costs if they changed the mix of inputs and/or paid lower input prices. It should be noted that, when claiming that a court should pay lower prices, these lower prices are observed in the data, so there are courts in Sweden that experience these lower prices.

The results are far from unexpected. For example, rental costs per square meter are typically higher in larger cities and for courts that are in a city centre. Moving a court out of the city centre would reduce the rent per square meter. In addition, to enable staff to travel to remotely located courts, there are wage compensation issues. On the other hand, a decision to locate a court in these places can be motivated

| Year | Mean (%) | SD | Min. (%) | No. of allocative-efficient units |
|------|----------|----|----------|----------------------------------|
| 2012 | 92.1     | 5.7| 76.4     | 6                                |
| 2013 | 92.5     | 5.6| 71.0     | 7                                |
| 2014 | 94.2     | 5.4| 79.4     | 11                               |
| 2015 | 91.8     | 5.7| 76.2     | 8                                |
| Average 2012–2015 | 92.6 | 5.6 | 71.0 | 32 |

We consider input prices and square meters used by districts courts as given within the existing contract period. However, it is perfectly reasonable that a district court can choose a new location with a new rent when the contract is renegotiated, making both square meters and the price courts pay choice variables in the longer run. We also assume that the same number of square meters will be equally productive, regardless of location. For example, if two district courts are producing the same amount of output with the same amount of staff and office space they will have the same technical efficiency. However, if one district court pay more in rent that court will be cost inefficient. It will not be due to overuse of inputs, rather it is due to the fact that they have chosen a location that is more expensive, i.e., the cost inefficiency is due to allocative inefficiency.
by regional policy concerns. However, even if there is a logical explanation for the amount of allocative inefficiency, the results give a good indication of the cost, for instance, of having a court located in the city centre instead of a more remote location. Figure 3 presents the distribution of allocative efficiency for Swedish district courts in 2015.14

In Fig. 3, the amount of allocative efficiency is reported, in ascending order, on the vertical axis, expressed as a percentage. The horizontal axis presents the courts in the data. Allocative inefficiency is viewed as deviation from 100%. For example, the district court located in Södertälje (left) has allocative efficiency of around 76%, which means that it has allocative inefficiency of 24%.

Figure 3 shows that eight courts were allocatively efficient in 2015. Further, the allocative inefficiency (i.e., deviation from 100%) is unevenly distributed among courts. Courts with allocative efficiency scores lower than 100% have a non-optimal input mix or are paying non-optimal prices. The court with the highest cost inefficiency, Sundsvall, had cost inefficiency of around 35%. However, as the results in Fig. 3 reveal, the amount of cost inefficiency that is due to paying too much for inputs or incorrect allocation only contributes around 15 percentage points. Thus, there are other reasons for the cost inefficiency observed.

14 The results for all the years are presented in Fig. 7 in the Appendix.
In contrast to allocative efficiency, the sources of technical inefficiency can be found in resource use. Table 3 presents the average technical efficiency for the studied period. The average amount of technical efficiency is, in an input-based framework, interpreted as the amount of the current vectors of inputs that are needed to produce the observed output. For the Swedish district courts, the current output could, on average, be produced using 95.6% of the current inputs. This means that there is potential to reduce the current input vector by 4.4% while still producing the same amount of output. The results are very stable over time, and the average efficiency scores differ by only 1 or 2 percentage points over the studied period. The minimum efficiency scores are between 0.627 and 0.775, implying saving potential for these courts of around 23–37%. The rankings of the district courts also remain relatively stable between years. For example, the district court in Sundsvall has the lowest efficiency score in 2 out of 3 years. The individual technical efficiency scores for each district court for 2015 are presented in Fig. 4.\footnote{The results for all the years are presented in Fig. 8 in the Appendix.}

In Fig. 4, the amount of long-run technical inefficiency, that is, 1-cost efficiency, is reported, in ascending order, on the vertical axis, expressed as a percentage. On the
horizontal axis are the courts in the data. For example, the district court located in Sundsvall (right) has long-run technical efficiency of around 75%, which mean that it has long-run technical inefficiency of 25%. Nineteen courts are technically efficient (i.e., cannot improve their performance by reducing the inputs at the given output level and current input mix). However, several courts show technical inefficiency. For example, the courts in Sundsvall and Lycksele both have the potential to reduce their inputs by around 25%. As mentioned previously, there are regulations of court sizes regarding minimum staff levels that are likely to cause at least some of this long-run inefficiency. However, there are also courts that are rather big that still show fairly high levels of technical inefficiency, which is the case for the court in Sundsvall, for instance.

### 6.4 Share of allocative and technical inefficiency

The results above reveal that different cost-inefficient courts have different problems. In Fig. 5, we illustrate the share of cost efficiency that relates to allocative as well as technical efficiency for 2015.

Eight courts were cost efficient in 2015. These courts are located to the left in Fig. 5. To read the figure, consider the court in Hudiksvall, the first court with a visible bar from the left. In 2015, the cost efficiency for Hudiksvall was 88%, the allocative efficiency was 98%, and the technical efficiency was 90%. According to Fig. 5, around 12% of the observed cost inefficiency is due to wrong input mix/input prices, and the other 88% is due to too much input use. The recommendation

---

*Fig. 5* Share of allocative and technical efficiency as a percentage of cost efficiency, 2015, N = 48
for Hudiksvall would therefore be to look more into the use of inputs rather than searching for improvement potential in the input mix or input prices. In contrast, the court located far to the right, Nyköping, has a different distribution. A closer look at Nyköping reveals that the court has cost efficiency of 88%; however, it is also technically efficient. This implies that the court needs to look at its input mix and/or input prices to reduce its cost inefficiency. As Fig. 5 reveals, the major share related to cost inefficiency is not due to excess input use. Instead, it relates to courts having the wrong mix of inputs and/or paying too much for their inputs.

7 Conclusions and concluding remarks

Even though there is a large body of literature on the technical efficiency of district courts, no studies have attempted to measure cost efficiency. Cost efficiency expands the concept of technical efficiency by introducing input prices and input mixes. This means that our study provides policy makers and the administration of district courts in Sweden with new information.

This study investigates cost efficiency for Swedish district courts over the period 2012–2015. Cost efficiency is decomposed into allocative and technical efficiency. We find average cost efficiency of 87%. This means that courts could produce the same amount of output using 87% of what they currently spend. Further, we observe a rather large variation between courts. At the same time, there are courts that are fully cost efficient and courts that have cost inefficiency of more than 40%.

Two reasons for the existence of cost inefficiency can be identified from the model and are labelled allocative and technical inefficiency. To become allocatively efficient, a court must fulfil two conditions. Firstly, it must have the right mix of inputs, and secondly, it must overspend on inputs, given the input mix. Our results show that only eight district courts in Sweden fulfil both conditions. Finally, to be cost efficient, a court must use the minimum requirement of inputs to produce the given volume of outputs. The Swedish courts are mainly performing well in this respect. In 2015, for example, as many as 26 courts were technically efficient. However, there are also some courts for which the results show large amounts of resource overuse. Two courts had technical inefficiency of around 25% in 2015.

To some extent, the results for certain courts presented above are not surprising, and there are logical explanations for them. To be able to operate a district court in Sweden, there is, for example, a minimum staff requirement. For some small courts, this means that, even if there are few court cases to process, it is impossible to reduce the number of staff. This will result in inefficiency. Another cost-driving feature is that courts are often located in city centres, where the rent per square meter is higher. According to the SNCA, the reason for this localisation is that individuals who are in need of, or subject to, court actions should not have to travel too far. In line with this type of argument, there are also courts, often small ones, located in quite remote areas, where the localisation is supported by regional policy concerns. That is, for some inefficient district courts, other political concerns might be driving their cost inefficiency. However, our results show the costs associated with legal restrictions on staff, regional policy, and proximity to citizens. This is information
that the government and the SNCA need to take into consideration in their future work when trying to improve the efficiency of Swedish district courts.

**Appendix: Distribution of costs and allocative and technical efficiency by year: courts in alphabetic order**

Figures 6, 7 and 8.

**Fig. 6** Cost efficiency of Swedish district courts, 2012–2015, N = 48, constant returns to scale (CRS)

**Fig. 7** Allocative efficiency of Swedish district courts, 2012–2015, N = 48, constant returns to scale
Funding  Open access funding provided by Blekinge Institute of Technology. The authors did not receive support from any organization for the submitted work.

Declarations

Conflict of interest  The authors have no relevant financial or non-financial interests to disclose.

Open Access  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

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

Azevedo, P., & Yeung, L. (2011). Measuring efficiency of Brazilian courts with data envelopment analysis (DEA). *IMA Journal of Management Mathematics*, 22(4), 343–356.

Bajrić, H., & Kadić, E. (2019). Efficiency assessment of municipal courts in federation of Bosnia and Herzegovina using data envelopment analysis. *International Journal for Quality Research*, 14(2), 347–368.

Bogetoft, P., & Otto, L.-O. (2020). Benchmark and frontier analysis using DEA and SFA, Version August 7, 2020. https://cran.r-project.org/web/packages/Benchmarking
Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.

Deyneli, F. (2012). Analysis of relationship between efficiency of justice services and salaries of judges with two-stage DEA method. *European Journal of Law and Economics*, 34(3), 477–493.

Eichhorn, W. (1978). *Functional equations in economics*. Addison-Wesley Publishing Company.

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

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

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

Färe, R., & Grosskopf, S. (1998). Reference guide to onfront, economic measurement and quality AB.

Färe, R., Grosskopf, S., & Knox Lovell, C. (1985). *The measurement of efficiency of production*. Kluwer-Nijhoff.

Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), 253–290.

Ferrandino, J. (2014). Testing the Packer theorem: The efficiency of Florida’s criminal circuit courts. *American Journal of Criminal Justice*, 39(2), 375–393.

Ferro, G., Oubiña, V., & Romero, C. (2020). Benchmarking labor courts: An efficiency frontier analysis. *International Journal for Court Administration*, 11(2), 7.

Ferro, G., Romero, C., & Romero-Gómez, E. (2018). Efficient courts? A frontier performance assessment. *Benchmarking: An International Journal*, 25(9), 3443–3458.

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

Finocchiaro Castro, M., & Guccio, C. (2018). Measuring potential efficiency gains from mergers of Italian first instance courts through nonparametric model. *Public Finance Review*, 46(1), 83–116.

Giacalone, M., Nissi, E., & Cusatelli, C. (2020). Dynamic efficiency evaluation of Italian judicial system using DEA-based Malmquist productivity indexes. *Socio-Economic Planning Sciences*. https://doi.org/10.1016/j.seps.2020.100952

Gupta, M., & Bolia, M. (2020). Efficiency measurement of Indian high courts using DEA: A policy perspective. *Journal of Policy Modeling*, 42(6), 1372–1393.

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

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

Ippoliti, R., & Tria, G. (2020). Efficiency of judicial systems: Model definition and output estimation. *Journal of Applied Economics*, 23(1), 385–408.

Kittelsen, S., & Forsund, F. (1992). Efficiency analysis of Norwegian district courts. *Journal of Productivity Analysis*, 3(3), 277–306.

Kopp, R., & Diedert, W. (1982). The decomposition of frontier cost function deviations into measures of technical and allocative efficiency. *Journal of Econometrics*, 19(2–3), 319–331.

Lewin, A., Morey, R., & Cook, T. (1982). Evaluating the administrative efficiency of courts. *Omega-International Journal of Management Science*, 10(4), 401–411.

Major, W. (2015). Data envelopment analysis as an instrument for measuring the efficiency of courts. *Operation Research and Decisions*, 25(4), 19–34.

Mattsson, P. (2018). *Essays on efficiency, productivity, and impact of policy* [Doctoral thesis, Växjö]. Linnaeus University Dissertations, Vol. 39.

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

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

Nissi, E., Giacalone, M., & Cusatelli, C. (2018). The efficiency of the Italian judicial system: A two stage data envelopment analysis approach. *Social Indicators Research*, 146, 395–407.
Pedraja-Chaparro, F., & Salinas-Jiménez, J. (1996). An assessment of the efficiency of Spanish courts using DEA. *Applied Economics*, 28(11), 1391–1403.

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

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

SFS 2011:203. (2011). *The Swedish Budget Act, Chapter 1 3§*. Ministry of Justice.

Silva, M. (2018). Output-specific inputs in DEA: An application to courts of justice in Portugal. *Omega-The International Journal of Management Science*, 79, 43–53.

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

Simar, L., & Wilson, P. (2011). Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis.*, 36, 33–53.

Simm, J., & Besstremyannaya, G. (2020). Robust data envelopment analysis (DEA) for R. Version February 6, 2020. [https://cran.r-project.org/web/packages/rDEA](https://cran.r-project.org/web/packages/rDEA)

Voigt, S. (2016). Determinants of judicial efficiency: A survey. *European Journal of Law and Economics*, 42(2), 183–208.

Zieschang, K. (1983). A note on the decomposition of cost efficiency into technical and allocative components. *Journal of Econometrics*, 23(3), 401–405.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.