August 2021

WORKING PAPER SERIES

2021-EQM-05

Exploring Horizontal Mergers in Swedish District Courts Using Convex and Nonconvex Technologies: Usefulness of a Conservative Approach

Xiaoqing Chen
Corresponding author: School of Economics and Management, Southeast University, Nanjing, Jiangsu, China, and IESEG School of Management, 3 rue de la Digue, F-59000 Lille, France, 230189623@seu.edu.cn

Kristiaan Kerstens
Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Economie Management, F-59000 Lille, France, k.kerstens@ieseg.fr

Qingyuan Zhu
Nanjing University of Aeronautics & Astronautics, College of Economics and Management, Nanjing, China, zqyustc@mail.ustc.edu.cn

IÉSEG School of Management Lille Catholic University 3, rue de la Digue F-59000 Lille Tel: 33(0)3 20 54 58 92
www.ieseg.fr

Exploring Horizontal Mergers in Swedish District Courts Using Convex and Nonconvex Technologies: Usefulness of a Conservative Approach*

Xiaoqing Chen† Kristiaan Kerstens‡ Qingyuan Zhu§

22nd August 2021

Abstract

Swedish district courts have undergone a major mergers and acquisitions program between 2000 and 2010 to centralize activity in larger and fewer courts. The purpose of this contribution is to conduct an efficiency analysis of these courts to identify the eventual efficiency gains. Distinguishing mainly between technical and scale efficiency and determining the returns to scale of individual observations, we try to find the potential rationales behind this merger wave. We are to the best of our knowledge the first to combine traditional convex with nonconvex nonparametric frontier methods to calculate efficiency before and after the mergers. It turns out that the nonconvex methods provide a more cogent ex post explanation of this merger wave.

Keywords: Horizontal merger, Technical efficiency, Scale efficiency, Data Envelopment Analysis, Free Disposal Hull.

*Thanks to Ane Elixabete Ripoll-Zarraga and Jafar Sadeghi for constructive comments on previous versions. X. Chen gratefully acknowledges the financial support of the China Scholarship Council for funding her stay in Lille (202006090203).

†Corresponding author: School of Economics and Management, Southeast University, Nanjing, Jiangsu, China, and IÉSEG School of Management, 3 rue de la Digue, F-59000 Lille, France, 230189623@seu.edu.cn.

‡Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Économie Management, F-59000 Lille, France, k.kerstens@ieseg.fr

§Nanjing University of Aeronautics & Astronautics, College of Economics and Management, Nanjing, China, zqyustc@mail.ustc.edu.cn.

1 Introduction

Mergers and acquisitions (M&As) reflect a popular strategic choice for growth and expansion of organisational boundaries. Horizontal M&As take place between organisations working in the same market, while vertical M&As involve organisations operating in different markets upstream or downstream (see Gaughan (2007) for a more complete taxonomy). From a regulatory perspective (e.g., Belleflamme and Peitz (2010) or Viscusi, Harrington, and Vernon (2005)), since horizontal M&As reduce the number of competitors, they raise the possibility of creating market power implying social welfare losses. However, since horizontal M&As redefine the organisational boundaries by integration of the production facilities, there is also the possibility of achieving social welfare gains by cost reductions (assuming these are passed onto the final consumers). The main reason for horizontal M&As is economies of scale (advantages of production in higher volumes) and economies of scope (gains by changing input and/or output mix).

Horizontal M&As may raise the price due to an effect on the market power. Mergers can lead to substantial price increases if it makes collusion stable where before it was unstable. M&As may create cost savings by reshuffling the production of outputs across production facilities by exploiting cost differences, by using scale economies at a single plant, by creating synergies by pooling certain functions, by creating a larger innovative capacity leading to future efficiency gains, or by eliminating any eventual existing inefficiencies. It is well-known that the cost savings effect is often overruled by the market power effect (e.g., Farrell and Shapiro (1990)).

Outcomes of horizontal M&As are empirically evaluated using various methodologies. In the industrial organization literature, it is common to distinguish between event studies for stock market listed firms to assess shareholder value, direct price comparisons before and after the merger, and merger simulations using pre-merger market information to calibrate some noncooperative oligopoly model (see, e.g. Belleflamme and Peitz (2010, Section 15.4) for a broad overview or Budzinski and Ruhmer (2010) for a survey on merger simulations). This literature also recognises that technical and cost inefficiencies contribute to cost savings of horizontal M&As (see, e.g., Caves (2007) or Viscusi, Harrington, and Vernon (2005, p. 88-89) for a general argument and Akhavein, Berger, and Humphrey (1997) for an empirical study).

Since greater cost savings facilitate M&A being approved by the authorities, M&A participants have an incentive to overstate any eventual cost savings. Thus, since M&A participants are incentivised to overstate cost savings, it is important to obtain a conservative estimate. In this respect, we opt for nonconvex in addition to the traditional convex deterministic nonparametric production frontier models: the former exactly provide the most conservative estimates of efficiency gains available in the literature.

Indeed, our empirical evaluation tool is based on applied production analysis. In particular, deterministic nonparametric production frontier models (sometimes labeled as Data Envelopment Analysis (DEA)) are used to provide inner approximations of the boundaries of production possibility sets subject to a set of minimal axioms on what is deemed feasible (see Ray (2004)). Efficiency measures are used to position observations with respect to the boundary of such deterministic nonparametric production frontiers: either the observation is part of the boundary and technically efficient, or the observation is situated in the interior of the technology and it is technically inefficient (see Ray (2004)). This literature has led to evolved efficiency decompositions that fundamentally distinguish between technical and cost (in case of the cost function) efficiencies. Cost efficiency requires a point minimizing the linear cost function on the production frontier: an observation can be cost inefficient if it is situated away from this tangency point. Allocative efficiency closes the eventual gap between both cost efficiency and technical efficiency (see, for instance, Färe, Grosskopf, and Lovell (1985)): it indicates to which extent an observation deviates from the cost minimising input mix. This methodology is popular and has led to a large variety of empirical applications in a multitude of sectors (see Daraio, Kerstens, Nepomuceno, and Sickles (2020) for a recent meta-review) and it is a standard tool in the analysis of industrial organization (e.g., Caves (2007)).

In this deterministic nonparametric production frontier literature, various strands of literature analyse the potential ex ante and effective ex post efficiency gains of horizontal M&As. We provide a selective review of this literature, while focusing mainly on our own methodological choices for this contribution. We already mentioned some studies focusing on the public court sector that is the focus of our empirical application, but only dig deeper into this literature in the main body of the text.

We focus on technical efficiency measured with respect to a flexible or variable returns to scale technology, overall technical efficiency evaluated with regard to a constant returns to scale technology, and scale efficiency as a ratio of overall technical efficiency and technical efficiency. Scale efficiency evaluates the optimal scale level compatible with a long-run competitive equilibrium. It can be complemented with qualitative information on global returns to scale for individual observations. The standard reaction to such information on scale properties is that observations exhibiting increasing returns to scale should consider expanding, while observations showing decreasing returns to scale should contemplate contracting.

Studies adopting a similar methodology include the following examples. Cummins, Tennyson, and Weiss (1999) apply this frontier approach to determine technical efficiency and returns to scale in M&A in the US life insurance industry and find that merged firms realise greater efficiency gains than those that do not, and that firms with increasing returns to scale are more likely to be acquisition targets, among others. Harris, Ozgen, and Ozcan (2000) examine US hospitals using intertemporal production frontiers and show that M&As increase efficiency levels and that scale efficiency rather than technical efficiency is the main source of improved performance. Similar studies on courts (e.g., Agrell, Mattsson, and Månsson (2020), Castro and Guccio (2018), Peyrache and Zago (2016), among others) are discussed later on when presenting our own empirical results. In a review Frantz (2015) states that there is no evidence that mergers improve technical efficiency, underscoring the regulatory need to scrutinize popular justifications.

In a similar vein, Bogetoft and Wang (2005) initiate a substantial literature by proposing a decomposition of the potential gains from merging into technical efficiency, size (scale), and harmony (mix) gains and illustrate this proposal using agricultural extension offices in Denmark showing that there are considerable expected gains. Kristensen, Bogetoft, and Pedersen (2010) conduct this decomposition to Danish hospitals and evaluate the potential gains from the planned M&As, thereby showing that many hospitals are technically inefficient and some merged hospitals are too large and experience decreasing returns to scale. One such study on courts is found in Mattsson and Tidanå (2019).

Other studies assess the effect of M&As on productivity growth. Krishnasamy, Ridzwa, and Perumal (2004) analyze productivity growth of ten merged Malaysian banks using the Malmquist productivity index in the short period 2000-2001 and find that productivity increases in eight out of ten banks. Monastyrenko (2017) computes an eco-efficiency Malmquist productivity index among European electricity producers in the period 2005–2013 and finds that the heavily regulated domestic horizontal M&As have no impact, while the horizontal cross-border M&As damage eco-efficiency in the short run and become only positive in the medium run. Analogous studies focusing on courts (e.g., Falavigna, Ippoliti, and Ramello (2018), Mattsson, Månsson, Andersson, and Bonander (2018), among others) are presented in the empirical section.

As is common with the analysis of the public sector, we opt for an input-oriented efficiency measure since the outputs are determined by the demand for justice of citizens. However, in the literature, one can find several instances of articles focusing on output-oriented efficiency in courts (e.g., Castro and Guccio (2018) or Giacalone, Nissi, and Cusatelli (2020)). For our large unbalanced panel of Swedish district courts earlier analysed by Agrell, Mattsson, and Månsson (2020), Mattsson, Månsson, Andersson, and Bonander (2018) and Mattsson and Tidanå (2019), the approaches are mixed: Agrell, Mattsson, and Månsson (2020) use an input orientation, while Mattsson, Månsson, Andersson, and Bonander (2018) and Mattsson and Tidanå (2019) opt for output-oriented efficiency. We empirically demonstrate that these district courts do in fact control their inputs.

Already Farrell (1959) points out that the convexity assumption maintained in almost all production models precludes the various reasons that may generate nonconvexities in technology. First, indivisibilities point to the fact that inputs and outputs in production are not perfectly divisible and thus not continuous (see Scarf (1986; 1994)). These same indivisibilities may also limit the up- and especially the downscaling of production processes. Second, economies of scale and increasing returns to scale may yield nonconvex technologies where organisations have an interest to continue scaling up production. Third, economies of specialisation instead of economies of diversification may reveal gains in switching costs and time and yield nonconvex technologies. Fourth, both negative and positive externalities in production yield nonconvexities in the technology of the affected organisations. More recently, network externalities and nonrival inputs (like ideas) can be added as additional sources of nonconvexities.

It is often -implicitly or explicitly- assumed that nonconvexities have no impact on the estimates of the parameters of interest in production and, e.g., cost approaches alike. However, a basic deterministic nonparametric production frontier imposing flexible or variable returns to scale and dispensing with convexity has been originally developed by Deprins, Simar, and Tulkens (1984) (sometimes labeled Free Disposal Hull (FDH)). Kerstens and Vanden Eeckaut (1999) extend this basic nonconvex frontier by introducing constant, non-increasing and non-decreasing returns to scale assumptions. Moreover, these same authors propose a new goodness-of-fit approach to infer the characteristics of global returns to scale for nonconvex technologies. All these nonconvex nonparametric frontier technologies are smaller than the corresponding convex nonparametric frontier models and thus yield more conservative estimates of efficiency.

Furthermore, seminal contributions to axiomatic production theory indicate that the cost function is convex in the outputs if and only if technology is convex (e.g., Jacobsen (1970, Corollary 5.5)). Thus, using contraposition, the cost function is nonconvex if and only if technology is nonconvex. In fact, Briec, Kerstens, and Vanden Eeckaut (2004) propose nonconvex nonparametric cost frontiers with any returns to scale assumption and prove that these are always larger than or equal to the convex corresponding counterparts with similar returns to scale assumption: these are only identical under a single output and constant returns to scale. Kerstens and Van de Woestyne (2021) illustrate the potentially very substantial impact of convexity on cost function estimates.

In this contribution, we address the following sequence of three questions with regard to our empirical application based on a large unbalanced panel of Swedish district courts. First, are there any substantial changes in productivity in this sample that may have an impact on the assessment of the horizontal mergers? In the case of small or negligible productivity change, then we can safely ignore it when assessing horizontal mergers. For this purpose, we compute a Malmquist productivity index and test whether it measures any significant productivity change. Second, what are the effects of horizontal mergers on the overall technical efficiency as well as the technical and scale efficiencies under convex and nonconvex technologies? This question is addressed by computing the overall technical efficiency, the technical efficiency as well as the scale efficiency under convex and nonconvex technologies characterised by constant returns to scale and variable returns to scale. This may shed some light on the driving factors behind horizontal mergers. Third, what are the effects of the horizontal mergers on the global returns to scale characterization of these observations? To address this question, we derive qualitative information regarding the global returns to scale from the observations involved in the horizontal mergers.

For these purposes, this contribution is structured as follows. Section 2 provides some basic definitions of the traditional convex and the less widely applied nonconvex technologies. It also defines input-oriented efficiency measures for measuring overall technical efficiency, technical efficiency, and scale efficiency and describes how to determine global returns to scale information. Finally, it also defines the input-oriented Malmquist productivity index. After developing this theoretical framework, Section 3 describes the secondary unbalanced panel data set of Swedish district courts as well as the historical horizontal merger process that took place during the years 2000 till 2009. Section 4 with the empirical illustrations first presents results on the Malmquist productivity index and checks whether there is any substantial technological change in these courts. Then, we present convex and nonconvex estimates of overall technical efficiency and its decomposition into technical and scale efficiency at the sample level, and at the level of the years when horizontal mergers happened and the years thereafter. We also investigate returns to scale information under convex and nonconvex estimates. Finally, we repeat the same analysis at the level of the pre-merger and the post-merger observations. Section 5 concludes and outlines future research issues.

2 Nonparametric Technologies and Efficiency

This study uses traditional convex and nonconvex nonparametric, deterministic frontier methods to determine the static input-oriented efficiency of each operating unit, and also adopts the Malmquist productivity index to analyze the dynamics of productivity change in discrete time. We first introduce the static efficiency methods, and then the dynamic productivity index.

2.1 Nonparametric Technology Frontiers: A Unified Representation

Consider a set of K observations $\boldsymbol{A} = \{(x_1, y_1), ..., (x_K, y_K)\} \in \mathbb{R}_+^{m+n}$. A production technology describes all available possibilities to transform input vectors $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{R}_+^m$ into output vectors $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathbb{R}_+^n$. The production possibility set or technology $\boldsymbol{S}$ summarizes the set of all feasible input and output vectors: $\boldsymbol{S} = \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m+n} : \boldsymbol{x} \text{ can produce } \boldsymbol{y}\}$. Given our focus on input-oriented efficiency measurement later on, this technology can be represented by the input correspondence $L : \mathbb{R}_+^n \to 2^{\mathbb{R}_+^m}$ where $L(\boldsymbol{y})$ is the set of all input vectors that yield at least the output vector $\boldsymbol{y}$:

$$L(\boldsymbol{y}) = \{\boldsymbol{x} : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S}\}. \tag{1}$$

Nonparametric specifications of technology can be estimated by enveloping these K observations in the set $\boldsymbol{A}$ while maintaining some basic production axioms (see Hackman (2008) or Ray (2004)). We are interested in defining minimum extrapolation technologies satisfying strong disposability in the inputs and outputs, all four traditional returns to scale hypotheses (i.e., constant, non-increasing, non-decreasing and variable (flexible) returns to scale), including those technologies that satisfy the assumption of convexity and those that do not.

A unified algebraic representation of convex and nonconvex technologies under different returns to scale assumptions for a sample of K observations is found in Briec, Kerstens, and Vanden Eeckaut (2004):

$$\boldsymbol{S}^{\Lambda,\Gamma} = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m+n} : \boldsymbol{x} \geq \sum_{k=1}^{K} \alpha\, z_k\, \boldsymbol{x}_k,\ \boldsymbol{y} \leq \sum_{k=1}^{K} \alpha\, z_k\, \boldsymbol{y}_k,\ \sum_{k=1}^{K} z_k = 1,\ \boldsymbol{z} \in \Lambda, \alpha \in \Gamma \right\}, \tag{2}$$

where

$$
\begin{aligned}
\text{(i) } \Gamma &\equiv \Gamma^{\text{CRS}} = \{\alpha : \ \alpha \geq 0\}\,; \\
\text{(ii) } \Gamma &\equiv \Gamma^{\text{NDRS}} = \{\alpha : \ \alpha \geq 1\}\,; \\
\text{(iii) } \Gamma &\equiv \Gamma^{\text{NIRS}} = \{\alpha : \ 0 \leq \alpha \leq 1\}\,; \\
\text{(iv) } \Gamma &\equiv \Gamma^{\text{VRS}} = \{\alpha : \ \alpha = 1\}\,; \ \text{ and} \\
\text{(v) } \Lambda &\equiv \Lambda^{\text{C}} = \{\boldsymbol{z} = (z_1, \ldots, z_k) : z_k \geq 0\}\,, \ \text{ and (ii) } \Lambda \equiv \Lambda^{\text{NC}} = \{\boldsymbol{z} : z_k \in \{0,1\}\}\,.
\end{aligned}
$$

First, there is the activity vector ($\boldsymbol{z}$) operating subject to a convexity (C) or nonconvexity (NC) constraint. Second, there is a scaling parameter (α) allowing for a particular scaling of all K observations spanning the technology. This scaling parameter is smaller than or equal to 1 or larger than or equal to 1 under non-increasing returns to scale (NIRS) and non-decreasing returns to scale (NDRS) respectively, fixed at unity under variable returns to scale (VRS), and non-negative under constant returns to scale (CRS).

2.2 Input-Oriented Efficiency Measures and Estimating Returns to scale

The radial input efficiency measure can be defined as:

$$E_i^{\Lambda,\Gamma}\left(\boldsymbol{x}, \boldsymbol{y} : \boldsymbol{S}^{\Lambda,\Gamma}\right) = \min\left\{\theta \mid (\theta\, \boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S}^{\Lambda,\Gamma}, \theta \geq 0\right\}. \tag{3}$$

This efficiency measure indicates the minimum contraction of an input vector by a scalar θ while still producing the same outputs compatible with the technology $\boldsymbol{S}$. Obviously, the resulting input combination is located at the boundary of the input correspondence or technology. For our purpose, the radial input efficiency has two key properties (see, e.g., Hackman (2008)). First, it is smaller than or equal to unity ($0 < E_i^{\Lambda,\Gamma}\left(\boldsymbol{x}, \boldsymbol{y} : \boldsymbol{S}^{\Lambda,\Gamma}\right) \leq 1$), whereby efficient production on the isoquant of the input correspondence $L(\boldsymbol{y})$ is represented by unity and $1 - E_i^{\Lambda,\Gamma}\left(\boldsymbol{x}, \boldsymbol{y} : \boldsymbol{S}^{\Lambda,\Gamma}\right)$ indicates the amount of inefficiency. Second, it has a cost interpretation.

Definition 2.1. Under the assumptions on the technology $\boldsymbol{S}^{\Lambda,\Gamma}$ defined in (2), the following input-oriented efficiency notions can be distinguished:

- Technical Efficiency is the quantity: $TE_i^{\Lambda}(\boldsymbol{x}, \boldsymbol{y}) = E_i^{\Lambda,\text{VRS}}\left(\boldsymbol{x}, \boldsymbol{y} : \boldsymbol{S}^{\Lambda,\text{VRS}}\right)$;
- Overall Technical Efficiency is the quantity: $OTE_i^{\Lambda}(\boldsymbol{x}, \boldsymbol{y}) = E_i^{\Lambda,\text{CRS}}\left(\boldsymbol{x}, \boldsymbol{y} : \boldsymbol{S}^{\Lambda,\text{CRS}}\right)$;

- Scale Efficiency is the quantity: $SCE_i^{\Lambda}(\boldsymbol{x},\boldsymbol{y}) = E_i^{\Lambda,\text{CRS}}\left(\boldsymbol{x},\boldsymbol{y}:\boldsymbol{S}^{\Lambda,\text{CRS}}\right)/E_i^{\Lambda,\text{VRS}}\left(\boldsymbol{x},\boldsymbol{y}:\boldsymbol{S}^{\Lambda,\text{VRS}}\right)$.

Since $E_i^{\Lambda,\text{CRS}}\left(\boldsymbol{x},\boldsymbol{y}:\boldsymbol{S}^{\Lambda,\text{CRS}}\right) \leq E_i^{\Lambda,\text{VRS}}\left(\boldsymbol{x},\boldsymbol{y}:\boldsymbol{S}^{\Lambda,\text{VRS}}\right) \leq 1$, evidently $0 < SCE_i^{\Lambda}(\boldsymbol{x},\boldsymbol{y}) \leq 1$ (see Färe, Grosskopf, and Lovell (1983)). Using Definition 2.1, the following identity readily follows:

$$OTE_i^{\Lambda}(\boldsymbol{x},\boldsymbol{y}) = TE_i^{\Lambda}(\boldsymbol{x},\boldsymbol{y}) \cdot SCE_i^{\Lambda}(\boldsymbol{x},\boldsymbol{y}). \tag{4}$$

This decomposition simply states that Overall Technical Efficiency evaluated under CRS is the product of Technical Efficiency evaluated under VRS and Scale Efficiency (see Färe, Grosskopf, and Lovell (1985).

Briefly discussing the computational methods for obtaining the radial input efficiency measure (3) for each evaluated observation relative to all technologies in (2), the convex case just requires solving a nonlinear programming problem (NLP): this is evidently simplified to the familiar linear programming (LP) problem found in the literature (see Hackman (2008) or Ray (2004)) by substituting $w_k = \alpha z_k$. For nonconvex technologies, nonlinear mixed integer programs must be solved in (2): however, Podinovski (2004), Leleu (2006) and Briec, Kerstens, and Vanden Eeckaut (2004) propose mixed integer programs, LP problems, and closed form solutions derived from an implicit enumeration strategy, respectively. Kerstens and Van de Woestyne (2014) review all methods in this nonconvex case in more detail and empirically document that implicit enumeration is by far the fastest solution strategy. Daraio, Kerstens, Nepomuceno, and Sickles (2019) provide a review of software options (with the main focus on the convex methods).

Proposition 2.1. *Following Briec, Kerstens, and Vanden Eeckaut (2004, Lemma 3), it is straightforward to establish the following relations between convex and nonconvex input-oriented efficiency components:*

- $TE_i^{C}(\boldsymbol{x},\boldsymbol{y}) \leq TE_i^{NC}(\boldsymbol{x},\boldsymbol{y})$;
- $OTE_i^{C}(\boldsymbol{x},\boldsymbol{y}) \leq OTE_i^{NC}(\boldsymbol{x},\boldsymbol{y})$;
- $SCE_i^{C}(\boldsymbol{x},\boldsymbol{y}) \gtreqless SCE_i^{NC}(\boldsymbol{x},\boldsymbol{y})$.

To clarify the relationship between convex and nonconvex decompositions (4), we start from the observation that nonconvex technologies are nested in the convex counterparts. As a consequence, nonconvex $OTE_i^{\Lambda}(\boldsymbol{x},\boldsymbol{y})$ and $TE_i^{\Lambda}(\boldsymbol{x},\boldsymbol{y})$ components are larger or equal than their convex counterparts. However, there is no a priori ordering between nonconvex and convex $SCE_i^{\Lambda}(\boldsymbol{x},\boldsymbol{y})$ components: while the underlying efficiency measures can be ordered, it is impossible to order the ratios between these efficiency measures.

In the literature, several methods are available to obtain qualitative information characterising returns to scale (see Seiford and Zhu (1999) for a review). Since none of these existing methods are suitable for nonconvex technologies, Kerstens and Vanden Eeckaut (1999, Proposition 2) generalize the existing goodness-of-fit method proposed by Färe, Grosskopf, and Lovell (1983) in a convex setting such that it becomes perfectly general. Obviously, this qualitative information holds for efficient points only: these are either efficient observations, or projection points in case of initially inefficient observations. Formally, it is possible to infer for any single observation whether it satisfies globally constant (CRS), increasing (IRS), or decreasing (DRS) returns to scale by simply identifying the technology yielding the maximal input efficiency score.

Proposition 2.2. *Using $E_i^{\Lambda,\Gamma}(\boldsymbol{x},\boldsymbol{y}:\boldsymbol{S}^{\Lambda,\Gamma})$ and conditional on an efficient point, technology $\boldsymbol{S}^{\Lambda,VRS}$ is characterized by:*

(a) CRS $\Leftrightarrow E_i^{\Lambda,NIRS}(\boldsymbol{x},\boldsymbol{y}:\boldsymbol{S}^{\Lambda,NIRS}) = E_i^{\Lambda,NDRS}(\boldsymbol{x},\boldsymbol{y}:\boldsymbol{S}^{\Lambda,NDRS})$;

(b) IRS $\Leftrightarrow E_i^{\Lambda,NIRS}(\boldsymbol{x},\boldsymbol{y}:\boldsymbol{S}^{\Lambda,NIRS}) < E_i^{\Lambda,NDRS}(\boldsymbol{x},\boldsymbol{y}:\boldsymbol{S}^{\Lambda,NDRS})$;

(c) DRS $\Leftrightarrow E_i^{\Lambda,NIRS}(\boldsymbol{x},\boldsymbol{y}:\boldsymbol{S}^{\Lambda,NIRS}) > E_i^{\Lambda,NDRS}(\boldsymbol{x},\boldsymbol{y}:\boldsymbol{S}^{\Lambda,NDRS})$.

Note that all three input efficiency measures coincide for observations subject to constant returns to scale. The maximal input efficiency measure simply reflects the best fit of a specific technology for the given observation and therefore serves to indicate the most appropriate returns to scale assumption. In fact, it is applicable to any specification of technology and it is simply more general.[1]

2.3 Input-Oriented Malmquist Productivity Index

In a discrete time framework, the input-oriented radial efficiency measure $E_i^t(\boldsymbol{x}^t,\boldsymbol{y}^t)$ indicates the minimum contraction of an input by a scalar θ while still remaining on the boundary of the technology in time period t:

$$E_i^t(\boldsymbol{x}^t,\boldsymbol{y}^t) = \min\left\{\theta \mid (\theta\, \boldsymbol{x}^t,\boldsymbol{y}^t) \in \boldsymbol{S}^t, \theta \geq 0\right\}. \tag{5}$$

[1] One can also distinguish a fourth case of sub-constant returns to scale that is only relevant for nonconvex technologies: see Cesaroni, Kerstens, and Van de Woestyne (2017) for more details and a first empirical exploration. More recently, Mostafaee and Soleimani-Damaneh (2020) propose an even more refined classification of returns to scale for nonconvex technologies.

To simplify the notation, we suppress the superscripts Λ, Γ to indicate convexity or not, and returns to scale on the efficiency measure $E_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t)$ and the technology $\boldsymbol{S}^t$. Drawing on Färe, Grosskopf, Norris, and Zhang (1994), the radial efficiency measure $E_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t)$ is defined as the inverse of the corresponding Shephardian distance function. Hence, for $(a, b) \in \{t, t+1\}$, the time-related versions of the radial input efficiency measure are given as follows:

$$E_i^a(\boldsymbol{x}^b, \boldsymbol{y}^b) = \min\left\{\theta \mid (\theta\, \boldsymbol{x}^b, \boldsymbol{y}^b) \in \boldsymbol{S}^a\right\} \tag{6}$$

if there is some θ such that $(\theta \boldsymbol{x}^b, \boldsymbol{y}^b) \in \boldsymbol{S}^a$ and $E_i^a(\boldsymbol{x}^b, \boldsymbol{y}^b) = +\infty$ otherwise.

Thus, we can use the radial input measure to define the input-oriented Malmquist productivity index for base period t as follows:

$$M_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}) = E_i^t(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})/E_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t), \tag{7}$$

where $E_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t)$ and $E_i^t(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})$ are input efficiency relating observations in period t and $t+1$, respectively, to a period t technology. When the value of the input-oriented Malmquist productivity index for this base period t is above (below) unity, then it reveals an increase (decrease) in productivity.

In the similar way, an input-oriented Malmquist productivity index with base period $t+1$ is also defined as:

$$M_i^{t+1}(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}) = E_i^{t+1}(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})/E_i^{t+1}(\boldsymbol{x}^t, \boldsymbol{y}^t). \tag{8}$$

In the same way, when the value of the Malmquist productivity index for this base period $t+1$ is above (below) unity, then it reveals an increase (decrease) in productivity.

Moreover, to avoid an arbitrary choice of base period, Färe, Grosskopf, Norris, and Zhang (1994) propose defining the input-oriented Malmquist productivity index as a geometric mean of a period t and a period $t+1$ index:

$$\begin{aligned} M_i^{t,t+1}(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}) &= \sqrt{M_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}) \cdot M_i^{t+1}(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})} \\ &= \sqrt{\frac{E_i^t(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})}{E_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t)} \cdot \frac{E_i^{t+1}(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})}{E_i^{t+1}(\boldsymbol{x}^t, \boldsymbol{y}^t)}} \end{aligned} \tag{9}$$

Once again, when the geometric mean input-oriented Malmquist productivity index is greater (less) than 1, then it points to an increase (decrease) in productivity.

The base period of this Malmquist productivity index changes over time. It can be conceptualized as an index computed in a two-year window sliding over the observations in time. Moreover, the Malmquist index (9) can be decomposed into two mutually exclusive components:

$$M_i^{t,t+1}(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}) = \frac{E_i^{t+1}(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})}{E_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t)} \cdot \sqrt{\frac{E_i^t(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})}{E_i^{t+1}(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})} \cdot \frac{E_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t)}{E_i^{t+1}(\boldsymbol{x}^t, \boldsymbol{y}^t)}} \tag{10}$$

The first component measures the change in technical efficiency over time, while the second component is related to the frontier shifts of the production technology (i.e., it captures technological change). If $M_i^{t,t+1}(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})$ is larger (smaller) than unity, then this indicates an improvement (deterioration) in productivity. A similar interpretation applies to the separate components.

Remark that the above definitions deviate from the original ones in Caves, Christensen, and Diewert (1982) in that the ratios have been inverted. This ensures that productivity indices above (below) unity reveal productivity growth (decline), which is in line with traditional productivity indices.

3 Data Sample: Unbalanced Panel of Swedish Courts

The sample is an unbalanced panel of 18 years (2000-2017) of Swedish district courts based on annual statistics adopted from three existing studies (in particular, Mattsson, Månsson, Andersson, and Bonander (2018), Mattsson and Tidanå (2019), Agrell, Mattsson, and Månsson (2020)).[2] In these articles, there are four inputs, including three labor inputs and one capital input, and three outputs as a production specification. More specifically, among the three labor inputs, there are judges, law clerks, and administrative employees (other personnel) measured in full-time equivalents. In addition, the court area is adopted as a proxy variable for capital, under the assumption that the size of the premises is proportional to other capital variables (for example, the number of computers and other equipment, as well as the operational expenditures such as heating, maintenance, and insurance). Moreover, these articles state that the incorporation of capital is important because, to some extent, it is possible to substitute capital for labor in the production of court decisions. The three outputs are decided criminal cases, decided civil cases, and decided petitionary matters. Agrell, Mattsson, and Månsson (2020, p. 662) discuss how these three output categories result from an

[2]We are grateful to Pontus Mattsson for making these data available for our research contribution.

aggregation procedure using self-reported time consumption starting from fourteen output categories. Bogetoft and Wittrup (2021) recently investigate the whole issue of case weighting to assess the workload in a court system. In this contribution, we use the same four inputs and three outputs as used in the three existing studies to perform our own analysis. For more institutional details on the Swedish court system and the role of district courts, the reader is referred to these three existing studies.

Table 1: Descriptive Statistics over the Years 2000-2017

Years	Outputs			Inputs			
	Civil cases	Criminal cases	Matters	Judges	Law clerks	Other personnel	Court area
2000	539	674.7	309.4	6.88	5.28	13.77	2545
	(997)	(938)	(554)	(12.54)	(7.28)	(22.85)	(3760)
2001	537	684.4	303.3	6.68	5.81	12.89	2406
	(1024)	(1001)	(577)	(12.51)	(8.78)	(24.24)	(3867)
2002	651	860.3	377.1	8.10	6.81	14.96	2790
	(1096)	(1275)	(681)	(12.44)	(9.04)	(25.72)	(4262)
2003	710	966.6	430.4	8.64	7.43	16.61	2976
	(1150)	(1329)	(782)	(13.57)	(9.43)	(27.52)	(4313)
2004	710	966.0	418.3	8.32	6.77	15.56	2882
	(1158)	(1342)	(712)	(12.83)	(8.57)	(26.61)	(4275)
2005	734	1031	440.5	8.21	6.96	16.04	2893
	(1153)	(1405)	(798)	(12.77)	(8.95)	(27.21)	(4548)
2006	901	1320	507.7	10.60	10.15	20.50	3587
	(1215)	(1525)	(767)	(14.36)	(12.52)	(30.04)	(4848)
2007	902	1369	489.0	10.41	10.42	21.64	3429
	(1032)	(1240)	(599)	(12.04)	(10.76)	(26.21)	(3796)
2008	1077	1621	508.3	11.41	11.51	23.51	3681
	(1209)	(1327)	(531)	(10.89)	(10.98)	(25.13)	(3641)
2009	1169	1665	542.4	11.41	10.84	23.38	3703
	(1307)	(1461)	(547)	(10.66)	(10.78)	(25.09)	(4198)
2010	1411	1951	645.9	12.86	13.63	25.97	4268
	(1469)	(1551)	(611)	(11.25)	(13.14)	(24.55)	(4176)
2011	1402	2045	677.5	13.67	15.05	26.47	4381
	(1457)	(1606)	(670)	(12.19)	(14.34)	(24.35)	(4183)
2012	1458	2111	499.2	14.42	15.47	26.42	4371
	(1485)	(1811)	(473)	(12.29)	(14.95)	(24.68)	(4196)
2013	1501	2051	513.7	14.77	15.78	27.34	4465
	(1531)	(1823)	(519)	(13.09)	(15)	(25.69)	(4218)
2014	1519	2035	512.2	15.52	16.06	27.73	4461
	(1531)	(1812)	(507)	(13.77)	(15.17)	(25.79)	(4180)
2015	1428	2043	507.9	15.76	15.81	28.02	4511
	(1513)	(1807)	(513)	(13.74)	(15.26)	(26.83)	(4175)
2016	1343	2067	481.9	16.01	15.86	27.58	4554
	(1375)	(1869)	(483)	(14.54)	(15.51)	(26.42)	(4086)
2017	1396	2144	465	15.56	15.32	27.71	4509
	(1414)	(1903)	(423)	(13.17)	(14.39)	(25.48)	(3703)
Number of changes	55.71	55.59	55.71	52.47	52.65	53.94	14.29
	(12.39)	(12.47)	(12.39)	(11.11)	(10.80)	(11.44)	(6.95)

Standard deviation is displayed in parentheses.

The descriptive statistics of the average level and standard deviations of outputs and inputs over the 18 years are reported in Table 1. As can be seen, the differences in outputs and inputs over time are, on average, quite large. More specifically, each of the outputs and the inputs increases in size over time. For example, the number of civil cases goes on average up from 539 to 1396 (2.59 times). Moreover, the number of full-time equivalent judges expands from about 6.88 to about 15.56 on average (2.26 times). Also note that the standard deviations of civil cases, criminal cases, and law clerks are almost as large as their means. However, the standard deviations of the output matters, as well as the inputs judges, other personnel, and court area remain rather stable with very little variation.

Moreover, to determine whether there are fixed inputs that do not change, we exclude the initial post-merger observations (that automatically imply a change in inputs and outputs) and count the number of changes among the observations for each input and each output over all years: we report the average number and standard deviation of changes for all inputs and outputs over all years in the last two lines of Table 1. Among the inputs judges, law clerks, other personnel and court area, there is a change of 52.47, 52.65, 53.94 and 14.29 observations on average. Thus, all inputs seem to change and thus can be treated as variable inputs.[3] More details are provided in Table A.1 in Appendix A.

Table 2 reports the structure of the unbalanced panel over the sample period in the first two columns, and it summarizes the number of courts involved in a merger, and the resulting mergers in the third and fourth columns. In this sample, there were initially 95 district courts in 2000. Then, a court reorganization through mergers is implemented with 36 mergers in total occurring between 2000 and 2009 and 83 courts being involved in a merger (see Agrell, Mattsson, and Månsson (2020) for details). Observe that most mergers have taken place in the two years 2001 and 2005 with no less than 42 (=24+18) merged courts resulting in 16 (=9+7) courts. Between 2000 and 2009, the number of district courts decreases from 95 to 48 and it remains the same thereafter until the end of the sample period. In 2017, the original amount of courts (95) has almost been halved (48).

Moreover, while in general a horizontal merger is the takeover of one or more smaller adjacent district courts by a relatively large district court, during this period some of the new courts consist of parts of the original courts rather than just two or more other courts. For instance, as mentioned in Agrell, Mattsson, and Månsson (2020, p. 673), there were five such merger scenarios in 2007: (i) Sollentuna and parts of Södra Roslagen are merged into Attunda; (ii) parts of Handen, Huddinge, and parts of Stockholm are merged into Södertrön; (iii) Nacka, parts of Handen, and parts of Stockholm are merged into a new court in Nacka; (iv) parts of Stockholm and parts of Södra Roslagen are merged into Solna; and (v) Solna and parts of Stockholm are merged into Stockholm.

Mattsson, Månsson, Andersson, and Bonander (2018, p. 110) describe how the Swedish government implemented several reforms for the district courts during the last 20 years, with

[3] Mattsson, Månsson, Andersson, and Bonander (2018, p. 116) mention that inputs are not easily changed in the short term. This may explain why these authors use an output-oriented Malmquist productivity index.

the major objective of increasing efficiency and productivity, while simultaneously maintaining a high degree of law and order. One such reform targeted the size of the district courts, based on the simple presumption that scale advantages exist. There is no knowledge about any study supporting this presumption at the time of implementation of this merger policy.

Table 2: Summary of Mergers of Swedish District Courts

Years	Number of Courts	Number of Merging Courts	Merged Courts
2000	95	4	2
2001	93	24	9
2002	78	9	3
2003	72	0	0
2004	72	8	4
2005	68	18	7
2006	57	4	2
2007	57	7	5
2008	53	0	0
2009	53	9	4
2010-2017	48	0	0
All	1082	83	36

Furthermore, the descriptive statistics of the averages and standard deviations of the inputs and outputs for all the courts, the ones included in a merger and the ones not included in a merger, as well as the pre-merger observations, hypothetical mergers, and the post-merger observations in the merging years are all reported in Table 3. First, based on the average values of the total DMUs for all, merging and non-merging years in the first six rows, it can be seen that the average values of all input and output indicators in the merging years are smaller than those under all years and even much smaller than those under the non-merging years. Thus, the mergers that took place during the merging years have led to an overall scale increase that becomes visible during the non-merging years.

Second, the descriptive statistics of averages and standard deviations of the pre-merger observations, hypothetical mergers, and post-merger observations in the merging years are also reported in the final six rows. We observe that the means of the pre-merger observations are smaller than those of the post-merger observations, and that the means of the hypothetical mergers are even bigger than those of both the pre-merger and post-merger observations. This indicates that the hypothetical mergers resulting from just adding merging observations have in fact been judged as being too big. These hypothetical mergers have never materialised and the real mergers that took place concern scaled down versions of the hypothetical mergers resulting in the post-merger observations. Comparisons of standard deviations also yield the same conclusion. This phenomenon reveals that the Swedish administrators did not just blindly combine pre-merger observations into hypothetical mergers, but that they carefully have tried to trim down the scale of operations below the hypothetical mergers. The whole merger operations thus seems a very careful operation, even though to our knowledge no formal modeling was involved at any time.

Table 3: Descriptive Statistics for All, Merging and Non-merging Years

Sample	Outputs			Inputs			
	Civil cases	Crime cases	Matters	Judges	Law clerks	Other personnel	Court area
All years	994.29	1405.81	459.44	10.89	10.47	20.70	3510.12
(All obs. n=1082)	(1299.41)	(1557.18)	(620.74)	(13.05)	(12.25)	(26.18)	(4177.73)
Merging years	753.37	1048.14	416.30	8.73	7.77	17.15	2998.55
(All obs. n=698)	(1132.78)	(1302.07)	(661.62)	(12.59)	(9.73)	(26.01)	(4144.80)
Non-merging years	1432.22	2055.94	537.86	14.82	15.37	27.15	4440.01
(All obs. n=384)	(1460.26)	(1761.16)	(530.44)	(12.96)	(14.62)	(25.26)	(4080.91)
Pre-merger obs.	460.61	253.14	645.55	5.86	4.83	11.59	2137.84
(Merging years)	(700.28)	(365.16)	(819.48)	(8.25)	(5.24)	(19.37)	(2655.24)
Hypothetical mergers	1305.56	592.21	1793.64	12.91	13.51	25.82	4348.34
(Merging years)	(1310.80)	(475.40)	(1303.59)	(10.12)	(11.23)	(23.30)	(3568.83)
Post-merger obs.	1174.68	545.06	1684.98	12.05	12.39	23.76	4021.58
(Merging years)	(1120.86)	(359.45)	(1140.89)	(8.90)	(9.52)	(19.66)	(2853.24)

4 Empirical Illustration

Our empirical analysis proceeds in three steps. The first step is the calculation of the Malmquist productivity index and its components to obtain information on whether there are substantial changes in performance over time. The second step is the determination of convex and nonconvex technical and scale efficiency scores from the static efficiency decomposition (4). Finally, the third step is the detailed comparison between pre-merger and post-merger observations.

4.1 Malmquist Productivity Index: Any Substantial Progress?

While the main aim of the paper is to focus on the technical and scale efficiency of court mergers, a natural question to ask in the context of a panel data set is to which extent productivity change is substantial for the court sector? If productivity change is important, then it cannot be ignored in the analysis of horizontal mergers. However, if productivity change is close to negligible, then it can eventually be ignored and the technical and scale efficiency analysis can be conducted based on a single intertemporal production frontier that spans all time periods available.

The literature on the productivity of courts is scant. The slightly dated survey of Voigt (2016) on judicial efficiency barely cites a few studies and no firm conclusions are available. Frontier-based productivity studies on courts are very scarce. First, Mattsson, Månsson, Andersson, and Bonander (2018) measure the productivity of a sub-sample of 48 Swedish district courts for the years 2012 to 2015 using an output-oriented Malmquist index and obtain a 1.7% average productivity decline per year. Second, Falavigna, Ippoliti, and Ramello (2018) use the Malmquist productivity index to the Italian first instance tax courts over 3 years (2009-2011) to assess a court reform and find that a reduction in the number of active sections has on average a minor positive impact on productivity. Looking in detail, these authors report a mildly negative impact in some regions and a mildly positive effect on performance in other regions. Third, Giacalone, Nissi, and Cusatelli (2020) adopt the Malmquist index to conduct an empirical analysis over the years 2011-2016 of a reform of Italian first instance courts in 2013: on average there is only a slightly positive effect on productivity although a majority of provinces experiences a positive technological change. Finally, Blank and van Heezik (2020) apply a parametric nonfrontier-based cost function model to time series data from 1980 to 2016 of the Dutch judiciary sector to measure productivity development and obtain a sharp decline in productivity throughout the period despite various policy measures and technological changes.

To conclude, the above productivity measurement review, mainly based on employing the Malmquist productivity index, finds overall minor evidence of productivity growth in courts, even when reforms are consciously aimed at improving performance. Therefore, it is perfectly legitimate in the context of our Swedish district courts to ask the following question: is there any substantial productivity change as measured by our input-oriented Malmquist productivity index (9)?

To allow for a comparison with Mattsson, Månsson, Andersson, and Bonander (2018) who focus on a sub-sample of 48 Swedish district courts for the years 2012 to 2015, we report empirical results for the input-oriented Malmquist productivity index computed relative to a convex CRS technology for the whole period 2000 to 2017 in the upper part of Table 4. There are three slight differences between the Malmquist productivity approach developed in Mattsson, Månsson, Andersson, and Bonander (2018) as presented in the lower part of Table 4 and our approach. First, Mattsson, Månsson, Andersson, and Bonander (2018) focus on the years 2012 to 2015 only, while we analyse the period from 2000 to 2017. Second, our input-oriented Malmquist productivity index has been defined (see supra) to be comparable to the output-oriented Malmquist index of Mattsson, Månsson, Andersson, and Bonander (2018) such that productivity improvements (deteriorations) are indicated by index numbers above (below) unity.[4] Third, Mattsson, Månsson, Andersson, and Bonander (2018) drop some outliers from their 48 Swedish district courts and also report bootstrapped results for their output-oriented Malmquist productivity index.

We report average values per comparison period in Table 4 for the Malmquist productivity index and its components. Furthermore, we also mention the minimum and maximum values per comparison period between brackets. Analysing Table 4 we can make the following observations. First, both productivity growth and decline occur over time, but these are in general rather small. Second, the gist of the results for the common subperiod 2012 to 2015 is similar between our results and the results reported in Mattsson, Månsson, Andersson, and Bonander (2018).

Table 4: Malmquist Productivity Index under a Convex CRS Technology

	Year	Malmquist Productivity Index (MPI)	Technical Change (TC)	Efficiency Change (EC)
Our results	2000-2001	1.042 (0.606-1.721)	1.033 (0.866-1.250)	1.008 (0.594-1.506)
	2001-2002	1.062 (0.558-1.616)	1.126 (0.797-1.856)	0.959 (0.616-1.542)
	2002-2003	1.037 (0.570-1.662)	0.919 (0.660-1.103)	1.137 (0.776-1.940)
	2003-2004	1.065 (0.672-1.270)	1.067 (0.985-1.179)	0.997 (0.682-1.210)
	2004-2005	1.128 (0.692-2.931)	1.210 (0.966-2.196)	0.933 (0.564-1.464)
	2005-2006	0.954 (0.092-2.038)	0.872 (0.104-1.081)	1.095 (0.805-2.039)
	2006-2007	1.009 (0.417-1.490)	0.959 (0.463-1.309)	1.052 (0.785-1.624)
	2007-2008	1.050 (0.752-1.294)	1.048 (0.887-1.142)	1.000 (0.787-1.168)
	2008-2009	1.055 (0.794-1.310)	1.056 (0.877-1.259)	1.000 (0.721-1.169)
	2009-2010	1.031 (0.592-1.489)	1.002 (0.908-1.078)	1.026 (0.613-1.434)
	2010-2011	0.979 (0.801-1.218)	0.986 (0.847-1.218)	0.994 (0.808-1.183)
	2011-2012	0.955 (0.684-1.238)	1.015 (0.684-1.238)	0.943 (0.747-1.128)
	2012-2013	0.977 (0.732-1.173)	0.923 (0.732-1.097)	1.063 (0.847-1.382)
	2013-2014	0.989 (0.815-1.135)	0.963 (0.819-1.071)	1.029 (0.845-1.302)
	2014-2015	0.955 (0.689-1.184)	0.975 (0.902-1.054)	0.978 (0.684-1.127)
	2015-2016	0.984 (0.777-1.318)	0.981 (0.839-1.104)	1.005 (0.840-1.431)
	2016-2017	1.022 (0.801-1.150)	1.021 (0.903-1.121)	1.002 (0.836-1.169)
	2000-2017	1.017 (0.590-1.485)	1.009 (0.699-1.256)	1.013 (0.738-1.404)
Mattsson et al. (2018)	2012-2013	0.980 (0.957-1.001)	0.929 (0.863-0.960)	1.057 (1.009-1.136)
	2013-2014	1.000 (0.983-1.020)	0.987 (0.941-1.025)	1.014 (0.967-1.069)
	2014-2015	0.970 (0.947-0.992)	0.992 (0.961-1.050)	0.978 (0.907-1.016)
	2012-2015	0.983 (0.962-1.004)	0.969 (0.921-1.011)	1.016 (0.960-1.073)
t-test †		1.570**(MPI)	0.499**(TC)	1.076**(EC)
p-value		0.135 (MPI)	0.624 (TC)	0.297 (EC)

† t-test: critical values at 1% level=2.55 (***); 5% level=1.73 (**); 10% level=1.33(*)

Finally, we perform a t-test to evaluate whether the average Malmquist productivity index (MPI) as well as its technical change (TC) and efficiency change (EC) components are significantly different from unity or not and we report the corresponding p-values. If the p-value is greater than 0.05, then it means that we cannot reject the null hypothesis that the population average is equal to unity at the 5% significance level. If the p-value is

[4]Recall that under a standard definition of the input-oriented Malmquist productivity index it is the inverse of the output-oriented Malmquist productivity index under CRS.

less than 0.05, then it means that we reject this same null hypothesis at the 5% significance level and we consider that the average MPI and/or its components differ from unity. From the last line of Table 4, the p-values obtained from the t-test are 0.135 for MPI, 0.624 for TC, and 0.297 for EC, which are all greater than 0.05. This indicates that the average MPI and its components are equal to unity and thus that there is no obvious improvement or deterioration in productivity. Thus, we can safely use an intertemporal frontier to measure the efficiency values under convex and nonconvex estimates and ignore technical change.

4.2 OTE Decomposition under C and NC: A First Analysis

The above-reported t-test of the Malmquist productivity index justifies the use of an intertemporal frontier approach that basically ignores the technical change. Hence, we use a pooled frontier for the whole period as a benchmark when measuring the overall technical efficiency based on CRS, technical efficiency based on VRS, and scale efficiency (SCE) as a ratio of both previous concepts under C and NC technologies. With 1082 observations, this is among the biggest samples analysed in court efficiency studies (see the Voigt (2016) survey).

At the sample level of the Swedish district courts, we first illustrate the differences in the efficiency estimates for OTE, TE and SCE, as well as the returns to scale (RTS) characteristics for convex and nonconvex technologies. The descriptive statistics for these efficiency concepts are shown in Table 5. The first line reports the number of efficient observations. Thereafter, we report the arithmetic average, standard deviation, and minimum and maximum of the efficiency scores. The final line lists the results for the Li-test statistic. Indeed, these efficiency measures are compared by means of a nonparametric test comparing two entire distributions as initially developed by Li (1996) and refined by Fan and Ullah (1999) and most recently by Li, Maasoumi, and Racine (2009). The Li-test statistic tests for the eventual significance of differences between two kernel-based estimates of density functions f and g of a random variable x. The null hypothesis states that both density functions are almost everywhere equal ($H_0 : f(x) = g(x)$ for all x). The alternative hypothesis negates this equality of both density functions ($H_1 : f(x) \neq g(x)$ for some x).[5]

Table 5 reports these descriptive statistics for both the nonconvex and convex efficiency estimates in the columns 3 to 5 and the columns 6 to 8, respectively. The final three columns report the difference in terms of the nonconvex estimates (i.e., $(E_i^{NC} - E_i^{C})/E_i^{NC}$).

[5] Matlab code developed by P.J. Kerstens based on Li, Maasoumi, and Racine (2009) is found at: `https://github.com/kepiej/DEAUtils`.

Table 5: Nonconvex and Convex Efficiency Estimates: Descriptive Statistics

Sample		Nonconvex			Convex			Δw.r.t NC		
		OTE	TE	SCE	OTE	TE	SCE	OTE	TE	SCE
All years (all obs.)	#Eff. Obs.	241	797	241	18	58	18	13.389	13.741	13.389
	Average	0.873	0.977	0.892	0.676	0.758	0.893	0.225	0.225	-0.001
	Stand. Dev	0.127	0.051	0.112	0.128	0.134	0.065	-0.007	-1.619	0.418
	Min	0.310	0.692	0.328	0.291	0.309	0.609	0.046	0.553	-0.857
	Max	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
	Li-test †	157.23*** (OTE)			457.19*** (TE)			113.18*** (SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.0000 (SCE)		
Merging years (all obs.)	#Eff. Obs.	136	508	136	13	35	13	10.462	14.514	10.462
	Average	0.850	0.974	0.871	0.648	0.730	0.888	0.237	0.250	-0.020
	Stand. Dev	0.138	0.056	0.123	0.131	0.137	0.068	0.050	-1.443	0.449
	Min	0.305	0.692	0.328	0.291	0.309	0.609	0.046	0.553	-0.858
	Max	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
	Li-test †	84.522*** (OTE)			296.383*** (TE)			56.331*** (SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.0000(SCE)		
Non-merging years (all obs.)	#Eff. Obs.	105	288	105	5	23	5	21.000	12.522	21.000
	Average	0.914	0.983	0.930	0.726	0.808	0.900	0.206	0.178	0.032
	Stand. Dev	0.090	0.040	0.076	0.105	0.111	0.060	-0.162	-1.798	0.208
	Min	0.535	0.764	0.576	0.409	0.477	0.702	0.235	0.375	-0.220
	Max	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
	Li-test †	77.237*** (OTE)			162.444*** (TE)			57.760*** (SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.0000 (SCE)		
Hypothetical mergers (merging years)	Average	1.006	1.550	0.689	0.814	0.919	0.896	0.191	0.407	-0.300
	Stand. Dev	0.135	0.492	0.155	0.115	0.177	0.086	0.143	0.639	0.448
	Min	0.657	0.901	0.311	0.601	0.628	0.577	0.086	0.304	-0.854
	Max	1.350	3.735	0.991	1.165	1.576	1.000	0.137	0.578	-0.009
	Li-test †	33.909*** (OTE)			47.069*** (TE)			27.734***(SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.0000 (SCE)		

† Li-test: critical values at 1% level=2.33 (***); 5% level=1.64(**); 10% level=1.28(*)

The first horizontal part contains the sample level results that are our focus. The second and third horizontal parts report results for merging and non-merging years. Finally, the fourth horizontal part reports the results for the hypothetical mergers during the merging years. All these results are sequentially commented upon below.

This empirical analysis at the sample level generates the following conclusions. First, among all 1082 observations, the number of efficient observations is 241 under CRS and 797 under VRS under NC, while the number of efficient observations is just 18 under CRS and 58 under VRS under C. Thus, the number of efficient observations is in both cases more than 13 times higher under NC than that under C. Secondly, NC frontier estimates of OTE and TE are on average substantially higher than their C counterparts, while -as expected- the VRS estimates are again higher than the CRS ones. More specifically, the average value of OTE for all observations is 0.873 and for TE it is 0.977 under NC, while the average value of OTE is 0.676 and TE is 0.758 under C, respectively. Looking at the OTE decomposition, it is clear that the major source of inefficiency differs under NC and C. Under NC, TE being close to unity on average, the problem of OTE inefficiency is mainly caused by a low SCE. Under C, the major source of inefficiency is clearly TE, with SCE being less of a problem.

The last two columns also indicate that the C estimates are on average among 22.5% lower in the CRS case and 22.5% lower in the VRS case. Thirdly, the Li-test statistic, which is valid for both dependent and independent variables, has a null hypothesis stating that there exists no difference between the C and NC efficiency distributions for a given return to scale assumption. The bottom line reporting the results of this Li-test statistic confirms that OTE, TE and SCE all differ significantly at the 1% significance level between the NC and C series.

Furthermore, the empirical analysis at the level of merging years and non-merging years generates the following conclusions. First, among the 698 merging year observations and the 384 non-merging year observations, the number of efficient observations is 136 under CRS and 508 under VRS under NC and just 13 under CRS and 35 under VRS under C in the merging years, and 105 under CRS and 288 under VRS under NC and just 5 under CRS and 23 under VRS under C in the non-merging years. Thus, the number of efficient observations is in both cases more than 10 to 14 times higher under NC than that under C in the merging years, and more than 12 to 21 times higher under NC than that under C in the non-merging years. Secondly, NC frontier estimates of OTE and TE are on average higher than their C counterparts and the VRS estimates are again higher than the CRS ones. Comparing the OTE decomposition between merging years and non-merging years, we have the following conclusions. Under NC, TE being nearly efficient, the problem of OTE inefficiency is mainly caused by a relatively low SCE during the merging years which is substantially improved during the non-merging years. Under C, the main source of inefficiency being TE, both TE and SCE improve from the merging years to the non-merging years. Third, the bottom lines reporting the results of the Li-test statistic confirm that OTE, TE and SCE all differ significantly between the NC and C series for both the merging years and non-merging years alike.

Finally, the empirical analysis of the hypothetical mergers during the merging years projected onto the intertemporal frontier composed of all years generates the following results. First, under NC the average values of OTE and TE are 1.006 and 1.550, which are both larger than unity. However, the mean value of SCE is only 0.689, which is smaller than unity. Under C, the mean values of OTE, TE and SCE are 0.814, 0.919 and 0.896, which are all smaller than unity. Thus, the hypothetical mergers are situated in front of the NC CRS and VRS frontiers and therefore generate a technological progress, which is absent under C. Second, the C estimates are on average among 40.7% lower in the VRS case and 19.1% lower in the CRS case. Finally, the Li-test statistic confirms that the OTE, TE and SCE all differ significantly between the NC and C series. Thus, these results confirm that the hypothetical mergers would have generated technological change by shifting the frontier under NC: this would have generated overcapacity and this has led the Swedish administration to downscale the hypothetical mergers towards the current post-merger observations.

A comparison with related literature on courts learns us the following lessons. Castro and Guccio (2014) analyse 27 out of 29 Italian judicial districts in 2006 and find that TE and SCE are on average of equal importance. Castro and Guccio (2018) scrutinise 165 Italian judicial counties for 2011 and find that TE is now the dominant source of poor performance. Peyrache and Zago (2016) use the directional distance function to evaluate the inefficiency and the optimal structure of the Italian court system thereby focusing on the aggregation of results across regional levels. However, this framework is practically incomparable with the static efficiency decomposition.

Turning to the articles on the Swedish district courts, the work by Agrell, Mattsson, and Månsson (2020) adopts three complementary frameworks that allow for no comparison: a global frontier under CRS (results only graphically displayed); a metafrontier approach; and a conditional difference-in-differences analysis. In a similar vein, Mattsson and Tidanå (2019) adopt an analysis based on Bogetoft and Wang (2005): therefore, a comparison is not possible.

Next, we analyse the returns to scale (RTS) characterization of all observations, as well as the observations in the merging and non-merging years. A detailed count of the number of observations for various RTS under C and NC efficiency measures is shown in Table 6. Recall that CRS stands for constant returns to scale, NIRS stands for non-increasing returns to scale (thus, in fact decreasing RTS (DRS)), while NDRS represents non-decreasing returns to scale (thus, in fact increasing RTS (IRS)).

For the total sample, we can infer two conclusions. First, the amount of CRS observations is substantially higher under NC compared to C. Second, under C the overwhelming majority of observations experiences DRS with very few observations undergoing IRS, while under NC a small majority of observations experiences IRS with a slightly smaller amount being DRS. When comparing the merging years and the non-merging years, one can deduce the following conclusions. First, the amount of CRS observations increases due to the merger under NC, while this amount is about stationary under C. Second, the relative number of both IRS and DRS observations decreases in favour of CRS under NC, while under C the amount of IRS observations is reduced to zero while the relative amount of DRS observations increases even further.

This markedly different analysis of RTS under NC and C is not unusual: similar results have earlier been reported in even more details in Cesaroni, Kerstens, and Van de Woestyne (2017). Castro and Guccio (2018) find that the majority of Italian courts are under IRS under one model specification and that the majority of courts are under DRS under another model specification.

Table 6: RTS Classification over All Years, Merging and Non-merging Years

	Sample	#CRS	#NDRS (IRS)	#NIRS (DRS)	Total #observations
Nonconvexity	All obs.	240 (22.181)	434 (40.111)	408 (37.708)	1082
	Merging years	134 (19.198)	289 (41.404)	275 (39.398)	698
	Non-merging years	106 (27.604)	145 (37.760)	133 (34.635)	384
Convexity	All obs.	18 (1.664)	14 (1.294)	1050 (97.043)	1082
	Merging years	13 (1.862)	14 (2.006)	671 (96.132)	698
	Non-merging years	5 (1.302)	0 (0.000)	379 (98.698)	384

4.3 OTE Decomposition under C and NC: Comparing Pre- and Post-Merger Observations

In addition to the empirical analysis at the sample level and at the level of merging years and non-merging years above, thanks to the level playing field created by the hypothesis of no technical change and the resulting intertemporal frontier we can now dig deeper in detail by focusing on the comparison between pre-merger and post-merger observations solely. In this subsection, we conduct a comparative analysis and statistical tests on the efficiency values between the pre-merger and post-merger observations.

Table 7: Pre- and Post-Merger Observations under C and NC: Descriptive Statistics

	Sample	Nonconvexity			Convexity			Δw.r.t NC		
		OTE	TE	SCE	OTE	TE	SCE	OTE	TE	SCE
Pre-merger Obs.	# Eff. obs.	2	21	2	0	0	0	1.000	1.000	1.000
	Average	0.800	0.976	0.820	0.641	0.717	0.896	0.199	0.266	-0.093
	Stand. Dev	0.116	0.039	0.116	0.117	0.124	0.057	-0.008	-2.193	0.505
	Min	0.556	0.868	0.556	0.405	0.442	0.763	0.272	0.491	-0.373
	Max	1.000	1.000	1.000	0.901	0.933	0.992	0.099	0.067	0.008
	Li-test †	7.689***(OTE)			16.598***(TE)			1.888***(SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.0320 (SCE)		
Post-merger obs.	# Eff. Obs.	7	25	7	0	1	0	1.000	0.960	1.000
	Average	0.879	0.974	0.902	0.664	0.766	0.870	0.244	0.213	0.035
	Stand. Dev	0.093	0.046	0.075	0.090	0.108	0.068	0.034	-1.375	0.099
	Min	0.669	0.827	0.716	0.488	0.597	0.725	0.270	0.278	-0.013
	Max	1.000	1.000	1.000	0.834	1.000	0.999	0.166	0.000	0.001
	Li-test †	15.175***(OTE)			15.980***(TE)			0.244***(SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.2830 (SCE)		

† Li-test: critical values at 1% level=2.33 (***); 5% level=1.64(**); 10% level=1.28(*)

Descriptive statistics are reported in Table 7. This empirical analysis allows us to infer the following conclusions. First, the number of efficient observations is zero across the board under C for pre-merger observations, while only a single observation becomes efficient for TE due to the mergers. By contrast, the number of efficient observations is 2 for OTE and SCE, and 21 for TE under NC for pre-merger observations, and this number increases after the mergers to 7 for OTE and SCE, and 25 for TE: the largest relative increase is clearly in OTE and SCE. Second, as expected the NC frontier estimates are on average substantially higher than their C counterparts (about 20% and more) except for the SCE component, while the VRS results are again higher than the CRS ones in the pre-merger case. This result is confirmed in the post-merger case: NC frontier estimates are between 21.3% and 24.4% higher than their C counterparts, and this is now also valid for the SCE component (3.5%). Looking in more detail at the OTE decomposition, we find that under NC the TE component is close to unity and the main source of OTE inefficiency is due to SCE inefficiency, and that the merger improves the OTE efficiency level substantially because the SCE efficiency improves. Under C, the TE inefficiency is worse than the SCE inefficiency, and the merger improves the OTE efficiency level less than in the NC case because the TE efficiency level improves. Third, the bottom line containing the results of the Li-test statistic confirms once more that OTE, TE and SCE differ significantly at the 1% significance level between the NC and C series.

In addition, to further explore whether the performance of the observations involved in the merger has improved after the merger or not, we establish the following definition.

Definition 4.1. When comparing pre-merger and post-merger observations, we define performance as follows:

- If the average efficiency of pre-merger observations is smaller than or equal to the efficiency of post-merger observations, then we consider the performance has been improved.
- If the average efficiency of pre-merger observations is bigger than the efficiency of post-merger observations, then we consider the performance has been deteriorated.

Implementing this Definition 4.1, we simply count the number of different observations complying with this definition to verify if the merging activity improves performance or not. Results are reported in Table 8.

Analysing Table 8 we can infer the following conclusions. First, for the three efficiency results of OTE, TE and SCE the large majority of the 36 observations improve under NC.

Table 8: Number of Observations with Improved or Deteriorated Performance

	Nonconvexity			Convexity		
	OTE	TE	SCE	OTE	TE	SCE
# Observations with improved performance	29	29	29	21	22	13
# Observations with decreased performance	7	7	7	15	14	23

Second, for the same three efficiency results under C only OTE and TE improve in the majority of cases (even though it is less pronounced than under the NC case), while for SCE performance deteriorates for the majority of cases.

Next, our analysis tests for the returns to scale (RTS) characterization of these pre-merger and post-merger courts. A count of the number of observations for various RTS under C and NC efficiency measures is shown in Table 9.

Table 9: RTS Classification between Pre- and Post-Merger Observations

	Sample	#CRS	#NDRS (IRS)	#NIRS (DRS)	Total # observations
Nonconvexity	Pre-merger Observations	13 (15.663)	31 (37.349)	39 (46.988)	83
	Post-merger Observations	6 (16.667)	12 (33.333)	18 (50.000)	36
Convexity	Pre-merger Observations	6 (7.229)	9 (10.843)	68 (81.928)	83
	Post-merger Observations	0 (0.000)	1 (2.778)	35 (97.222)	36

For the numbers of the pre-merger observations under different returns to scale, we can make the following observations. First, among the 83 pre-merger observations 6 observations experience CRS under C and 13 observations under NC. Thus, under NC more observations are able to obtain an optimal size compatible with a long-run zero profit equilibrium. Second, under C only 9 observations experience IRS, while the largest group of observations (68) is characterised by DRS: thus, few observations can potentially benefit from a merger and the largest group of observations is actually already too big. Under NC, 31 observations experience IRS, while a slightly larger group of 39 observations experiences DRS: thus, substantially more observations can potentially benefit from a merger under NC. Third, both C and NC methods agree that the largest group of observations experiences DRS.

Switching to the post-merger observations under different returns to scale, the following conclusions are justified. First, among the 36 post-merger observations, 0 observation experiences CRS under C and 6 observations experience CRS under NC. Again, under NC more observations are able to obtain an optimal size. Second, under C only 1 observation experiences IRS, while the remaining group of 35 observations is characterised by DRS: thus, almost all observations have actually become too big. Under NC, 12 observations experience

IRS, while a slightly larger group of 18 observations experiences DRS: thus, fewer observations have actually become too big. Third, both C and NC methods indicate that by far the largest group of observations experiences DRS.

Hence, under C most pre-merger and almost all post-merger observations are DRS: this indicates a kind of overshooting of the goals of the merger wave. However, under NC the number of CRS, IRS, and DRS cases are more balanced: this would have allowed to better select the IRS observations for the merger, and it signifies there is less overshooting of the goals of the merger wave.

5 Conclusions

Inspired by other contributions utilizing the traditional static input-oriented decomposition of overall technical efficiency to assess the benefits of horizontal mergers, we have applied this rather well-known methodology to a large unbalanced panel of Swedish district courts observed over the years 2000 till 2017. To the best of our knowledge, we are the first study assessing the benefit of horizontal mergers under both convex and nonconvex nonparametric, deterministic frontier specifications. As argued in the introduction, there is a need for conservative estimates of cost savings, since in general these savings are often overcompensated by a market power effect: as shown by Definition 2.1, nonconvex estimates of efficiency gains are more conservative than traditional convex ones. Obviously, in the public sector a market power effect can be safely ignored, but the need for conservative estimates of cost savings remains.

Drawing upon the literature on court productivity and following up on the earlier study of Mattsson, Månsson, Andersson, and Bonander (2018) for a sub-sample of years, we find using an input-oriented Malmquist productivity index that the average productivity change is negligible. This serves to justify the use of an intertemporal or pooled frontier approach over all years that basically ignores any technical change in the sample.

The OTE decomposition under C and NC at the sample level yields the following conclusions. First, there are much more efficient observations under NC compared to C. Second, the major source of OTE inefficiency is SCE under NC and TE under C. Third, according to the Li-test OTE, TE and SCE all differ significantly between the NC and C series. When comparing merging years and non-merging years, about the same conclusions emerge: SCE improves over time under NC, and especially TE improves over time under C.

Turning to the characterization of RTS at the sample level, there are far more CRS observations under NC than under C, and most observations are DRS under C and IRS under NC. Comparing merging years and non-merging years, the amount of CRS observations increases due to the merger under NC, while it is about stationary under C. Furthermore, the relative number of IRS and DRS observations decreases in favour of CRS under NC, while under C the amount of IRS observations becomes null while the relative amount of DRS observations increases further.

Focusing on the analysis of pre- and post-merger observations solely, the following conclusions are supported by the data. First, the number of efficient observations increases under NC, and does only marginally so under C. Second, under NC the OTE decomposition improves because SCE improves, while under C the OTE decomposition improves because TE improves and SCE even slightly deteriorates. Implementing Definition 4.1 confirms improvement across the board under NC, and improvements in OTE and TE jointly with a deterioration of SCE under C. Turning to the characterization of RTS among these pre- and post-merger observations, under C most pre-merger and almost all post-merger observations are DRS, while under NC the number of CRS, IRS, and DRS cases are more balanced.

Therefore, the main contributions in this work can be summarized as follows. First, contrasting the traditional C with the less popular NC methodology, it is fair to state that the former has much more difficulty compared to the latter to make sense of the administrative decision to merge Swedish district courts. Under C, only TE tends to improve and most observations are DRS, while under NC one could have selected among IRS observations for the merger. Under C, there is a kind of overshooting of the traditional goals of the merger wave. Second, these empirical results make the NC methods a worthwhile alternative when one aims at a conservative estimate of the savings associated with horizontal mergers. Third, our results are complementary to the three existing studies analysing this merger wave among Swedish district courts.

References

Agrell, P., P. Mattsson, and J. Månsson (2020): "Impacts on Efficiency of Merging the Swedish District Courts," *Annals of Operations Research*, 288(2), 653–679.

Akhavein, J., A. Berger, and D. Humphrey (1997): "The Effects of Metamergers on Efficiency and Prices: Evidence from a Bank Profit Function," *Review of Industrial Organization*, 12(1), 95–139.

Belleflamme, P., and M. Peitz (2010): *Industrial Organization: Markets and Strategies.* Cambridge University Press, Cambridge.

Blank, J., and A. van Heezik (2020): "Policy Reforms and Productivity Change in the Judiciary System: A Cost Function Approach Applied to Time Series of the Dutch Judiciary System between 1980 and 2016," *International Transactions in Operational Research*, 27(4), 2002–2020.

Bogetoft, P., and D. Wang (2005): "Estimating the Potential Gains from Mergers," *Journal of Productivity Analysis*, 23(2), 145–171.

Bogetoft, P., and J. Wittrup (2021): "Benefit-of-the-Doubt Approach to Workload Indicators: Simplifying the Use of Case Weights in Court Evaluations," *Omega*, 103, 102375.

Briec, W., K. Kerstens, and P. Vanden Eeckaut (2004): "Non-Convex Technologies and Cost Functions: Definitions, Duality and Nonparametric Tests of Convexity," *Journal of Economics*, 81(2), 155–192.

Budzinski, O., and I. Ruhmer (2010): "Merger Simulation in Competition Policy: A Survey," *Journal of Competition Law & Economics*, 6(2), 277–319.

Castro, M., and C. Guccio (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.

Castro, M. F., and C. Guccio (2018): "Measuring Potential Efficiency Gains from Mergers of Italian First Instance Courts through Nonparametric Model," *Public Finance Review*, 46(1), 83–116.

Caves, D., L. Christensen, and W. Diewert (1982): "The Economic Theory of Index Numbers and the Measurement of Inputs, Outputs and Productivity," *Econometrica*, 50(6), 1393–1414.

Caves, R. (2007): "In Praise of the Old I.O.," *International Journal of Industrial Organization*, 25(1), 1–12.

Cesaroni, G., K. Kerstens, and I. Van de Woestyne (2017): "Global and Local Scale Characteristics in Convex and Nonconvex Nonparametric Technologies: A First Empirical Exploration," *European Journal of Operational Research*, 259(2), 576–586.

Cummins, J., S. Tennyson, and M. Weiss (1999): "Consolidation and Efficiency in the US Life Insurance Industry," *Journal of Banking & Finance*, 23(2–4), 325–357.

Daraio, C., K. Kerstens, T. Nepomuceno, and R. Sickles (2019): "Productivity and Efficiency Analysis Software: An Exploratory Bibliographical Survey of the Options," *Journal of Economic Surveys*, 33(1), 85–100.

——— (2020): "Empirical Surveys of Frontier Applications: A Meta-Review," *International Transactions in Operational Research*, 27(2), 709–738.

Deprins, D., L. Simar, and H. Tulkens (1984): "Measuring Labor Efficiency in Post Offices," in *The Performance of Public Enterprises: Concepts and Measurements*, ed. by M. Marchand, P. Pestieau, and H. Tulkens, pp. 243–268. North Holland, Amsterdam.

Falavigna, G., R. Ippoliti, and G. B. Ramello (2018): "DEA-based Malmquist Productivity Indexes for Understanding Courts Reform," *Socio-Economic Planning Sciences*, 62, 31–43.

Fan, Y., and A. Ullah (1999): "On Goodness-of-fit Tests for Weakly Dependent Processes Using Kernel Method," *Journal of Nonparametric Statistics*, 11(1), 337–360.

Färe, R., S. Grosskopf, and C. Lovell (1983): "The Structure of Technical Efficiency," *Scandinavian Journal of Economics*, 85(2), 181–190.

——— (1985): *The Measurement of Efficiency of Production*. Kluwer, Boston.

Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994): "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries," *American Economic Review*, 84(1), 66–83.

Farrell, J., and C. Shapiro (1990): "Horizontal Mergers: An Equilibrium Analysis," *American Economic Review*, 80(1), 107–126.

Farrell, M. (1959): "The Convexity Assumption in the Theory of Competitive Markets," *Journal of Political Economy*, 67(4), 377–391.

Frantz, R. (2015): "Antitrust and X-Efficiency," *Antitrust Bulletin*, 60(3), 221–230.

Gaughan, P. (2007): *Mergers, Acquisitions, and Corporate Restructurings*. John Wiley & Sons, Hoboken, 4th edn.

Giacalone, M., E. Nissi, and C. Cusatelli (2020): "Dynamic Efficiency Evaluation of Italian Judicial System using DEA Based Malmquist Productivity Indexes," *Socio-Economic Planning Sciences*, 72, 100952.

Hackman, S. (2008): *Production Economics: Integrating the Microeconomic and Engineering Perspectives*. Springer, Berlin.

Harris, J., H. Ozgen, and Y. Ozcan (2000): "Do Mergers Enhance the Performance of Hospital Efficiency?," *Journal of the Operational Research Society*, 51(7), 801–811.

Jacobsen, S. (1970): "Production Correspondences," *Econometrica*, 38(5), 754–771.

Kerstens, K., and I. Van de Woestyne (2014): "Solution Methods for Nonconvex Free Disposal Hull Models: A Review and Some Critical Comments," *Asia-Pacific Journal of Operational Research*, 31(1).

——— (2021): "Cost Functions are Nonconvex in the Outputs when the Technology is Nonconvex: Convexification is Not Harmless," *Annals of Operations Research*, p. forthcoming.

Kerstens, K., and P. Vanden Eeckaut (1999): "Estimating Returns to Scale Using Nonparametric Deterministic Technologies: A New Method Based on Goodness-of-Fit," *European Journal of Operational Research*, 113(1), 206–214.

Krishnasamy, G., A. Ridzwa, and V. Perumal (2004): "Malaysian Post Merger Banks' Productivity: Application of Malmquist Productivity Index," *Managerial Finance*, 30(4), 63–74.

Kristensen, T., P. Bogetoft, and K. M. Pedersen (2010): "Potential Gains from Hospital Mergers in Denmark," *Health Care Management Science*, 13(4), 334–345.

Leleu, H. (2006): "A Linear Programming Framework for Free Disposal Hull Technologies and Cost Functions: Primal and Dual Models," *European Journal of Operational Research*, 168(2), 340–344.

Li, Q. (1996): "Nonparametric Testing of Closeness between Two Unknown Distribution Functions," *Econometric Reviews*, 15(1), 261–274.

Li, Q., E. Maasoumi, and J. Racine (2009): "A Nonparametric Test for Equality of Distributions with Mixed Categorical and Continuous Data," *Journal of Econometrics*, 148(2), 186–200.

Mattsson, P., J. Månsson, C. Andersson, and F. Bonander (2018): "A Bootstrapped Malmquist Index Applied to Swedish District Courts," *European Journal of Law and Economics*, 46(1), 109–139.

Mattsson, P., and C. Tidanå (2019): "Potential Efficiency Effects of Merging the Swedish District Courts," *Socio-Economic Planning Sciences*, 67, 58–68.

Monastyrenko, E. (2017): "Eco-efficiency Outcomes of Mergers and Acquisitions in the European Electricity Industry," *Energy Policy*, 107, 258–277.

Mostafaee, A., and M. Soleimani-Damaneh (2020): "Global Sub-Increasing and Global Sub-Decreasing Returns to Scale in Free Disposal Hull Technologies: Definition, Characterization and Calculation," *European Journal of Operational Research*, 280(1), 230–241.

Peyrache, A., and A. Zago (2016): "Large Courts, Small Justice!: The Inefficiency and the Optimal Structure of the Italian Justice Sector," *Omega*, 64, 42–56.

Podinovski, V. (2004): "On the Linearisation of Reference Technologies for Testing Returns to Scale in FDH Models," *European Journal of Operational Research*, 152(3), 800–802.

Ray, S. (2004): *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*. Cambridge University Press, Cambridge.

Scarf, H. (1986): "Testing for Optimality in the Absence of Convexity," in *Social Choice and Public Decision Making: Essays in Honor of Kenneth J. Arrow, Volume I*, ed. by W. Heller, R. Starr, and S. Starrett, pp. 117–134. Cambridge University Press, Cambridge.

——— (1994): "The Allocation of Resources in the Presence of Indivisibilities," *Journal of Economic Perspectives*, 8(4), 111–128.

Seiford, M., and J. Zhu (1999): "An Investigation of Returns to Scale in Data Envelopment Analysis," *Omega*, 27(1), 1–11.

Viscusi, W., J. Harrington, and J. Vernon (2005): *Economics of Regulation and Antitrust*. MIT Press, Cambridge, 4th edn.

Voigt, S. (2016): "Determinants of Judicial Efficiency: A Survey," *European Journal of Law and Economics*, 42(2), 183–208.

Appendix A Additional Descriptive Statistics

Table A.1: Number of times observations change

Years	Outputs			Inputs			
	Civil cases	Criminal cases	Matters	Judge	Law clerks	Other personnel	Court area
2000-2001	91	91	91	89	90	91	35
2001-2002	69	69	69	67	68	68	18
2002-2003	69	69	69	62	55	61	10
2003-2004	72	72	72	55	56	64	12
2004-2005	64	64	64	55	51	57	23
2005-2006	49	49	49	41	46	47	10
2006-2007	53	53	53	51	52	53	13
2007-2008	48	48	48	46	48	48	12
2008-2009	52	52	52	51	52	52	22
2009-2010	44	44	44	43	43	43	16
2010-2011	48	48	48	46	47	47	16
2011-2012	48	48	48	48	48	48	14
2012-2013	48	48	48	48	48	48	11
2013-2014	48	47	48	48	48	48	9
2014-2015	48	48	48	47	48	47	9
2015-2016	48	48	48	48	48	48	6
2016-2017	48	47	48	47	47	47	7
Average	55.71	55.59	55.71	52.47	52.65	53.94	14.29
Standard Deviation	12.39	12.47	12.39	11.11	10.80	11.44	6.95

To ascertain whether or not there are any fixed inputs that do not change, we exclude the initial post-merger observations and count the number of changes among the observations for each input and each output over all years: we report the specific numbers per year as well as the average number of changes for all inputs and outputs over all years in the Table A.1.

Among the inputs judges, law clerks, other personnel and court area, one can observe variations over all years. In particular, in the non-merging years, there are 48 units. And from the Table A.1, we find that almost all the outputs as well as the three labor inputs have changed. Moreover, the input court area also experiences a smaller number of changes. Therefore, all inputs are changing over the years and can be considered as variable inputs.