A bootstrapped Malmquist index applied to Swedish district courts

Pontus Mattsson¹ · Jonas Månsson^{1,2,3} · Christian Andersson² · Fredrik Bonander²

Published online: 26 April 2018

© The Author(s) 2018, Corrected publication May/2018

Abstract This study measures the total factor productivity (TFP) of the Swedish district courts by applying data envelopment analysis to calculate the Malmquist productivity index (MPI) of 48 Swedish district courts from 2012 to 2015. In contrast to the limited international literature on court productivity, this study uses a fully decomposed MPI. A bootstrapping approach is further applied to compute confidence intervals for each decomposed factor of TFP. The findings show a 1.7% average decline of TFP, annually. However, a substantial variation between years can be observed in the number of statistically significant courts below and above unity. The averages of the components show that the negative impact is mainly driven by negative technical change. Large variations are also observed over time where the small courts have the largest volatility. Two recommendations are: (1) that district courts with negative TFP growth could learn from those with positive

The original version of this article was revised: The equation 7 was incomplete in the published PDF and has been updated.

✉ Jonas Månsson
jonas.mansson@lnu.se

Pontus Mattsson
pontus.mattsson@lnu.se

Christian Andersson
christian.andersson@riksrevisionen.se

Fredrik Bonander
fredrik.bonander@riksrevisionen.se

¹ Department of Economics and Statistics, Linnaeus University, 351 95 Växjö, Sweden

² Swedish National Audit Office (SNAO), Nybrogatan 55, 114 90 Stockholm, Sweden

³ Thammasat Centre for Efficiency and Productivity Analysis, Thammasat University, Bangkok, Thailand

TFP growth; and (2) that the back-up labour force could be developed to enhance flexibility.

Keywords Bootstrap · Data envelopment analysis (DEA) · District courts · Malmquist productivity index (MPI) · Total factor productivity (TFP)

JEL Classification D24 · K49 · O33

1 Introduction

Efficiency and productivity in the public sector are major concerns for most parliaments and governments. It is explicitly stated in the Swedish Budget Act (SFS 2011:203), that all state services should be provided with a high level of efficiency. When dealing with state-provided production, this means that it should be efficient; that is, it should use the least amount of inputs at the given output level or produce the maximum amount of production at the given resource level. The motivation for studying productivity development is therefore, from a policy perspective, straightforward. Swedish applications can, for instance, be found for higher education (Andersson et al. 2017a), employment offices (Andersson et al. 2014), day care (Bjurek et al. 1992), elderly care, and primary and secondary education (Arnek et al. 2016). Economic development is concluded to benefit from an efficient judicial system, internationally (Feld and Voigt 2003; Messick 1999). Furthermore, there is an ongoing debate in the media, as well as in research regarding punishments, criminality, and the judicial system in Sweden (Sturup et al. 2018; Tyrefors Hinnerich et al. 2017). For instance, the share of solved suspicious declined from 2004 to 2014, according to Nordén (2015).

The Swedish Government has launched a number of reforms for the district courts during the last 20 years, with the major objective of increasing efficiency and productivity, while maintaining a high degree of law and order (Swedish Agency for Public Management 2007). One such reform has targeted the size of the courts, based on the assumption that scale advantages exist. In 1999, there were 96 district courts in Sweden, but today only 48 exist. Despite the reforms, the Swedish National Court Administration (SNCA) reports that the productivity has declined. However, problems can occur in the study by SNCA, since productivity (i.e., labour productivity) is measured by partial measures, which ignore substitution between inputs. At the same time, new technologies have been introduced in the courts (i.e., the possibility of conducting hearings through video conferences), and these changes are not captured by the productivity measures used.

The aim of this study is to measure the Swedish district courts' total factor productivity (TFP) from 2012 to 2015. To compute TFP, distances obtained by the data envelopment analysis (DEA) framework, proposed by Charnes et al. (1978), is used. TFP is measured by the Malmquist productivity index (MPI), which was first proposed by Caves et al. (1982) and first applied in a DEA framework by Färe et al.

(1994a).¹ Following Wheelock and Wilson (1999), the MPI is decomposed into four parts: changes in (1) pure technical efficiency; (2) scale efficiency; (3) pure technology; and (4) the scale of the technology. A commonly known issue with all types of DEA analysis is that no statistical inference is possible (Simar and Wilson 2000). In this study a bootstrap approach is used to determine the confidence intervals of the different components presented above (Efron 1979). Another issue with DEA is the influence of outliers (Kapelko and Oude Lansink 2015). The analysis of outliers is, to a large extent, omitted in previous studies on court performance. In this study, an outlier detection analysis is performed to investigate whether the results depend on a few extreme observations.²

The findings indicate a decline in TFP of 1.7% on average. However, a substantial variation between courts and between years is present. Between 36 and 57% of the courts have a significantly negative change in TFP, while the share of courts with a significantly positive TFP change is 16–36%. The reason for this is because courts that improve 1 year may show a decline in TFP the following year. Looking at the components, the negative impact is driven by a decline in pure technical change (TC) of 4.7% in 2012–2013. Further, the TFP is significantly negative during 2014–2015. During this period, the number of courts observed to have a significant decline in TFP is larger, and the numbers of courts with a positive and significant TFP development are fewer, in comparison with the rest of the years. The correlation analysis concludes that the rate of change in the caseload has a significantly positive correlation with TFP, which indicates flexibility problems. The caseload variable is defined as the number of pending cases and matters by the start of the year, plus new cases and matters during the year.

This paper is organised as follows. Section 2 provides a brief summary of the Swedish judicial system. Section 3 presents the previous TFP and efficiency literature regarding courts. Section 4 describes the methodology. Section 5 examines the data, including outlier detection. The results are reported in Sect. 6. Finally, Sect. 7 concludes and discusses the policy implications.

2 The Swedish judicial system: a short description

The Ministry of Justice is responsible for matters that are related to the judicial system, which include legislation on the fields of civil law and criminal law, for example.³ However, it is not allowed to interfere in the day-to-day work, since the aim of the Swedish legal justice system is to provide fair trials. This requires independence and autonomy between courts, in relation to the Parliament, Government, and other authorities. The judicial process differs, depending on

¹ The first application of the MPI using a DEA framework was presented in a working paper (Färe 1988). However, this study was not published in a scientific outlet until 1994 (Färe et al. 1994a), after several publishing attempts (Grosskopf 2003). Therefore, the first published study using the DEA, for computing the MPI, is Färe et al. (1992).

² This study is a development of the report by Andersson et al. (2017b).

³ This section builds on information presented by the Ministry of Justice (2015) and information obtained from the SNCA.

Fig. 1 Description of the judicial process

whether it is a criminal case, a civil case, or a matter. The different processes are described in Fig. 1. Each stage has the general purpose of dealing with cases and matters in an efficient manner and in compliance with the rule of law.

Criminal cases, to the left in Fig. 1, are first handled by the police. These cases start with a police report, followed by a preliminary investigation. In the next step, the case can either be closed or sent to the prosecutor, who will decide whether the case will be prosecuted. If the case continues to prosecution, it will end up in a district court. Initially, a dispute is handled by the municipalities; however, if it remains unsolved, it becomes a civil case in a district court. Civil cases are related to a dispute between individuals or business firms. Matters are regulated in the Court Matters Act (SFS 1996:242) and can be separated into four categories: (1) debt clearances; (2) debt enforcements; (3) bankruptcies or company reconstructions; and (4) other matters. Categories 1–3 relate to payment problems, as shown in Fig. 1. Debt clearances and debt enforcements are decided by the Swedish Enforcement Agency and adjudicated, as a matter, by the district courts if the decision of the Swedish Enforcement Agency is appealed (SFS 1981:774). A decision of bankruptcy must be decided by the district court, which is also the case if a business firm applies for bankruptcy. The fourth category, named ‘other matters’ in Fig. 1, includes a variety of matters, for example: estate administrators, parking remarks, heritages, and custodians.⁴

⁴ The exemplified matters in ‘other matters’ are the largest; however, many other matters (e.g., adoptions, divorces, building permits, etc.) are also included in this category.

There are three different types of courts that build up the Swedish court system, namely, the general courts, the administrative courts, and special tribunals. The general courts consist of the district courts, the Courts of Appeal, and the Supreme Court. Each of the instances is important, since different instances provide possibilities to appeal to achieve a fair trial, which is a fundamental right in any legal justice system. The Supreme Court, which is the last instance, has the main mission to provide the district court system with legal practice to enhance the uniformity of actions in legal decisions.

This study focuses on the district courts, which have the mission to serve as the first instance in the legal system. Each district court mainly handles cases related to their catchment area, which corresponds to the surrounding geographical area. However, there are five courts that specialise in land and environment cases. These courts deal, for example, with environmental and water issues, property registration, and building matters. Within each court, there are Chief Judges, Senior Judges, and Judges who are considered as permanent judges, the former being the head of the court. Each judge is appointed by the government. There are also law clerks who work as non-permanent judges, including both recent law graduates in the training programme to become a permanent judge and regularly employed law clerks that are not included in the judge training. The work tasks of the law clerks normally consist of preparing cases, but can also include deciding simple cases as a non-permanent judge. Finally, Lay Judges have experience from other occupations and politics and are chosen by the Municipal Council, but they are not educated in law. They work as judges for a period of 4 years.

3 Literature review

There is existing literature that focuses on the labour productivity of courts (Blank et al. 2004; SNCA 2015), but there is only a limited amount of literature that considers court TFP. Kittelsen and Førsund (1992) are the first to investigate TFP change over time. The efficiency scores of Farrell (1957) are used to calculate the MPI, which is decomposed into change in efficiency and technology, with the first year as the base (Caves et al. 1982). In terms of the decomposed factors, the catching-up was 4% and the technology shifted 2% from 1983 to 1988. Kittelsen and Førsund (1992) perform an outlier detection analysis, in which the MPI and its components are shown in a histogram, with the labour share on the x-axis. Based on the diagrams, three courts are considered to be outliers, due to a large improvement or decline in TFP.⁵ Fauvrelle and Almeida (2016) calculate the MPI and decompose it into TC and efficiency change (EC).

Following Färe et al. (1994b), EC is further decomposed into a pure EC and a scale component.⁶ The results show, on average, a positive TFP change of 1.5%,

⁵ Kittelsen and Førsund (1992) also note that an outlier with an exceptionally low efficiency score of 0.35 was present. This court is highly specialised, but it is argued that very efficient courts may be more problematic.

⁶ In the USA, a Ph.D. thesis by Ferrandino (2010) investigates TFP change and decomposes it into the same components as Fauvrelle and Almeida (2016).

which is decomposed into a decline of 1.7% in TC, a pure EC of 3.3%, and a scale EC of 0.7%.⁷ Both Fauvrelle and Almeida (2016) and Kittelsen and Førsund (1992) use the averages of TFP change and decompose them into, at most, three components. However, neither of them investigates whether the changes are statistically significant. Finally, Falavigna et al. (2017) contribute to the literature by applying a bootstrapped MPI in a two-stage analysis, proposed by Simar and Wilson (2007), to investigate the impact of structural changes in Italian district courts during 2009–2011. The MPI is found to be 0.3%, EC – 0.1%, and TC 0.4%. Further, they conclude that the role of judges is correlated with court productivity and efficiency.

While only a few studies exist on TFP change, efficiency is measured more extensively. Such research is important for this study, since it deals with the question of which inputs and outputs are best for measuring performance.⁸ Lewin et al. (1982) are the first to investigate inefficiency in district courts, using DEA.⁹ Lewin et al. (1982), as well as all the other studies, use the number of employees as an input. In some studies, employees are measured as the number of judges (Falavigna et al. 2017; Ferrandino 2014; Finocchiaro Castro and Guccio 2014). In other studies, the personnel are separated into judges and office staff (Major 2015; Santos and Amado 2014). The caseload of a court is another input included in some studies. The caseload consists of pending and new cases; that is, the demand of justice services (Kittelsen and Førsund 1992; Schneider 2005). For instance, Nissi and Rapposelli (2010) and Schneider (2005) argue the importance of including the caseload, since an underestimation of productivity will occur, because the employees cannot perform their job without incoming or pending cases. However, this is a slightly contradicting argument when analysing TFP, since courts should be able to adjust inputs when justice demands change. This will be discussed, in more detail, in Sect. 5.

Moreover, Beenstock and Haitovsky (2004) argue that individual productivity increases if the work pressure is high. However, the caseload can also, as Kim and Min (2016) argue, correlate negatively with quality. For example, if the caseload is low, more time can be spent on each case, which, on average, generates a more precise judgement. Outputs normally consist of the number of decided cases (Falavigna et al. 2017; Nissi and Rapposelli 2010). In some studies, cases are separated by type; for example, criminal cases and civil cases (Finocchiaro Castro and Guccio 2016). However, due to data limitations, the studies cannot separate outputs within each category based on the spent resources. This aggregation equalises, for example, a murder with a car crime. Different types of crimes require different amounts of resources, due to their dissimilarity in complication. A problem

⁷ Moreover, an investigation is performed of what is described as ‘... productivity or technical efficiency ...’ (Schneider 2005, p. 133). It is, in fact, TE that is calculated but referred to as court productivity each time, except the previously cited occasion. Furthermore, Schneider (2005) investigates the determinants of efficiency, such as the share of Ph.D. judges and the share of judges older than 60, among others.

⁸ For a general descriptive survey of judicial efficiency, see Voigt (2016).

⁹ Other previous input–output-oriented studies exist, concerning court performance (Nardulli 1978) but not concerning TE. There is also older research on courts that, for instance, focuses on organisational perspectives (Eisenstein and Jacob 1977; Feeley 1973).

with this will occur, in court performance analysis, if there are differences in the mixture of crime types between courts.

Quality variables are argued to be important in some studies (Yeung and Azevedo 2011). Some attempts to investigate the impact of quality variables on performance can be found in the literature. Examples are judges' salaries and education, of which the former have a significantly positive effect on efficiency (Deyneli 2012). Furthermore, Schneider (2005) concludes that more PhD holders, as judges, increase the efficiency. Falavigna et al. (2015) use court delay, as an undesirable output, in a directional distance function.¹⁰ Finally, Andersson et al. (2017b) include a quality measure that relates to the number of changed decisions by a superior court, but does not find any significant correlation with efficiency. All of these studies focus on TE, which is basically a measurement of similarity. For instance, Espasa and Esteller-Moré (2015) argue that the efficiency can be high, even if the courts perform poorly, as long as they are congested. Thus, this is not a good measurement of performance improvements over time; for example, lower inefficiency over time could occur due to a decline in performance of the best district courts.

To sum up, there is no research regarding TFP in Sweden and very little literature, internationally. Furthermore, the international studies do not, due to data limitations, investigate the potential heterogeneity in resource spending within the output categories. Moreover, statistical inference is left out, with the exception of Falavigna et al. (2017), and TFP is at most decomposed into three components.

4 Methodology

Different approaches can be applied in productivity and efficiency studies. Stochastic frontier analysis (SFA) is a widely used parametric methodology (Krüger 2012; Kumbhakar and Lovell 2003). SFA has the advantage of allowing for statistical noise, directly. However, the disadvantage is that it requires a specific functional form. Another option is the DEA approach, which has the advantage of relying on few assumptions and the capability of handling multiple outputs and inputs. Furthermore, DEA is relevant when analysing the public sector, in which the outputs are not sold on the market (Førsund 2016). However, DEA also has some disadvantages. Firstly, it does not give information about inference. To some extent this can be handled by using resampling methods, such as the bootstrap procedure proposed by Simar and Wilson (1998a). The second disadvantage of DEA is its sensitivity to outliers. This shortage is, to a large extent, neglected in previous literature on court performance.

¹⁰ The inefficiency is between 9 and 11% in Norway, Denmark, and Germany (Kittelsen and Førsund 1992; Rigsrevisionen 2000; Schneider 2005) and between 18 and 52% in Italy, Portugal, and Brazil (Nissi and Rapposelli 2010; Pedraja-Chaparro and Salinas-Jimenez 1996; Peyrache and Zago 2016; Yeung and Azevedo 2011).

4.1 Outlier detection

There is no optimal procedure to detect outliers, since no generally accepted definition of an outlier can be found (Davies and Gather 1993). However, plenty of methods are applied in different areas. For example, the outlier detection method, by Wilson (1993), is useful when the data checking is costly (i.e., when the data-set is large). Kapelko and Oude Lansink (2015) use a specific deviation from the median. In DEA, it is important to identify observations that substantially push the frontier, as proposed by Banker and Gifford (1988). This procedure, referred to as the method of super-efficiency, is further concluded to perform well in practical applications, using experiments by Banker and Chang (2006) and Banker et al. (2017), which concludes robustness using different scale assumptions. The focus in this paper is TFP change, and a super-efficient unit 1 year may change the results. This paper identifies an observation as a potential outlier if the output-based super-efficiency score, assuming constant returns to scale (CRS), is below 0.75. This limit is used in, for example, the robustness investigation by Agrell and Niknazar (2014) and the empirical application by Edvardsen et al. (2017).¹¹ Finally, when a potential outlier is identified, a closer look at the specific observation should be taken to produce arguments for why it is an outlier (Simar 2003).

4.2 DEA and the Malmquist productivity index

The point of reference can be taken either from an input perspective (i.e., minimise the inputs to produce a given level of output) or from an output perspective (i.e., maximise the output given the level of inputs). As in most studies of district courts, an output-based perspective is assumed. There are, in the scope of courts, two reasons for choosing an output-based perspective. First, inputs are not easily changed in the short-run. Second, the individual court has no incentives to change its inputs, since the budget for employees is given for a specific year. Thus, the maximum output should be carried out using a given level of inputs. The production technology in time period t , for the 48 Swedish district courts, is defined as:

$$S_t = \{(\mathbf{x}_i^t, \mathbf{y}_i^t) | \mathbf{x}_i \text{ can produce } \mathbf{y}_i \text{ at time } t\}, \quad (1)$$

where S_t represents the technology. Each court, i , uses a vector of inputs, $\mathbf{x}^t$, to produce a vector of outputs, $\mathbf{y}^t$, in period t . Using the output distance function,¹² the technical efficiency (TE) can, in time period t , be written as:

$$D_O^t(\mathbf{x}_i^t, \mathbf{y}_i^t) = \inf\{\theta : (\mathbf{x}_i^t, \mathbf{y}_i^t/\theta) \in S^t\}, \quad (2)$$

¹¹ The super-efficiency scores are calculated using the package ‘benchmarking’ in R (Bogetoft and Otto 2015).

¹² See Shephard (1970) for the output distance function. The input distance function is defined in Shephard (1953).

where θ is a scalar and the distance is $D'_O(\mathbf{x}_i^t, \mathbf{y}_i^t)$. From the distance function, a measure of TE is obtained as $TE = 1/D'_O(\mathbf{x}_i^t, \mathbf{y}_i^t)$.¹³ If TE is equal to unity, the court is on the frontier, meaning that it is technically efficient. However, if TE is larger than unity, the court is inefficient; for instance, TE equal to 1.1 means that the output can be increased by 10%, given the amount of inputs. To calculate the standard MPI, introduced by Caves et al. (1982), the same calculation needs to be performed for the following period: $t + 1$. This is shown in Eq. 3.¹⁴

$$D'_C(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) = \inf\{\theta : (\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}/\theta) \in S^t\} \quad (3)$$

In Eq. 3, and hereafter, the C subscript represents CRS. Similarly, Eq. 3 can be written in the variable returns to scale (VRS) case, which is defined as $D'_V(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})$ where V is the VRS representation. If the technology is not CRS, the MPI does not accurately measure TFP, according to Griffel-Tatjé and Lovell (1995). However, Wheelock and Wilson (1999) state that using the CRS assumption, if the true technology is VRS, will generate inconsistent distances that give arguments for not restricting the calculation to one scale assumption.¹⁵ Using Eqs. 2 and 3, assuming the technology of period t as the reference, Caves et al. (1982) define the MPI as:

$$M^{t,t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t, \mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) = \frac{D'_C(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})}{D'_C(\mathbf{x}_i^t, \mathbf{y}_i^t)}, \quad (4)$$

where the MPI is the ratio of the output distance functions in each period, respectively. This paper uses the most common version of the MPI, based on Caves et al. (1982). To avoid an assumption of the benchmark technology, Eq. 4 is often defined as the geometric mean of two indices.¹⁶

4.3 Decomposition

Decomposition of the productivity index was first proposed by Nishimizu and Page (1982), who define TFP as the sum of the EC and TC. The geometric mean of the two indices is, following Caves et al. (1982) and Färe et al. (1992, 1994a), obtained by rewriting Eq. 4 as:

¹³ Hereafter, the 'O' subscript for the output distance function is omitted to avoid notational clutter.

¹⁴ The name of the index comes from the early work on price indexes by Malmquist (1953). In productivity analysis, it was Caves et al. (1982) who adopted the methodology developed by Malmquist.

¹⁵ See also Grosskopf (2003) for an overview of different arguments on technology assumptions and decompositions.

¹⁶ The geometric mean is commonly used. However, it is worth to point out the potential issue of non-compatibility with the circularity assumption, which was pointed out already by Gini (1931). One possibility is to choose a base technology, but then the index depends on the chosen base (Berg et al. 1992; Pastor and Lovell 2007). Another is to use the 'transitivized' MPI by Balk and Althin (1996). However, the view of Färe (2008), in the spirit of Fisher, is that the natural order of time can be followed (i.e., circularity should not be a problem). For an empirical application of an alternative version of the MPI that fulfills this in a fairly similar application, see e.g., Førsund et al. (2015).

$$M^{t,t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t, \mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) = \left\{ [M^t(\mathbf{x}_i^t, \mathbf{y}_i^t, \mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) \times M^{t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t, \mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})] \right\}^{\frac{1}{2}} \\ = \left[\frac{D_C^t(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})}{D_C^t(\mathbf{x}_i^t, \mathbf{y}_i^t)} \times \frac{D_C^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})}{D_C^{t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t)} \right]^{1/2}, \quad (5)$$

where the distance functions are defined, assuming CRS. Based on the geometric mean defined in Eq. 5, the MPI can be decomposed into TC and EC. Following Wheelock and Wilson (1999), the decomposition is, while allowing for VRS, written as¹⁷:

$$M^{t,t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t, \mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) = \left(\frac{D_V^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})}{D_V^t(\mathbf{x}_i^t, \mathbf{y}_i^t)} \right) \times \left(\frac{D_V^t(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) D_V^t(\mathbf{x}_i^t, \mathbf{y}_i^t)}{D_V^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) D_V^{t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t)} \right)^{1/2} \\ = \text{EC} \times \text{TC}. \quad (6)$$

EC is interpreted as changes in the relative efficiency of a court (i.e., movements towards or away from the frontier), while TC measures the shift of the frontier itself.¹⁸ EC or TC that is larger (or smaller) than unity, indicates an improvement (or decline) in EC or TC, between period t and period $t + 1$.¹⁹ Allowing both TC and EC to have either VRS or CRS, makes the decomposition shown in Eq. 7 possible.

$$M^{t,t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t, \mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) = \left(\frac{D_C^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})}{D_C^t(\mathbf{x}_i^t, \mathbf{y}_i^t)} \right) \times \left(\frac{D_V^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) / D_C^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})}{D_V^t(\mathbf{x}_i^t, \mathbf{y}_i^t) / D_C^t(\mathbf{x}_i^t, \mathbf{y}_i^t)} \right) \\ \times \left(\frac{D_C^t(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) D_C^t(\mathbf{x}_i^t, \mathbf{y}_i^t)}{D_C^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) D_C^{t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t)} \right)^{\frac{1}{2}} \\ \times \left(\frac{D_V^t(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) / D_C^t(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})}{D_V^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1}) / D_C^{t+1}(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})} \times \frac{D_V^t(\mathbf{x}_i^t, \mathbf{y}_i^t) / D_C^t(\mathbf{x}_i^t, \mathbf{y}_i^t)}{D_V^{t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t) / D_C^{t+1}(\mathbf{x}_i^t, \mathbf{y}_i^t)} \right)^{\frac{1}{2}} \\ = \Delta \text{PureEff} \times \Delta \text{ScaleEff} \times \Delta \text{PureTech} \times \Delta \text{ScaleTech}. \quad (7)$$

$\Delta \text{PureTech}$ and $\Delta \text{PureEff}$ are both defined on the best-practice technologies, according to Ray and Desli (1997) and Färe et al. (1994b), respectively. The scale EC measures the movement towards or away from the technically optimal scale.

¹⁷ The inclusion of multiple inputs and outputs, when assuming VRS in TE computations, was first proposed by Banker (1984) and empirically applied by Banker et al. (1984). A graphical description of the DEA frontier, with CRS and VRS, is presented in Fig. 2.

¹⁸ Non-homogeneity occurs, since another scale assumption other than CRS is carried out. This is discussed in Griffel-Tatjé and Lovell (1995).

¹⁹ Färe et al. (1994b) relax the CRS assumption to allow for VRS and decompose EC into a pure effect and a scale effect, respectively. According to Ray and Desli (1997), the decomposition presented by Färe et al. (1994b) is wrong, because the EC is assumed to exhibit VRS while the technology has CRS. To clarify, if CRS holds, there will not be any scale effect, since scale optimality, pioneered by Frisch (1965), is assumed. On the other hand, if VRS is assumed, TC, as defined by Färe et al. (1994b), does not measure the shift in the CRS frontier. However, the error in calculation that Ray and Desli (1997) mention is, according to Simar and Wilson (1998b), only an error in the definition of Eqs. 6 and 7 in Färe et al. (1994b). In other words, the definitions by Färe et al. (1994b) assume CRS, but their calculations allow for VRS. According to Simar and Wilson (1998b), Eq. 6 in Färe et al. (1994b) should be written as Eq. 6 in this paper.

Finally, the scale of the technology (i.e., $\Delta\text{ScaleTech}$), proposed by Wheelock and Wilson (1999), represents the scale bias of TC (i.e., the geometric mean of two scale efficiency ratios). This means that any change in $\Delta\text{ScaleTech}$ occurs from a change in the shape of the technology. The first ratio consists of the change in the scale of the technology between t and $t + 1$. The reasoning of the second ratio is similar, specifically a change in the scale of the technology between t and $t + 1$, relative to the location of the production unit in period t .²⁰ Problems with this decomposition can occur when cross-period distance functions are calculated using the VRS assumption, since it can generate missing values for some components.²¹ Finally, this decomposition is criticised slightly for its confusing interpretation. For example, Wheelock and Wilson (1999) interpret what we call $\Delta\text{ScaleTech}$ as the shape of the technology, while Zofio and Lovell (1998) interpret it as the scale bias of the technology (Balk 2001; Ray 2001).²²

To examine TFP and its decomposed factors, the efficiency needs to be calculated. The reciprocal to the output-based Farrell (1957) measure of TE is formulated by Färe et al. (1994b) as:

$$[D'(\mathbf{x}_i^t, \mathbf{y}_i^t)]^{-1} = \text{TE} = \text{Max } \theta \quad (8)$$

Subject to

$$\sum_{k=1}^K \mathbf{z}_k \mathbf{x}_{k,n} \leq \mathbf{x}_{j,n}, \quad n = 1, \dots, N \quad (9)$$

$$\sum_{k=1}^K \mathbf{z}_k \mathbf{y}_{k,n} \geq \theta_j \mathbf{y}_{j,m}, \quad m = 1, \dots, M \quad (10)$$

$$\sum_{k=1}^K \mathbf{z}_k \geq 0(\text{CRS}), \quad (11)$$

$$\sum_{k=1}^K \mathbf{z}_k = 1(\text{VRS}), \quad (12)$$

where $\mathbf{z}_k$ is $N \times 1$ the vector of intensity variables (i.e., weights). The objective is to maximise θ which corresponds to minimising the value of the distance function,

²⁰ This decomposition is simultaneously proposed by Gilbert and Wilson (1998), Simar and Wilson (1998b), and Zofio and Lovell (1998).

²¹ The VRS assumption is more sensitive to infeasible values when the cross-period distance functions are computed (Briec and Kerstens 2009). This can be avoided by only using a CRS technology. However, CRS is only relevant when long-run equilibrium, in size, appears to be a reasonable assumption (Chambers and Pope 1996). This is not a valid assumption in our case.

²² An alternative decomposition can be found in Lovell (2003). In the Lovell (2003) case, both the output and the input mix are considered, respectively. This is, however, not relevant to district courts, since a single court does not have the power to decide the output mix.

$D'(\mathbf{x}_i^t, \mathbf{y}_i^t)$.²³ For example, if $\mathbf{y}_0$ is an arbitrarily chosen level of output, the maximum output, given the level of inputs, is calculated as $\mathbf{y}_0 * TE^t$ or similarly as $\mathbf{y}_0 / D'(\mathbf{x}_i^t, \mathbf{y}_i^t)$.²⁴

To compute the MPI, four single-period problems are required, assuming CRS and VRS, as well as four mixed-period problems, under CRS and VRS.²⁵ These calculations will generate the average MPI. To improve the robustness of the calculated MPIs and draw conclusions based on statistical inference, a bootstrap approach is applied.

4.4 Bootstrapping the Malmquist productivity index

Statistical inference for DEA is most commonly based on bootstrapping (Efron 1979). Bootstrapping, and other resampling techniques, simulates the data-generating process multiple times by resampling from the data and applying the original estimator to each simulated sample. This generates an approximation of the sample distribution that can be used to create inference that is meaningful in a statistical sense; for example, the confidence intervals of the DEA efficiency scores. These confidence intervals are based on a large number of bootstrap draws (Simar and Wilson 1998a). Further, the efficiency scores can be bias-corrected, as proposed by Simar and Wilson (1999). However, the rule of thumb is not to correct for this bias, unless $s^2 < \frac{1}{3} \left(Bias_B \left[\hat{\theta}(\mathbf{x}_i^t, \mathbf{y}_i^t) \right] \right)^2$, where s^2 is the variance of the bootstrapped values (Simar and Wilson 2000). The procedure can be summarised in four steps: (1) calculate the MPIs as previously described; (2) generate an i.i.d. bootstrap sample from the original sample; (3) calculate the MPIs based on the bootstrap sample; and (4) repeat steps 2 and 3 a sufficient number of times (e.g., 2000 repetitions in our study) to generate standard deviations to construct the confidence intervals of the MPI and its decomposed factors.²⁶

To summarise the methodology, the MPI, including all the decomposed factors and their confidence intervals, will be computed and bootstrapped. This provides the possibility to evaluate changes based on statistical significance due to the bootstrapping. The decomposition can serve as a good starting point for

²³ A distance, $D'(\mathbf{x}_i^t, \mathbf{y}_i^t) \leq 1$, is fulfilled if the technology is defined on the production set, as in Eq. 3. However, it can be above 1 in the mixed-linear problem, indicating technical progress.

²⁴ The returns to scale subscript is omitted to avoid notational clutter.

²⁵ The necessary linear problems to be solved are $D'_C(\mathbf{x}_i^t, \mathbf{y}_i^t)$, $D_{C+1}'(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})$, $D'_C(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})$, and $D_{C+1}'(\mathbf{x}_i^t, \mathbf{y}_i^t)$ for CRS. Similarly they are $D'_V(\mathbf{x}_i^t, \mathbf{y}_i^t)$, $D_{V+1}'(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})$, $D'_V(\mathbf{x}_i^{t+1}, \mathbf{y}_i^{t+1})$, and $D_{V+1}'(\mathbf{x}_i^t, \mathbf{y}_i^t)$ for VRS. These are calculated for each change in time, generating 8*3 problems for each court.

²⁶ This calculation is performed using the FEAR software in R (Wilson 2008), and the bootstrap procedure used in this paper is proposed by Simar and Wilson (1999). A discussion about weaknesses with this bootstrap procedure can be found in Olesen and Petersen (2016).

investigating the sources of the TFP change. This can be achieved without problems of sample noise, making the results statistically robust.²⁷

5 Data

The data used were obtained from the SNCA and cover the time period 2012–2015. However, data on hearing times are available for a longer time period (2007–2015) and will be used for the weighting of the outputs. The available data is more detailed than the data used in previous research. For example, different cases and matters are reported in 292 sub-categories, which will be taken into consideration. The choice of input and output variables is based on what representatives of the courts have stated in interviews as reasonable resource and performance measures, as well as economic theory and previous research.²⁸ This very detailed information implies that the complexity of cases and matters within a given sub-category will not vary much, since cases and matters within the same sub-category will be similar.

In the first step, the outputs are aggregated into decided civil cases, decided criminal cases, and decided matters, which are described in Fig. 1. Simply adding these three groups together, however, will introduce aggregation errors. There are differences between courts, regarding the type of cases and matters that they handle. Some courts handle more resource-intensive cases than others. For example, a murder case would most likely require more resources than a traffic case. This heterogeneity within different categories of cases and matters is not taken into account in previous studies (Kittelsen and Førsund 1992; Santos and Amado 2014; Yeung and Azevedo 2011). To compensate for these facts, the outputs are weighted by the hearing time (i.e., the time in the courtroom).²⁹ This means that courts with a large share of cases, from a complicated category, are not negatively affected in terms of TFP. The weights are based on the average hearing time in each sub-category; for example, criminal cases alone consist of almost 40 sub-categories, and each sub-category receives its own weight.³⁰

On the input side, labour is the largest cost share for Swedish district courts, with 70%, and the rental cost is about 13% for the period 2012–2015 (SNCA 2012, 2013, 2014, 2015, 2016). Labour is divided into three categories, specifically the number of hours worked for: (1) judges; (2) law clerks; and (3) other personnel.

²⁷ Førsund (2016) argues that separation of the pure and the scale effects are problematic in the decompositions. Furthermore, the interpretation of multiplicative decomposition is argued to be unclear. Furthermore, no clear-cut, regarding causality, can be identified. However, other authors such as Lovell (2003) and Simar and Wilson (1998b) argue that the adopted decomposition is meaningful.

²⁸ Wagner and Shimshaks's (2007) strategy of model selection is also used to assess the plausibility of the model.

²⁹ Andersson et al. (2017b) conducted a Confirmatory Factor Analysis (CFA) to investigate if hearing time could be used as an approximation of resources used. The result of the CFA was that hearing time got a loading of 0.96, indicating that the hearing time is a good approximation for resources used.

³⁰ To illustrate how the weights are constructed: (1) Economic crimes had a mean hearing time of 215 min, during the time period 2007–2015; and (2) Drug offences, on the other hand, had a mean hearing time of 57 min during the same time period. This implies that 3.8 drug offences are assumed to use as much resources as one economic crime.

The reason for dividing labour into three categories is that the different types of staff conduct different tasks at the Swedish district courts, and the staff composition varies between courts, according to the SNCA. A measure of capital is omitted from previous research (Ferrandino 2014; Kittelsen and Førsumd 1992), except for Elbialy and García-Rubio (2011), who incorporate computers. To incorporate capital, the office space of the court is used, following the assumption that the amount of capital (e.g., computers and office equipment) is proportional to the size of the premises. We argue that including a measure of capital is important, since it is, to some extent, possible to substitute labour with capital in court production. An example is the incorporation of video conferences, which, according to the SNCA, decreases the travelling time for judges.

Furthermore, the caseload, as described in the previous literature, is an important source of performance for several reasons; for example, if there is no caseload, there will not be any output. We argue that the caseload is an important variable to incorporate when calculating performance that focuses on improvements of technology, management, and so on. However, it is not recommendable to include the caseload in the main analysis when TFP is investigated, since an important factor of TFP changes is flexibility in inputs, i.e., adjustable inputs depending on changes in justice demand. Thus, the caseload is only included in a second-stage correlation analysis. The caseload in year t is defined as the stock of open cases and matters at the end of year $t - 1$ plus the incoming cases and matters in the present year. A potential problem with this method is that the incoming cases, at the end of 2015, are not included. The correlation is invariant for addition and, therefore, only a problem when the difference is non-random between courts. In our case, however, it is more relevant to assume randomness between courts, meaning that it does not affect the correlations.³¹

5.1 Outlier detection

The chosen limit of the super-efficiency scores is 0.75. This means that if an observation is identified with a super-efficiency score below the limit in any of the years it will be under consideration to become eliminated from the main analysis. Four district courts are below this limit during, at least, one of the years. These are Eksjö, Uddevalla, Gotland and Nacka.³² Gotland is unlikely to be super-efficient, in general, based on interviews with representatives for the district courts. However, 2014 is an exception, according to the representatives at the SNCA. Furthermore, from 2013 to 2014, Gotland has a TFP growth of 12%. Therefore, Gotland is eliminated from the main analysis. Thus, four courts are eliminated, meaning that there are 44 courts left in the sample. The descriptive statistics of the outputs and inputs, after the elimination of outliers, are reported in Table 1.

³¹ A measure of quality can potentially be included in the analysis. Andersson et al. (2017b) incorporates the change frequency in higher instances in a second-stage analysis and concludes zero correlation with the efficiency scores.

³² Eksjö had a super-efficiency score of 0.70 in 2012; Uddevalla had a score of 0.73 in 2013; Gotland had a score of 0.66 in 2014; and Nacka had a score of 0.623 in 2015.

Table 1 Descriptive statistics, excluding outliers

	Mean (SD)			
	2012	2013	2014	2015
<i>Outputs</i>				
Number of decided civil cases (weighted)	3722 (3565)	3886 (3733)	3928 (3705)	3762 (3642)
Number of decided criminal, real estate, and environmental cases (weighted)	3975 (3276)	3904 (3283)	3853 (3144)	3901 (3 227)
Number of decided matters (weighted)	845 (612)	869 (735)	889 (799)	875 (764)
<i>Inputs</i>				
Regular and irregular judges of any kind ^a	17.28 (14.36)	17.67 (15.21)	18.53 (16.03)	18.81 (15.92)
Law clerks and employees of the law clerk training program ^a	18.71 (17.98)	19.02 (17.93)	19.34 (18.02)	19.00 (18.18)
Other personnel ^a	32.28 (30.29)	33.43 (31.50)	33.78 (31.58)	34.21 (32.87)
Office space (square meters)	4530 (4318)	4640 (4332)	4622.5 (4300.1)	4649 (4305)
<i>Correlation variable</i>				
Caseload	7079 (6039)	6996 (6108)	6754 (5840)	5158 (4464)

^aMeasured as full-time equivalent

Table 1 show that the differences in output over time are, on average, quite small. However, each of the inputs increased in size over time. For example, the number of full-time equivalent judges increased from 17.28 to 18.81 (9%). The caseload declined over time, but the dramatic drop in the last period is because all incoming cases were not included as previously described. Also note the large standard deviations, which are almost as large as the means. For instance, the Stockholm district court is, in terms of hours worked for judges, almost 31 times larger than Lycksele. Finally, the non-weighted caseload declined during the studied time period.

6 Results

The results, concerning the MPI and its decomposed factors, are reported first, namely, EC and TC. EC and TC are then further decomposed into a pure and scale effect. Then a correlation analysis is performed, based on the MPI and its components. Finally, observations are concluded as outliers are eliminated from the main results.³³

³³ Inclusion of all observations generates similar results. These are made available on request.

Table 2 Malmquist index and its decomposed factors after eliminating outliers

Year	Malmquist productivity index (MPI)	Technical change (TC)	Efficiency change (EC)	Number of courts with TFP below unity*	Number of courts with TFP above unity*
2012–2013	0.980 (0.957–1.001)	0.929** (0.863–0.960)	1.057** (1.009–1.136)	18 (41%)	16 (36%)
2013–2014	1.000 (0.983–1.020)	0.987 (0.941–1.025)	1.014 (0.967–1.069)	16 (36%)	13 (30%)
2014–2015	0.970** (0.947–0.992)	0.992 (0.961–1.050)	0.978 (0.907–1.016)	25 (57%)	7 (16%)
2012–2015	0.983 (0.962–1.004)	0.969 (0.921–1.011)	1.016 (0.960–1.073)	20 (45%)	12 (27%)

The bootstrapped confidence intervals at 5% level of significance are reported in parentheses, where ** symbolizes significance at the 5% level. *Below (above) unity means significantly below (above) the 5% level of significance

6.1 Malmquist index and its decomposed factors

The MPI and its components are reported in Table 2.³⁴

In Table 2 the TFP change, measured as the MPI, is negative for 2012–2013 and 2014–2015, respectively. Column 3 reports that TC is significantly negative during the first period and, on average, below zero in the following periods. EC contributes positively to the TFP growth during the period 2012–2014. For the last period, 2014–2015, both TC and EC affect TFP negatively, which generates a statistically significant decline in TFP. We do not aim to identify causes of the different components of TFP change; however, argumentation of potential sources of the results is provided. A negative TC, which is observed for each period, has its original interpretation from other sectors, where it can occur from an absence of reinvestment in capital so less outputs can be produced. However, this is not likely for district courts. Instead, an inward shift of the frontier will most likely occur, due to two reasons. First, if the turnaround time increases, due to more complicated cases, that will generate a lower output, which occurs as a negative TC in the model. This means that all courts are affected the same (i.e., efficiency remains constant, but all courts are closer to the origin). However, the turnaround time decreased during this time period, in terms of both criminal cases and civil cases, according to the SNCA (2014, 2016), which means that the source is something else.

As a second attempt to interpret a decline in TC or EC, depending on if the affected courts operate on the frontier, it is worth studying Table 1 of the descriptive statistics. In Table 1, it can be observed that the average number of decided cases and matters fluctuate between 1 and 2% for 2012–2014; thus, the changes are stable. However, for 2014–2015, the decided criminal cases and matters are in the same range, as previously described, but the civil cases declined by 4.2%, on average. Thus, the produced output decreases in total, driven by a lower number of civil cases. This, however, only

³⁴ The bias is not corrected for, since $s^2 = 0.00089$ and $\frac{1}{3} \left(Bias_B \left[\hat{\theta}(\mathbf{x}_i', \mathbf{y}_i') \right] \right)^2 = 0.00036$, calculated as an annual average (Simar and Wilson 2000).

concerns decided cases. However, the number of incoming civil cases is reported to decline by 5% during the period 2014–2015.³⁵ Thus, a potential explanation for the negative TFP change that is driven by both a decline in TC and EC, depending on if the courts operate on the frontier, is likely to be due to the decline in the caseload during this period. This, in itself, should not decrease TFP if the inputs are fully flexible. However, it does decrease if the inputs are not flexible enough to compensate for the lower workload level. The MPI and its confidence intervals are graphically reported for each court and each year, excluding the outliers (see Figs. 3, 4, and 5 in the “Appendix”).³⁶ Furthermore, the geometric mean of the MPI and its components are provided for the individual courts in Table 6.

In columns 5 and 6 in Table 2, it can be observed that the number of courts, with a significantly negative TFP change, is fairly stable during the period 2012–2014; however, the number increases in 2014–2015. Furthermore, the number of courts with significantly positive TFP growth decreases in each year. Both the fact that the production frontier moves towards the origin and the fact that fewer courts have a significant and positive TFP change indicates that this result is not driven by a few observations. Additional to the caseload, other differences may generate differences in TFP change. For example, the organisation within the courts may be an issue. Thus, adjusting the organisation to the best performing court can generate a better development. To gain more information, TC is decomposed into pure TC and scale TC.

6.1.1 Decomposition of TC

The decomposition of TC into pure TC and scale TC is performed according to Eq. 7 in Sect. 4. The results are reported in Table 3.

TC is defined as the product of pure TC and scale TC.³⁷ Pure TC means that the best firms, assuming CRS, have a significant decline in 2012–2013 and a smaller negative change in the following periods. The movement of the technology, from the optimal scale (i.e., scale TC), generates regress of 2.2% for the same time period. This indicates that the largest decline has its source in pure TC, meaning that the frontier moves inwards; however, the shape of scale TC also contribute negatively. In November 2011, there was a reform so that a type of matter, handled only by a few courts, was moved to the Swedish mapping, cadastral, and land registration authority. In particular, these courts have a large decline in pure TC; for example, Ångermanland had a negative pure TC of 34%. The source of this decline is that cases were moved in the end of 2011, which generated a smaller stock and less incoming cases during 2012–2013. Therefore, less outputs are produced; meanwhile, the inputs are not changed accordingly, even though the mentioned change was known by the courts at

³⁵ It can also be noted that the incoming cases in this category decrease by 8% from 2013 to 2015 (SNCA 2016).

³⁶ Figures of TFP change of all courts, including the outliers, can be made available upon request.

³⁷ As noted in the methodology, a VRS assumption, using cross-period distance functions, is sensitive to infeasible values. For 2012–2013, Lycksele received missing values for pure TC and scale TC. During 2013–2014, Lycksele and Haparanda obtained missing values. Finally, missing values for these components are received for Lycksele, Haparanda, and Gällivare. Each of these courts shows the same pattern, specifically large volatility in TFP.

Table 3 Decomposition of technical change into pure technical change and scale technical change

Year	Technical change (TC)	Pure TC	Scale TC
2012–2013	0.929** (0.863–0.960)	0.953** (0.871–0.989)	0.978 (0.944–1.023)
2013–2014	0.987 (0.941–1.025)	0.994 (0.937–1.034)	0.993 (0.967–1.025)
2014–2015	0.992 (0.961–1.050)	0.993 (0.963–1.057)	1.002 (0.967–1.026)
2012–2015	0.969 (0.921–1.011)	0.980 (0.923–1.026)	0.991 (0.959–1.025)

The bootstrapped confidence intervals at 5% level of significance are reported in parentheses, where ** symbolizes significance at the 5% level

least 1 year in advance, indicating a flexibility problem. In contrast to the previous interpretation over time (i.e., that the result is not driven by a few courts), it can be concluded that the significant decline in pure TC during 2012–2013 is driven by district courts where different types of matters were moved to another authority. To investigate the components of EC, its decomposition is now reported.

6.1.2 Decomposition of EC

EC is decomposed into pure EC and scale EC, according to Färe et al. (1994b). The results are reported in Table 4.

EC is positive for 2012–2013 and 2013–2014, respectively. The positive effect has its source in the positive pure EC and scale EC for both periods. This indicates that district courts, on average, become more homogeneous, since their efficiency measures the distance from the frontier. However, based on previous arguments regarding TC, it is not necessarily the case that a positive EC of 5.7% has its source in better performance of inefficient courts. Instead, using the decomposition of MPI, inefficiency is reduced when courts on the frontier move towards the origin, meaning that such courts are closer in distance to the previously inefficient courts. In other words, the positive and significant EC during 2012–2013 is, most likely, due to a movement inwards of the frontier.

During the period 2014–2015, EC is negative, indicating greater heterogeneity between courts; that is, the average court is further away from the production frontier. This effect comes almost equally from both components of EC. Thus, it can be concluded that most changes in TFP, during the last time-period, occur from the different components of EC. As previously described, a decline in EC can also occur from a lower justice demand for courts that do not operate on the frontier, ex-ante. However, it can also occur from organisational issues; for example, if high-skilled employees leave the court and there are difficulties finding replacement staff.

6.2 Correlation analysis

A few of the previous studies argue the importance of incorporating the justice demand to avoid underestimating a court's TFP. However, as stated, the justice demand should

Table 4 Decomposition of efficiency change into pure efficiency change and scale efficiency change

Year	Efficiency change (EC)	Pure EC	Scale EC
2012–2013	1.057** (1.009–1.136)	1.023 (0.971–1.117)	1.033 (0.970–1.073)
2013–2014	1.014 (0.967–1.069)	1.005 (0.953–1.070)	1.009 (0.971–1.039)
2014–2015	0.978 (0.907–1.016)	0.990 (0.906–1.028)	0.988 (0.959–1.028)
2012–2015	1.016 (0.960–1.073)	1.006 (0.943–1.071)	1.010 (0.967–1.046)

The bootstrapped confidence intervals at 5% level of significance are reported in parentheses, where ** symbolizes significance at the 5% level

not affect TFP if the inputs are fully flexible; that is, there should be a zero correlation if this is fulfilled. The interpretation of the previously presented result indicates that the MPI and its components are not independent of the changes in workload. In Table 5, the MPI and decomposed factors are correlated with the rate of change in the caseload.

From Table 5, it can be observed that the MPI and its components, in all cases, are positive. To a large extent, the MPI and its components also have statistically significant correlation with the rate of change in the caseload. A positive correlation can mean either the inputs do not decrease enough when the demand for justice services declines or the employees work harder when the demand increases, generating increased output for the given inputs. Each of these reasons indicate a slack in the courts; that is, more can be produced without increasing the inputs. Schneider (2005) argues that the exclusion of the caseload generates an underestimation of TFP.

However, despite the positive correlation concluded in this section, we argue that the correct measure of TFP is what we reported in the main analysis. Nevertheless, the caseload can, at least, partly explain the results indicating that the inputs are not flexible enough. The positive relationship between the MPI and the rate of change in the caseload is also in line with Beenstock and Haitovsky (2004), who argue that individual productivity increases when the work pressure is high. These results strengthen the previous argument of low flexibility in inputs. However, it should be interpreted carefully, since no causality can be concluded, and other factors are likely to affect the TFP change, which is not included here.

7 Conclusion and policy recommendations

This paper aimed to investigate the development of TFP from 2012 to 2015. The differences in comparison with previous research are: (1) more detailed data are used, which allow the outputs to be weighted based on the hearing time; and (2) TFP is decomposed into four components, in contrast to a maximum of three in the earlier literature.

The findings indicate a 1.7% decline in TFP, which is measured as an annual geometric mean. However, a substantial variation between courts is found; for example, 36–57% of the courts have a negative change in TFP, while 16–36% of the courts have a positive TFP change, depending on the year. The negative TFP change is mainly driven

Table 5 Spearman correlations between MPI, TC and EC and the rate of change in the caseload

Year	Malmquist productivity index (MPI)	Technical change (TC)	Efficiency change (EC)
2012–2013	0.672*** (0.000)	0.328** (0.030)	0.452*** (0.002)
2013–2014	0.511*** (0.000)	0.337** (0.025)	0.176 (0.253)
2014–2015	0.553*** (0.000)	0.270* (0.077)	0.450*** (0.002)

P values are reported in parentheses. ***symbolizes significance at 1%, **significance at 5% and *significance at 10%

by a decline in TC during the first period. Looking at the components of TC, it can be observed that most of the decline has its source in pure TC that is argued to be assigned to a decline in the caseload. However, the period 2014–2015 has a negative TFP change, occurring from a decline in pure TC, pure EC, and scale EC. Likely, this decline is also due to a smaller demand of justice services. However, the different components are differently affected, depending on where the court operates in relation to the frontier. Furthermore, the correlation between TFP and the rate of change of caseload is concluded to be positive and significant, which strengthens the previous argument and, therefore, indicates a non-sufficient level of flexibility in inputs.

The policy conclusion is that there is room for improvements. A recommendation is that district courts with negative TFP could learn from those with positive TFP in aspects of organisation and internal development of working methods. Furthermore, since the smallest courts have the largest volatility in TFP change, smoother changes can be achieved by merging courts, which would improve TFP. However, merging is, to some extent, constrained by the social and geographical issues that need to be taken into consideration. To avoid this issue, a less controversial policy implication that achieves more flexibility in the Swedish district courts is to develop the back-up labour force, introduced in 2012, to include other personnel than judges. This will allow the inputs to be adjusted when the demand fluctuates, which generates a higher degree of flexibility on the regional level.

In particular, this will enhance the flexibility of the small courts. The smallest courts have close to the minimum number of employees. Small courts are, by construction, more sensitive to changes in the workload, since a small change in the justice demand generates a large share of the percentage. Therefore, the issue of large volatility in the justice demand could, at least, partly be solved by an expansion of the back-up labour force to enhance flexibility. Furthermore, more flexible inputs across Sweden could potentially make it possible to eliminate the requirement of a minimum number of employees in each court. Instead, the volatility in the smallest courts can be served by flexible personnel (e.g., the back-up labour force).

Finally, peer comparisons of courts could be used in many potential aspects of the work for improving efficiency and productivity. For example, differences can be present that are not directly possible to determine in this study, such as organisational problems. This is, however, an aspect that can be taken into consideration in future research.

Acknowledgements We are grateful for received comments during the 15th International Conference on DEA in Prague and a seminar at Linnaeus University. Finally, we would like to thank the anonymous referees for insightful comments.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

Illustration of DEA

Figure 2 below shows the production, under constant returns to scale (CRS) and variable returns to scale (VRS), when one input is used to produce one output.

Figure 2 shows the production frontier for CRS and VRS during year t and year $t + 1$, respectively. Year $t + 1$ is on a higher level of output, given the level of input, meaning that technical change (TC) has occurred. Furthermore, increasing returns to scale (IRS) can be observed, assuming VRS, to the left of the tangency between the CRS and the VRS frontier, indicating that the firms using less than this level of input are too small. Similarly, to the right-hand side of the tangency point, there are decreasing returns to scale, meaning that firms observed there would increase their productivity by becoming smaller.

Results including all courts

To incorporate information of the results including all courts, the geometric means of all years for the individual courts, excluding the outliers, are reported in Table 6.

Finally, the total factor productivity (TFP) change and its confidence intervals are reported for each court and year, excluding the outliers, in Figs. 3, 4, and 5.

Fig. 2 Graphical illustration of DEA. Drawn based on Coelli et al. (2005), figure 11.1

Table 6 Geometric mean of MPI and its components, excluding the outliers, 2012–2015

Court	MPI	TC	Pure TC	Scale TC	EC	Pure EC	Scale EC
Alingsås	0.936 (0.915–0.945)	0.992 (0.947–1.024)	0.973 (0.910–1.001)	1.020 (0.989–1.069)	0.943 (0.895–0.988)	0.952 (0.912–1.012)	0.990 (0.929–1.027)
Attunda	1.029 (1.011–1.036)	1.029 (0.983–1.092)	1.022 (0.969–1.079)	1.006 (0.977–1.047)	1.000 (0.909–1.045)	1.000 (0.917–1.053)	1.000 (0.956–1.027)
Blekinge	0.940 (0.923–0.972)	0.940 (0.870–0.995)	0.946 (0.871–0.997)	0.994 (0.974–1.033)	1.000 (0.925–1.094)	1.000 (0.930–1.101)	1.000 (0.957–1.014)
Borås	0.997 (0.975–1.019)	0.968 (0.928–1.003)	0.975 (0.937–1.010)	0.993 (0.975–1.010)	1.030 (0.980–1.085)	1.027 (0.976–1.078)	1.003 (0.984–1.023)
Eskilstuna	0.969 (0.938–1.003)	0.969 (0.897–1.008)	0.966 (0.896–1.009)	1.002 (0.983–1.018)	1.000 (0.939–1.088)	1.000 (0.931–1.084)	1.000 (0.987–1.020)
Falun	0.981 (0.947–1.005)	0.981 (0.928–1.034)	0.987 (0.932–1.038)	0.994 (0.982–1.010)	1.000 (0.927–1.066)	1.000 (0.929–1.072)	1.000 (0.983–1.004)
Gällivare	0.970 (0.944–0.994)	0.932 (0.886–0.975)	–	–	1.042 (0.976–1.099)	1.000 (0.916–1.097)	1.042 (0.920–1.124)
Gävle	1.029 (0.977–1.066)	1.007 (0.961–1.062)	1.009 (0.967–1.067)	0.998 (0.982–1.008)	1.022 (0.912–1.087)	1.013 (0.900–1.073)	1.009 (1.001–1.018)
Göteborg	1.017 (1.009–1.027)	1.002 (0.958–1.045)	1.020 (0.945–1.091)	0.983 (0.908–1.045)	1.015 (0.969–1.059)	1.000 (0.917–1.076)	1.015 (0.940–1.089)
Halmstad	0.968 (0.949–0.985)	0.966 (0.930–0.994)	0.969 (0.937–1.000)	0.996 (0.982–1.005)	1.002 (0.965–1.042)	1.000 (0.961–1.040)	1.001 (0.994–1.009)
Haparanda	0.998 (0.968–1.018)	0.970 (0.931–1.004)	–	–	1.029 (0.978–1.072)	1.000 (0.917–1.097)	1.029 (0.906–1.115)
Helsingborg	0.981 (0.967–0.997)	0.978 (0.932–1.013)	0.985 (0.938–1.022)	0.992 (0.969–1.019)	1.003 (0.962–1.054)	1.000 (0.955–1.058)	1.003 (0.968–1.023)

Table 6 continued

Court	MPI	TC	Pure TC	Scale TC	EC	Pure EC	Scale EC
Hudiksvall	0.987 (0.967–1.015)	1.002 (0.952–1.048)	0.981 (0.929–1.015)	1.022 (1.001–1.051)	0.985 (0.950–1.030)	0.984 (0.950–1.036)	1.001 (0.965–1.029)
Hässelholm	0.980 (0.958–1.001)	0.940 (0.904–0.978)	1.003 (0.960–1.064)	0.938 (0.881–0.974)	1.043 (0.987–1.091)	1.031 (0.938–1.076)	1.011 (0.960–1.082)
Jönköping	0.992 (0.979–1.004)	0.979 (0.946–1.005)	0.985 (0.946–1.015)	0.994 (0.964–1.029)	1.013 (0.982–1.047)	1.009 (0.976–1.053)	1.004 (0.961–1.036)
Kalmar	1.018 (1.005–1.037)	0.989 (0.950–1.026)	1.000 (0.967–1.032)	0.989 (0.954–1.018)	1.030 (0.991–1.069)	1.020 (0.983–1.057)	1.009 (0.977–1.045)
Kristianstad	1.011 (0.998–1.046)	0.975 (0.938–1.019)	0.970 (0.937–1.011)	1.004 (0.989–1.020)	1.038 (0.993–1.092)	1.033 (0.984–1.080)	1.005 (0.995–1.024)
Linköping	1.012 (1.002–1.032)	0.979 (0.947–1.008)	0.998 (0.968–1.034)	0.980 (0.940–1.011)	1.034 (1.004–1.073)	1.021 (0.985–1.059)	1.013 (0.976–1.063)
Luleå	0.997 (0.954–1.029)	0.978 (0.9351.036)	0.985 (0.943–1.045)	0.993 (0.981–1.004)	1.020 (0.924–1.081)	1.018 (0.921–1.080)	1.002 (0.990–1.008)
Lund	0.989 (0.972–1.004)	0.989 (0.920–1.045)	0.986 (0.920–1.037)	1.003 (0.979–1.027)	1.000 (0.927–1.067)	1.000 (0.928–1.069)	1.000 (0.973–1.029)
Lycksele	1.004 (0.993–1.020)	0.931 (0.865–0.971)	–	–	1.079 (1.030–1.153)	1.000 (0.915–1.094)	1.079 (0.963–1.190)
Malmö	0.993 (0.977–1.020)	0.983 (0.941–1.009)	0.989 (0.904–1.045)	0.994 (0.919–1.070)	1.010 (0.978–1.061)	1.000 (0.926–1.088)	1.010 (0.910–1.095)
Mora	0.877 (0.846–0.912)	0.905 (0.843–0.943)	0.882 (0.817–0.916)	1.026 (0.986–1.072)	0.969 (0.913–1.051)	1.000 (0.944–1.093)	0.969 (0.924–1.015)
Norrköping	1.004 (0.986–1.019)	0.964 (0.926–0.983)	0.973 (0.940–0.995)	0.991 (0.974–1.004)	1.041 (1.018–1.083)	1.032 (1.002–1.069)	1.009 (0.995–1.028)

Table 6 continued

Court	MPI	TC	Pure TC	Scale TC	EC	Pure EC	Scale EC
Norrälje	0.846 (0.814–0.890)	0.846 (0.789–0.887)	0.867 (0.784–0.924)	0.976 (0.934–1.012)	1.000 (0.922–1.100)	1.000 (0.917–1.094)	1.000 (1.000–1.023)
Nyköping	0.987 (0.968–1.007)	0.987 (0.913–1.049)	0.985 (0.903–1.047)	1.003 (0.996–1.019)	1.000 (0.914–1.084)	1.000 (0.917–1.088)	1.000 (0.985–1.001)
Skaraborg	1.027 (1.014–1.040)	1.003 (0.964–1.038)	1.013 (0.977–1.053)	0.990 (0.970–1.008)	1.024 (0.981–1.068)	1.009 (0.955–1.046)	1.015 (0.995–1.044)
Skellefteå	0.922 (0.899–0.944)	0.920 (0.880–0.959)	0.950 (0.869–1.008)	0.969 (0.920–1.034)	1.002 (0.941–1.060)	1.000 (0.912–1.093)	1.002 (0.918–1.060)
Solna	0.975 (0.965–0.992)	0.975 (0.898–1.036)	0.975 (0.887–1.038)	1.000 (0.988–1.025)	1.000 (0.927–1.079)	1.000 (0.924–1.087)	1.000 (0.979–1.010)
Stockholm	1.028 (1.013–1.052)	0.973 (0.934–1.020)	1.055 (0.948–1.129)	0.922 (0.844–1.018)	1.057 (1.002–1.102)	1.000 (0.918–1.090)	1.057 (0.936–1.145)
Sundsvall	0.927 (0.902–0.951)	0.941 (0.900–0.969)	0.944 (0.911–0.968)	0.996 (0.975–1.014)	0.985 (0.952–1.027)	0.981 (0.948–1.020)	1.004 (0.980–1.027)
Södertälje	1.013 (0.979–1.040)	0.979 (0.941–1.013)	0.970 (0.931–1.005)	1.009 (0.989–1.032)	1.035 (0.989–1.078)	1.039 (0.985–1.084)	0.996 (0.979–1.014)
Södertörn	0.982 (0.961–1.014)	0.982 (0.938–1.016)	0.966 (0.863–1.026)	1.016 (0.948–1.102)	1.000 (0.967–1.051)	1.000 (0.923–1.092)	1.000 (0.917–1.074)
Umeå	1.028 (1.008–1.053)	0.975 (0.943–1.009)	0.983 (0.953–1.016)	0.991 (0.977–1.009)	1.055 (1.017–1.093)	1.056 (1.021–1.099)	0.999 (0.974–1.013)
Uppsala	0.967 (0.955–0.975)	0.974 (0.910–1.019)	0.975 (0.888–1.028)	0.999 (0.966–1.038)	0.993 (0.937–1.057)	0.996 (0.931–1.089)	0.997 (0.948–1.028)
Varberg	0.946 (0.928–0.956)	0.966 (0.924–0.994)	0.961 (0.917–0.990)	1.004 (0.989–1.019)	0.980 (0.945–1.020)	0.987 (0.951–1.033)	0.993 (0.972–1.010)

Table 6 continued

Court	MPI	TC	Pure TC	Scale TC	EC	Pure EC	Scale EC
Vänersborg	0.986 (0.972–0.998)	0.915 (0.859–0.957)	0.995 (0.937–1.038)	0.919 (0.868–0.962)	1.078 (1.024–1.139)	1.004 (0.951–1.059)	1.074 (1.020–1.128)
Värmland	0.997 (0.971–1.018)	0.965 (0.914–1.026)	1.001 (0.936–1.073)	0.965 (0.939–0.998)	1.034 (0.941–1.094)	1.000 (0.904–1.070)	1.034 (0.989–1.051)
Västmanland	0.967 (0.949–0.983)	0.950 (0.910–0.973)	0.963 (0.922–0.988)	0.987 (0.943–1.028)	1.018 (0.985–1.061)	1.005 (0.973–1.049)	1.012 (0.962–1.063)
Växjö	0.998 (0.981–1.014)	0.992 (0.932–1.059)	1.005 (0.942–1.079)	0.988 (0.954–1.011)	1.005 (0.921–1.072)	1.000 (0.907–1.068)	1.005 (0.984–1.037)
Ystad	1.015 (0.992–1.045)	1.007 (0.965–1.047)	1.006 (0.962–1.046)	1.001 (0.990–1.011)	1.008 (0.961–1.065)	1.006 (0.955–1.064)	1.002 (0.989–1.017)
Ångermanland	0.909 (0.884–0.930)	0.909 (0.850–0.955)	0.916 (0.860–0.967)	0.993 (0.961–1.017)	1.000 (0.927–1.067)	1.000 (0.917–1.073)	1.000 (0.964–1.027)
Örebro	0.956 (0.936–0.983)	0.980 (0.935–1.016)	0.961 (0.906–0.995)	1.019 (0.981–1.069)	0.976 (0.931–1.029)	1.000 (0.948–1.062)	0.976 (0.924–1.022)
Östersund	0.991 (0.974–1.007)	0.974 (0.923–1.031)	0.972 (0.930–1.003)	1.002 (0.964–1.047)	1.017 (0.952–1.066)	1.017 (0.980–1.059)	1.001 (0.948–1.045)

The courts with missing observations for pure and scale TC have a missing value for at least one of the observed periods

Fig. 3 MPI and its confidence interval, excluding outliers, 2012–2013

Fig. 4 MPI and its confidence interval, excluding outliers, 2013–2014

Fig. 5 MPI and its confidence interval, excluding outliers, 2014–2015

References

- Agrell, P. J., & Niknazar, P. (2014). Structural and behavioral robustness in applied best-practice regulation. *Socio-Economic Planning Sciences*, 48(1), 89–103.
- Andersson, C., Antelius, J., Månsson, J., & Sund, K. (2017a). Technical efficiency and productivity for higher education institutions in Sweden. *Scandinavian Journal of Educational Research*, 61(2), 205–223.
- Andersson, C., Bonander, F., & Månsson, J. (2017b). *Tingsrätters effektivitet och produktivitet (The efficiency and productivity of Swedish district courts)*. Stockholm: Swedish National Audit Office.
- Andersson, C., Månsson, J., & Sund, K. (2014). Technical efficiency of Swedish employment offices. *Socio-Economic Planning Sciences*, 48(1), 57–64.
- Arnek, M., Melin, T., & Norrlid, A. (2016). *Mer än tur i struktur – en ESO-rapport om kommunal effektivitet (More than luck in structure—An ESO-report regarding efficiency within municipalities)*. Stockholm: Elanders Sverige AB.
- Balk, B. M. (2001). Scale efficiency and productivity change. *Journal of Productivity Analysis*, 15(3), 159–183.
- Balk, B. M., & Althin, R. (1996). A new, transitive productivity index. *Journal of Productivity Analysis*, 7(1), 19–27.
- Banker, R. D. (1984). Estimating most productive scale size using data envelopment analysis. *European Journal of Operational Research*, 17(1), 35–44.
- Banker, R. D., & Chang, H. (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. *European Journal of Operational Research*, 175(2), 1311–1320.
- Banker, R. D., Chang, H., & Zheng, Z. (2017). On the use of super-efficiency procedures for ranking efficient units and identifying outliers. *Annals of Operations Research*, 250(1), 21–35.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092.
- Banker, R. D., & Gifford, J. L. (1988). *A relative efficiency model for the evaluation of public health nurse productivity*. Pittsburgh: Carnegie Mellon University.
- Beenstock, M., & Haitovsky, Y. (2004). Does the appointment of judges increase the output of the judiciary? *International Review of Law and Economics*, 24(3), 351–369.

- Berg, S., Førsund, F., & Jansen, E. (1992). Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980–89. *The Scandinavian Journal of Economics*, 94(Supplement), S211–S228.
- Bjurek, H., Kjulin, U., & Gustafsson, B. (1992). Efficiency, productivity and determinants of inefficiency at public day care centers in Sweden. *The Scandinavian Journal of Economics*, 94(Supplement), S173–S187.
- Blank, J., van der Ende, M., van Hulst, B., & Jagtenberg, R. (2004). *Benchmarking in an international perspective: An international comparison of the mechanisms and performance of the judiciary system*. Rotterdam: Netherlands Council for the Judiciary.
- Bogetoft, P., & Otto, L. (2015). Package ‘benchmarking’. Retrieved January 04, 2016 from <https://cran.r-project.org/web/packages/Benchmarking/Benchmarking.pdf>.
- Briec, W., & Kerstens, K. (2009). Infeasibility and directional distance functions with application to the determinateness of the Luenberger productivity indicator. *Journal of Optimization Theory and Applications*, 141(1), 55–73.
- Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica: Journal of the Econometric Society*, 50(6), 1393–1414.
- Chambers, R. G., & Pope, R. D. (1996). Aggregate productivity measures. *American Journal of Agricultural Economics*, 78(5), 1360–1365.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
- Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. New York: Springer.
- Davies, L., & Gather, U. (1993). The identification of multiple outliers. *Journal of the American Statistical Association*, 88(423), 782–792.
- 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.
- Edvardsen, D. F., Førsund, F. R., & Kittelsen, S. A. (2017). Productivity development of Norwegian institutions of higher education 2004–2013. *Journal of the Operational Research Society*, 68(4), 399–415.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26.
- Eisenstein, J., & Jacob, H. (1977). *Felony justice: An organizational analysis of criminal courts*. Boston: Little, Brown.
- Elbially, N., & García-Rubio, M. A. (2011). *Assessing judicial efficiency of Egyptian first instance courts: A DEA analysis* (No. 19-2011). Joint discussion paper series in economics.
- Espasa, M., & Esteller-Moré, A. (2015). Analyzing judicial courts' performance: Inefficiency vs. congestion. *Revista de Economía Aplicada*, 23(69), 61–82.
- Falavigna, G., Ippoliti, R., Manello, A., & Ramello, G. B. (2015). Judicial productivity, delay and efficiency: A directional distance function (DDF) approach. *European Journal of Operational Research*, 240(2), 592–601.
- Falavigna, G., Ippoliti, R., & Ramello, G. B. (2017). DEA-based Malmquist productivity indexes for understanding courts reform. *Socio-Economic Planning Sciences*, 1–13 (forthcoming).
- Färe, R. (1988). *Fundamentals of production theory*. Berlin: Springer-Verlag.
- Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1992). Productivity changes in Swedish pharmacies 1980–1989: A non-parametric Malmquist approach. In T. R. Gulledge & C. A. K. Lovell (Eds.), *International Applications of Productivity and Efficiency Analysis* (pp. 81–97). Dordrecht: Springer.
- Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1994a). Productivity developments in Swedish hospitals: A malmquist output index approach. In A. Charnes, W. W. Cooper, A. Y. Lewin, & L. M. Seiford (Eds.), *Data envelopment analysis: Theory, methodology, and applications* (pp. 253–272). Dordrecht: Springer.
- Färe, R., Grosskopf, S., & Margaritis, D. (2008). Efficiency and productivity: Malmquist and more. In H. O. Fried, C. A. K. Lovell, & S. S. Schmidt (Eds.), *The measurement of productive efficiency and productivity growth* (pp. 522–622). New York: Oxford University Press.
- Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994b). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 84(1), 66–83.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), 253–290.

- Fauvreille, T. A., & Almeida, A. T. (2016). *Determinants of judicial efficiency change: Evidence from Brazil*. SSRN. Retrieved December 28, 2017 from <https://ssrn.com/abstract=2745269>.
- Feeley, M. M. (1973). Two models of the criminal justice system: An organizational perspective. *Law & Society Review*, 7(3), 407–425.
- Feld, L. P., & 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.
- Ferrandino, J. A. (2010). *Integrating theory, practice and policy: The technical efficiency and productivity of Florida's circuit courts*. Orlando: University of Central Florida.
- Ferrandino, J. (2014). Testing the Packer theorem: The efficiency of Florida's criminal circuit courts. *American Journal of Criminal Justice*, 39(2), 375–393.
- Finocchiaro Castro, M. F., & 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. F., & Guccio, C. (2016). Measuring potential efficiency gains from mergers of Italian first instance courts through nonparametric model. *Public Finance Review*, 46(1), 83–116.
- Førsund, F. R. (2016). Productivity interpretations of the Farrell efficiency measures and the Malmquist index and its decomposition. In J. Aparicio, C. Lovell, & J. Pastor (Eds.), *Advances in efficiency and productivity* (Vol. 249, pp. 121–147)., International series in operations research & management science Cham: Springer.
- Førsund, F. R., Edvardsen, D. F., & Kittelsen, S. A. (2015). Productivity of tax offices in Norway. *Journal of Productivity Analysis*, 43(3), 269–279.
- Frisch, R. (1965). *Theory of production*. Dordrecht: D. Reidel.
- Gilbert, R. A., & Wilson, P. W. (1998). Effects of deregulation on the productivity of Korean banks. *Journal of Economics and Business*, 50(2), 133–155.
- Gini, C. (1931). On the circular test of index numbers. *Metron*, 9(2), 3–24.
- Grifell-Tatjé, E., & Lovell, C. K. (1995). A note on the Malmquist productivity index. *Economics Letters*, 47(2), 169–175.
- Grosskopf, S. (2003). Some remarks on productivity and its decompositions. *Journal of Productivity Analysis*, 20(3), 459–474.
- Kapelko, M., & Oude Lansink, A. (2015). An international comparison of productivity change in the textile and clothing industry: A bootstrapped Malmquist index approach. *Empirical Economics*, 48(4), 1499–1523.
- Kim, D., & Min, H. (2016). Appeal rate and caseload: Evidence from civil litigation in Korea. *European Journal of Law and Economics*, 44(2), 1–22.
- Kittelsen, S. A. C., & Førsund, F. R. (1992). Efficiency analysis of Norwegian district courts. *Journal of Productivity Analysis*, 3(3), 277–306.
- Krüger, J. J. (2012). A Monte Carlo study of old and new frontier methods for efficiency measurement. *European Journal of Operational Research*, 222(1), 137–148.
- Kumbhakar, S. C., & Lovell, C. A. K. (2003). *Stochastic frontier analysis*. Cambridge, UK: Cambridge University Press.
- Lewin, A. Y., Morey, R. C., & Cook, T. J. (1982). Evaluating the administrative efficiency of courts. *Omega*, 10(4), 401–411.
- Lovell, C. A. K. (2003). The decomposition of Malmquist productivity indexes. *Journal of Productivity Analysis*, 20(3), 437–458.
- Major, W. (2015). Data envelopment analysis as an instrument for measuring the efficiency of courts. *Operations Research and Decisions*, 25(4), 19–34.
- Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estadística*, 4(2), 209–242.
- Messick, R. E. (1999). Judicial reform and economic development: A survey of the issues. *The World Bank Research Observer*, 14(1), 117–136.
- Ministry of Justice. (2015). *The Swedish judicial system*. Retrieved September 20, 2017 from <http://www.regeringen.se/49e9fa/contentassets/24ef36d9206a41fb80d6b74030fd9ef6/the-swedish-judicial-system.pdf>.
- Nardulli, P. F. (1978). *The courtroom elite: An organizational perspective on criminal justice*. Cambridge, MA: Ballinger Publishing Company.
- Nishimizu, M., & Page, J. M. (1982). Total factor productivity growth, technological progress and technical efficiency change: Dimensions of productivity change in Yugoslavia, 1965–78. *The Economic Journal*, 92(368), 920–936.

- Nissi, E., & Rapposelli, A. (2010). A data envelopment analysis of Italian courts efficiency. *Statistica Applicata - Italian Journal of Applied Statistics*, 22(2), 199–210.
- Nordén, E. (2015). *Utvecklingen av personuppklarade misstankar och bifalla åtal 2004–2014* (The development of solved suspicious and approved prosecutions 2004–2014). Kortanalys. Brottsförbyggande rådet, Enheten för kommunikation. Retrieved December 28, 2017 from https://bra.se/download/18.31d7fffa1504bbffea0a7ff1/1450358617656/2015_Utvecklingen+av+personuppklarade.pdf.
- Olesen, O. B., & Petersen, N. C. (2016). Stochastic data envelopment analysis—A review. *European Journal of Operational Research*, 251(1), 2–21.
- Pastor, J. T., & Lovell, C. A. K. (2007). Circularity of the Malmquist productivity index. *Economic Theory*, 33(3), 591–599.
- Pedraja-Chaparro, F., & Salinas-Jimenez, 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.
- Ray, S. C. (2001). *On an extended decomposition of the Malmquist productivity index*. Paper presented at the seventh European workshop on efficiency and productivity analysis. Universidad de Oviedo. Retrieved December 20, 2017 from <https://pdfs.semanticscholar.org/de87/c7db60ccc86026796d2261355cd4ecb9fd25.pdf>.
- Ray, S. C., & Desli, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: Comment. *The American Economic Review*, 87(5), 1033–1039.
- Rigsrevisionen. (2000). *Beretning til statsrevisorerne om retternes produktivitet mv* (Report to the public auditors about the productivity of the judiciary). Retrieved December 28, 2017 from <http://www.rigsrevisionen.dk/media/1877192/5-99.pdf>.
- Santos, S. P., & Amado, C. A. F. (2014). On the need for reform of the Portuguese judicial system—Does data envelopment analysis assessment support it? *Omega*, 47, 1–16.
- Schneider, M. R. (2005). Judicial career incentives and court performance: An empirical study of the German labour courts of appeal. *European Journal of Law and Economics*, 20(2), 127–144.
- SFS 1981:774. *The Swedish Enforcement Code* (Den svenska utsökningsbalken). Stockholm: Ministry of Justice.
- SFS 1996:242. *The Swedish Court Matters Act* (Den svenska lagen om domstolsärenden). Stockholm: Ministry of Justice.
- SFS 2011:203. *The Swedish Budget Act, Chapter 1, 3§* (Den svenska budgetlagen, kapitel 1, 3§). Stockholm: Ministry of Justice.
- Shephard, R. W. (1953). *Cost and production functions*. Princeton, NJ: Princeton University Press.
- Shephard, R. W. (1970). *Theory of cost and production functions*. Princeton, NJ: Princeton University Press.
- Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3), 391–424.
- Simar, L., & Wilson, P. W. (1998a). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44(1), 49–61.
- Simar, L., & Wilson, P. W. (1998b). *Productivity growth in industrialized countries*. Université Catholique de Louvain, Center for Operations Research and Econometrics [CORE]. Retrieved December 28, 2017 from <https://pdfs.semanticscholar.org/d7fd/bbc6dbb17a6e3b5a49a4dfe494e9692a3f76.pdf>.
- Simar, L., & Wilson, P. W. (1999). Estimating and bootstrapping Malmquist indices. *European Journal of Operational Research*, 115(3), 459–471.
- Simar, L., & Wilson, P. W. (2000). Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis*, 13(1), 49–78.
- Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of econometrics*, 136(1), 31–64.
- SNCA (Domstolsverket). (2012). *Årsredovisning för Sveriges domstolar* (Annual report for Swedish district courts). Jönköping.
- SNCA (Domstolsverket). (2013). *Årsredovisning för Sveriges domstolar* (Annual report for Swedish district courts). Jönköping.
- SNCA (Domstolsverket). (2014). *Årsredovisning för Sveriges domstolar* (Annual report for Swedish district courts). Jönköping.

- SNCA (Domstolsverket). (2015). *Årsredovisning för Sveriges domstolar* (Annual report for Swedish district courts). Jönköping.
- SNCA (Domstolsverket). (2016). *Årsredovisning för Sveriges domstolar* (Annual report for Swedish district courts). Jönköping.
- Sturup, J., Rostami, A., Gerell, M., & Sandholm, A. (2018). Near-repeat shootings in contemporary Sweden 2011 to 2015. *Security Journal*, 31(1), 73–92.
- The Swedish Agency for Public Management. (2007). Sammanslagna tingsrätter – en utvärdering (p. 09).
- Tyrefors, B. H., Palme, M., & Priks, M. (2017). *Age-dependent court sentences and crime bunching: Empirical evidence from Swedish administrative data* (IFN Working Paper No. 1163). Research Institute of Industrial Economics. Retrieved December 28, 2017 from <http://www.ifn.se/wfiles/wp/wp1163.pdf>.
- Voigt, S. (2016). Determinants of judicial efficiency: A survey. *European Journal of Law and Economics*, 42(2), 183–208.
- Wagner, J. M., & Shimshak, D. G. (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *European Journal of Operational Research*, 180(1), 57–67.
- Wheelock, D. C., & Wilson, P. W. (1999). Technical progress, inefficiency, and productivity change in US banking, 1984–1993. *Journal of Money, Credit and Banking*, 31(2), 212–234.
- Wilson, P. W. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business & Economic Statistics*, 11(3), 319–323.
- Wilson, P. W. (2008). FEAR: A software package for frontier efficiency analysis with R. *Socio-Economic Planning Sciences*, 42(4), 247–254.
- Yeung, L. L., & Azevedo, P. F. (2011). Measuring efficiency of Brazilian courts with data envelopment analysis (DEA). *IMA Journal of Management Mathematics*, 22(4), 343–356.
- Zofio, J. L., & Lovell, C. A. K. (1998). *Yet another Malmquist productivity index decomposition*, Mimeo. Department of Economics, Universidad Autonoma de Madrid, Madrid, University of Georgia, Athens, Georgia.