Annals of Operations Research (2025) 351:1319–1351
https://doi.org/10.1007/s10479-025-06508-9

ORIGINAL RESEARCH

Exploring horizontal mergers in Swedish district courts using technical and scale efficiency: rejecting convexity in favour of nonconvexity

Xiaoqing Chen[1] · Kristiaan Kerstens[2] · Qingyuan Zhu[3]

Received: 11 December 2023 / Accepted: 24 January 2025 / Published online: 14 February 2025
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2025

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 historical merger wave aimed at increasing the size of operations. A battery of recent test statistics rejects convexity in favour of nonconvexity.

Keywords Data envelopment analysis · Free disposal hull · Horizontal mergers · Technical efficiency · Scale efficiency

We are grateful for two most constructive referees. The usual disclaimer applies. This study draws on the European Cooperation in Science and Technology (COST) Action 'Efficient Justice for All: Improving Court Efficiency through EU Benchmarking' (CA20131). We thank Ane Elixabete Ripoll-Zarraga, Shirong Zhao and Jafar Sadeghi for constructive comments on previous versions. X. Chen gratefully acknowledges the financial support of National Natural Science Foundation of China (72301145), the Philosophy and Social Science Foundation of the Jiangsu Higher Education Institutions of China (2023SJYB0177), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (23KJB120008), and the Startup Foundation for Introducing Talent of NUIST (2023r030).

✉ Kristiaan Kerstens
k.kerstens@ieseg.fr

Xiaoqing Chen
XQchen1001@163.com

Qingyuan Zhu
zqyustc@mail.ustc.edu.cn

1 School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China

2 Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Économie Management, F-59000 Lille, France

3 College of Economics and Management, Nanjing University of Aeronautics & Astronautics, Nanjing, China

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 & Peitz, 2010) or Viscusi et al. (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) for an early study and Weinberg (2008) for a survey).

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 mergers, and merger simulations using pre-merger market information to calibrate some noncooperative oligopoly models (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 et al. (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.

In this empirical 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 and specify an intertemporal production frontier whereby we amalgamate all observations over all years confounded. To answer this question, we turn to the recent study of Chen et al. (2024) which studies exactly the same sample using a Malmquist productivity index and a Hicks-Moorsteen total factor productivity index under both variable and constant returns to scale and under convexity and nonconvexity. These authors find that under none of these configurations there is any productivity or total factor productivity that is significantly different from the status quo of no productivity growth. Thus, for this specific sample this lack of productivity growth justifies the specification of an intertemporal production frontier: all observations form a single frontier.

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. Overall, we find that convexity is unable to rationalise the horizontal merger event, while nonconvexity is perfectly capable to make sense of it.

For these purposes, this empirical contribution is structured as follows. Section 2 provides a selective literature review both with regard to the literature on nonparametric frontier methods on mergers and acquisitions and on the need to test for convexity within these methodologies. Section 3 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. After developing this theoretical framework, Sect. 4 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. The historical merger process clearly aimed at increasing the scale of operations. But, the mergers fail to satisfy additivity, a necessary condition for convexity. Section 5 with the empirical illustrations first presents 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. Convexity fails to find improved scale efficiency, while nonconvexity does find scale improvement. 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 6 provides the conclusions.

2 Selective literature review

2.1 Nonparametric frontier methods on mergers and acquisitions

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 et al. (1994)): 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 et al. (2020) for a 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 mention some studies focusing on the public court sector that is the focus of our empirical application later on in this subsection, but only dig deeper into this literature in the main body of the text.

In terms of the efficiency decompositions alluded to above, 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 (see Banker et al. (1984) and Färe et al. (1983) for this decomposition).[1] 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 et al. (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 et al. (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 et al. (2020), Castro and Guccio (2018), Gorman and Ruggiero (2009), 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 the media.

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 et al. (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).

[1] Färe et al. (1983) predates Banker et al. (1984) in decomposing overall technical efficiency into technical efficiency and scale efficiency. Both articles contain a different method to determine returns to scale. Furthermore, Färe et al. (1983) in addition consider a structural or congestion efficiency component. However, a Google Scholar search on 26 November 2023 reveals that in terms of citations the Banker et al. (1984) article is cited 25905 times while the Färe et al. (1983) contribution is only cited 254 times: likely a clear case of the Matthew effect at work.

Other studies assess the effect of M&As on productivity growth. Rezitis (2008) provides a parametric analysis of Malmquist productivity for ten Greek banks over the period 1993 till 2004 and finds that the decrease in productivity for the five post merging banks is due to an increase in technical inefficiency and the disappearance of economies of scale, while technical change is unaffected compared to the pre-merging level. 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., Giacalone et al. (2020); Mattsson et al. (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 (see, e.g., the court survey of Aiello, Bonanno, and Foglia (2024, p. 18)). 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 et al. (2020)). For our large unbalanced panel of Swedish district courts earlier analysed by Agrell et al. (2020), Mattsson et al. (2018) and Mattsson and Tidanå (2019), the approaches are mixed: Agrell et al. (2020) use an input orientation, while Mattsson et al. (2018) and Mattsson and Tidanå (2019) opt for output-oriented efficiency. We empirically demonstrate that these district courts do in fact control their inputs: this need not imply that courts themselves implement input changes, but rather that their administrative authority continuously adds and/or subtracts (re-shuffles) resources. Silva (2018) seems to be the first to assess efficiency by three methods reflecting relationships between inputs and outputs in Portuguese courts: separate assessments, ratios, and differences. In some of these models outputs are generated by output-specific inputs rather than by all inputs jointly.

2.2 Nonparametric frontier methods: the need for testing convexity

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.

Convexity is then maintained in economics and part of operations research because of the assumption of perfect time divisibility (for instance, Shephard (1970, p. 15)), or simply because of analytical convenience. Hence, if time is not perfectly divisible (i.e., positive setup times exist), then nonconvexities may matter.[2] It is often -implicitly or explicitly- assumed

[2] Scarf (1986, p. 120) is very sharp in his critique of convexity: "Unfortunately, convexity of the production set is not a strikingly realistic description of economic reality. Convexity requires that the production possibility set exhibit constant or decreasing returns to scale: That you or I can manufacture automobiles in our own backyards with the same degree of efficiency as that achieved by the Ford Motor Co. Economies of scale

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 et al. (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 if technology is convex, then the cost function is convex in the outputs (e.g., Jacobsen (1970, Proposition 5.2, (Q.9))). Thus, using contraposition, if technology is nonconvex, then the cost function is nonconvex in the outputs. In what can be considered a refined statement, Briec et al. (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.

While it is intuitively clear that convexity can have a potentially drastic impact on technologies, Cesaroni et al. (2017) empirically illustrate that overall technical efficiency, technical efficiency, and scale efficiency are quite different under convexity and nonconvexity, and that returns to scale are also impacted with even contradictory indications as a possibility. Kerstens and Van de Woestyne (2021) systematically review evidence and illustrate the potentially very substantial impact of convexity on cost function estimates, as well as on the determination of scale economies again documenting the possibility of contradictory indications as an empirical regularity. Dang (2022) affirms the findings for the Chilean hydro-electric power generation plants in Atkinson and Dorfman (2009).[3] In particular, using a recent machine learning method to estimate second derivatives, Dang (2022) reports curvature violations rejecting convexity of technology.

Kerstens et al. (2019) empirically compare a large series of technical and economic (i.e., cost-based) capacity notions on both convex and nonconvex technologies. Having defined these capacity notions in detail, an empirical comparison using a secondary data set leads to two key empirical conclusions. First, all these different technical and economic capacity notions follow different distributions. Second, these distributions almost always differ under convex and nonconvex technologies. Finally, Baležentis et al. (2024) employ an additive Luenberger-Hicks-Moorsteen productivity indicator and these authors even report opposite signs between convex and nonconvex productivity measures for a substantial part of the sample in each of the analysed years in their panel data. Thus, it is undeniable that convexity matters both theoretically and from an empirical point of view.

Obviously, such evidence on the impact of convexity on the technology, on the cost function, on technical and economic capacity notions, and on one specific productivity indicator remains just casual evidence at best. The convexity assumption is central to economics and to the DEA literature in operations research and it is clear that it is not gone be abandoned unless the evidence against it is massive and robust. Therefore, in this contribution we focus on yet

based on large indivisible pieces of machinery or forms of productive organization such as the assembly line, which are not economically merited at small scales of operation, are a major ingredient of the industrial revolution of the last 100 years. And their workings cannot be captured, either theoretically or computationally, by the competitive paradigm."

[3] One of the two secondary data sets analysed in Kerstens and Van de Woestyne (2021).

another potential empirical application area of relevance for industrial organization, namely the evaluation of horizontal mergers using familiar technical and scale efficiency concepts. To the best of our knowledge, we are the first study to combine convex and nonconvex methods to study these technical and scale efficiencies for horizontal court mergers. Furthermore, for replication purposes we opt to use data on Swedish district courts that have been analysed before by Mattsson et al. (2018); Mattsson and Tidanå (2019); Agrell et al. (2020).

Anticipating our empirical results, our study of these data reveal that the horizontal mergers have been crafted with the explicit goal to magnify the scale of the court operations. However, this merger process does not respect the simple addition of observations, thus failing a necessary condition for convexity. Then, the fundamental question is whether this scale improvement is picked up by the convex and nonconvex models we are employing. The empirical evidence indicates that nonconvex models find scale efficiency improvements but no technical efficiency changes, while convex models report technical efficiency improvements but no scale efficiency changes. Furthermore, the use of state of the art statistical tests makes us reject convexity in favour of nonconvexity: this combination of tests has to our knowledge never been used in a court setting.

3 Nonparametric technologies, efficiency and statistical tests

This study uses traditional convex and nonconvex nonparametric, deterministic frontier methods to determine the static input-oriented efficiency of each operating unit. In this section we introduce the static efficiency methods.

3.1 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: $\boldsymbol{S} = \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m+n} : \boldsymbol{x}$ 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 inputs and outputs, all four traditional returns to scale hypotheses (i.e., constant, non-increasing, non-decreasing, and variable (flexible) returns to scale), and technologies that satisfy the convexity assumption 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 et al. (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

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

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) or decreasing returns to scale (DRS) and non-decreasing returns to scale (NDRS) or increasing returns to cale (IRS) respectively, fixed at unity under variable returns to scale (VRS), and non-negative under constant returns to scale (CRS).

3.2 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 technology $\boldsymbol{S}$. Obviously, the resulting input combination is located at the boundary of the input correspondence. 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 3.1 Under the assumptions on the technology $\boldsymbol{S}^{\Lambda,\Gamma}$ defined in (2) and following, e.g., Färe et al. (1983), the following input-oriented efficiency notions can be distinguished:

- Technical Efficiency (TE) 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 (OTE) 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 (SCE) 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 $E_i^{\Lambda,\text{CRS}}\left(\boldsymbol{x}, \boldsymbol{y} : \boldsymbol{S}^{\Lambda,\text{CRS}}\right) \leq E_i^{\Lambda,\text{VRS}}\left(\boldsymbol{x}, \boldsymbol{y} : \boldsymbol{S}^{\Lambda,\text{VRS}}\right) \leq 1$, clearly $0 < SCE_i^{\Lambda}(\boldsymbol{x}, \boldsymbol{y}) \leq 1$ (see Banker et al. (1984) or Färe et al. (1983)). Using Definition 3.1, the following identity 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 OTE evaluated under CRS is the product of TE evaluated under VRS and SCE (see Färe et al. (1994)).

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 et al. (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. Furthermore, implicit enumeration makes use of (scaled) vector dominance that compare an inefficient observation with an observation that somehow dominates it: this facilitates learning. Daraio et al. (2019) provide a review of software options (with the main focus on the convex methods).

Proposition 3.1 *(Briec, Kerstens, and Vanden Eeckaut (2004, Lemma 3)) It is straightforward to establish the following relations between convex and nonconvex input-oriented efficiency components:*

- $TE_i^C(\boldsymbol{x}, \boldsymbol{y}) \leq TE_i^{NC}(\boldsymbol{x}, \boldsymbol{y})$;
- $OTE_i^C(\boldsymbol{x}, \boldsymbol{y}) \leq OTE_i^{NC}(\boldsymbol{x}, \boldsymbol{y})$;
- $SCE_i^C(\boldsymbol{x}, \boldsymbol{y}) \overset{\geq}{\underset{<}{=}} SCE_i^{NC}(\boldsymbol{x}, \boldsymbol{y})$.

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 the Seiford and Zhu (1999) review).[4] Since none of these existing methods are suitable for nonconvex technologies, Kerstens and Vanden Eeckaut (1999, Proposition 2) generalize the existing convex goodness-of-fit method of Färe et al. (1983) 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 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 3.2 *(Kerstens and Vanden Eeckaut (1999, Proposition 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 two 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.[5]

---

[4] Quantitative scale information would, e.g., be scale elasticities: these are not trivial to obtain given the lack of continuity in the convex case, and these are impossible to obtain in the nonconvex case in the current state of knowledge.

[5] One can also distinguish a fourth case of sub-constant returns to scale that is only relevant for nonconvex technologies: see Cesaroni et al. (2017) for more details and a first empirical exploration.

3.3 Statistical testing

These efficiency measures in Definition 3.1 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 et al. (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). In order to get around the issue of spurious mass at the boundary, Simar and Zelenyuk (2006) further refine this Li-test statistic for nonparametric frontier estimators. Their Algorithm I ignores the boundary estimates, and their Algorithm II smooths the boundary estimates by adding uniform noise that is one order of magnitude less than the noise added by the specific estimator. According to Monte Carlo evidence, algorithm II appears to perform somewhat better overall, despite the fact that the test statistic's strength diminishes as the production specification includes more dimensions. To put it briefly, we use the Li et al. (2009) version of this test that has been modified using Simar and Zelenyuk (2006) Algorithm II.[6]

In addition to this Li-test, following recent studies (e.g., O'Loughlin and Wilson (2021) and Wilson and Zhao (2023)) we report two further tests developed by Kneip et al. (2016) and Simar and Wilson (2020) to specifically test the convexity of the production possibility set in Swedish district courts against its nonconvex alternative. More specifically, Kneip et al. (2016) utilize the central limit theorems for the means of efficiency estimates to develop statistical tests of various model features (here convexity). These tests of convexity involve comparing the mean of the nonconvex input-oriented efficiency measure $TE_i^{NC}(\boldsymbol{x}, \boldsymbol{y})$ (denoted μ_i^{NC}) and the mean of the convex input-oriented efficiency measure $TE_i^{C}(\boldsymbol{x}, \boldsymbol{y})$ (denoted μ_i^{C}). The null hypothesis states that these two means are equal ($H_0 : \mu_i^{NC} = \mu_i^{C}$). The alternative hypothesis negates this equality of the two means and regards the mean of $TE_i^{NC}(\boldsymbol{x}, \boldsymbol{y})$ to be larger than the mean of $TE_i^{C}(\boldsymbol{x}, \boldsymbol{y})$ ($H_1 : \mu_i^{NC} > \mu_i^{C}$).

These tests require the two sample means under comparison to be independent of each other: this requires randomly splitting the original sample into two independent parts. The resulting test statistic is asymptotically normally distributed, and this test is one-sided. Though this Kneip et al. (2016) test is valid for a single split of the original sample, different results can be obtained from different splits of the sample. Therefore, Simar and Wilson (2020) suggest splitting multiple times and then using bootstrap methods to implement tests based on sample means of the test statistics from each split, or based on the p-values obtained on each split. This results in two tests: one based on the mean of test statistics over multiple sample-splits (denoted KSW-test#1), and the other based on a Kolmogorov-Smirnov test of uniformity of the p-values (denoted KSW-test#2). Moreover, the critical values for these two tests are determined from the bootstrap as the sample splits are not independent. We use 100 multiple sample-splits and 1000 bootstrap replications. The reader is referred to Kneip et al. (2016) and Simar and Wilson (2020) for additional discussion and technical details.[7]

---

6 Matlab code developed by P.J. Kerstens based on Li et al. (2009) and Simar and Zelenyuk (2006) is found at: https://github.com/kepiej/DEAUtils.

7 Computations are done using the `FEAR` package in the software `R`.

4 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 five existing studies (in particular, Mattsson et al. (2018), Mattsson and Tidanå (2019), Agrell et al. (2020), Chen and Kerstens (2023), and Chen et al. (2024)).[8] Chen and Kerstens (2023, Section 2) review in rather detail what is known so far on the gains and efficiency of horizontal mergers in Swedish district courts.

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 et al. (2020, p. 662) discuss how these three output categories result from an 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 five existing studies to perform our own analysis.[9] For more institutional details on the Swedish court system and the role of district courts, the reader is referred to these five existing studies. Voigt (2016) and more recently Aiello et al. (2024) survey the literature on court performance.

The descriptive statistics of the average level and standard deviations of outputs and inputs over the 18 years are reported in Table 1.[10] 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 538.67 to 1395.79 (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

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

[9] One referee has suggested the use of weight restrictions on the inputs and/or the outputs. While convex models with weight restrictions are quite common, this comes close to putting input (output) prices on the inputs (outputs). We see two practical objections to this suggestion. First, to our knowledge no court frontier performance study has so far incorporated weight restrictions (see the surveys of Voigt (2016) or Aiello et al. (2024)): thus, articulating these weight restrictions may not be that trivial in a court context and there is no evidence at all that ignoring weight restrictions would somehow jeopardize our focus on testing convexity of the court technology. Second, apart from the problem of the non-uniqueness of multipliers in a convex technology, the weights in the basic nonconvex model are either zero or infinity and thus cannot be meaningfully compared. Therefore, we are unaware of meaningful nonconvex production models with weight restrictions and this would somehow undermine the purpose of comparing convex and nonconvex approaches. This is certainly a promising avenue for future work.

[10] We impose the following regularity conditions on the data for inputs and outputs (see Färe, Grosskopf, and Lovell (1994, p. 44–45)): (1) each unit uses nonnegative amounts of each input to produce nonnegative amounts of each output; (2) there is an aggregate production of positive amounts of every output, and an aggregate utilisation of positive amounts of every input; and (3) each unit employs a positive amount of at least one input to produce a positive amount of at least one output. Hence, we eliminate in total two units for which all input dimensions are zero.

Table 1 Descriptive statistics over the years 2000–2017

Years	# Courts	Outputs			Inputs			
		Civil cases	Criminal cases	Matters	Judges	Laws	Personnel	Area
2000	95	538.67	674.67	309.40	6.88	5.28	13.77	2545.44
		(996.86)	(937.82)	(554.18)	(12.54)	(7.28)	(22.85)	(3759.61)
2001	95	535.51	685.28	303.39	6.69	5.87	13.02	2424.51
		(1013.39)	(992.05)	(571.52)	(12.38)	(8.70)	(24.00)	(3834.70)
2002	78	655.33	866.38	380.71	8.15	6.91	15.09	2812.35
		(1093.84)	(1272.21)	(679.70)	(12.42)	(9.01)	(25.67)	(4251.33)
2003	72	709.96	966.61	430.37	8.64	7.43	16.61	2976.49
		(1149.60)	(1329.20)	(781.78)	(13.57)	(9.43)	(27.52)	(4312.72)
2004	72	709.58	965.93	418.32	8.32	6.77	15.56	2881.77
		(1158.24)	(1341.78)	(712.02)	(12.83)	(8.57)	(26.61)	(4274.58)
2005	68	748.47	1054.16	449.85	8.39	7.14	16.44	2986.43
		(1149.72)	(1400.59)	(797.01)	(12.74)	(8.93)	(27.17)	(4553.03)
2006	57	900.63	1320.42	507.70	10.60	10.15	20.50	3587.35
		(1214.52)	(1525.47)	(766.84)	(14.36)	(12.52)	(30.04)	(4847.75)
2007	57	901.79	1369.44	488.95	10.41	10.42	21.64	3428.62
		(1031.68)	(1240.41)	(599.25)	(12.04)	(10.76)	(26.21)	(3795.91)
2008	53	1076.62	1621.05	508.33	11.41	11.51	23.51	3681.07
		(1208.89)	(1326.91)	(531.12)	(10.89)	(10.98)	(25.13)	(3641.46)
2009	54	1161.06	1651.82	540.55	11.30	10.78	23.38	3688.02
		(1300.05)	(1459.17)	(544.53)	(10.63)	(10.74)	(24.97)	(4179.92)
2010	48	1410.55	1951.05	645.91	12.86	13.63	25.97	4267.81
		(1468.59)	(1550.58)	(611.20)	(11.25)	(13.14)	(24.55)	(4176.22)

Table 1 continued

Years	# Courts	Outputs			Inputs			
		Civil cases	Criminal cases	Matters	Judges	Laws	Personnel	Area
2011	48	1402.48	2045.46	677.53	13.67	15.05	26.47	4380.81
		(1456.66)	(1606.21)	(669.73)	(12.19)	(14.34)	(24.35)	(4182.80)
2012	48	1458.05	2111.06	499.21	14.42	15.47	26.42	4370.77
		(1484.60)	(1811.46)	(473.05)	(12.29)	(14.95)	(24.68)	(4196.37)
2013	48	1500.93	2051.45	513.67	14.77	15.78	27.34	4464.98
		(1530.99)	(1823.05)	(519.15)	(13.09)	(15.00)	(25.69)	(4218.27)
2014	48	1519.02	2034.52	512.18	15.52	16.06	27.73	4461.00
		(1531.44)	(1811.65)	(507.30)	(13.77)	(15.17)	(25.79)	(4179.80)
2015	48	1428.08	2042.89	507.86	15.76	15.81	28.02	4510.96
		(1513.24)	(1806.69)	(513.45)	(13.74)	(15.26)	(26.83)	(4174.75)
2016	48	1342.87	2067.05	481.88	16.01	15.86	27.58	4554.46
		(1374.82)	(1869.23)	(482.61)	(14.54)	(15.51)	(26.42)	(4086.36)
2017	48	1395.79	2144.09	464.61	15.56	15.32	27.71	4509.27
		(1414.38)	(1903.39)	(423.44)	(13.17)	(14.39)	(25.48)	(3702.75)
# Changes		58.59	58.59	58.59	55.18	55.06	56.94	17.53
		(14.87)	(14.87)	(14.87)	(12.91)	(12.05)	(13.70)	(9.66)
Stockholm vs. other courts before merger in 2007		8454.05	5567.17	9478.61	100.17	74.92	198.66	33520.44
		(835.31)	(737.18)	(1381.73)	(10.09)	(8.45)	(13.50)	(3385.71)
		vs	*vs*	*vs*	*vs*	*vs*	*vs*	*vs*

Table 1 continued

Years	# Courts	Outputs			Inputs			
		Civil cases	Criminal cases	Matters	Judges	Laws	Personnel	Area
		582.55	327.44	827.80	7.02	6.28	13.60	2471.10
		(610.37)	(300.19)	(751.07)	(6.94)	(4.96)	(14.86)	(2116.80)
Stockholm vs. other courts after merger in 2007		6714.53	3042.37	6408.98	57.80	70.78	128.44	2496.2
		(391.11)	(444.14)	(268.98)	(6.09)	(8.41)	(6.67)	(1562.52)
		vs	*vs*	*vs*	*vs*	*vs*	*vs*	*vs*
		1252.78	482.83	1872.16	13.16	13.28	24.21	3850.85
		(1205.06)	(387.93)	(1584.30)	(10.96)	(11.43)	(20.65)	(2800.06)

Average on first line. Standard deviation is displayed in parentheses personnel, and court area remain rather stable with very little variation. The number of courts monotonously decreases over time: at the end we retain about half of the starting number of courts.

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 first two lines of the lower part of Table 1. For example, when the number of judges changes in adjacent years for a particular court, then the number of input changes is recorded as 1. If there are 90 courts with changes in the number of judges in adjacent years, then we count the number of changes as 90, and so on, until the number of changes in all adjacent years is obtained. Then, we compute the arithmetic average and standard deviation to get the result of "# Changes" in the third horizontal part of Table 1. Among the inputs judges, law clerks, other personnel and court area, there is a change of 55.18, 55.06, 56.94 and 17.53 observations on average. Thus, all inputs seem to change and thus can be treated as variable inputs.[11] Remark that this tendency of all inputs changing need not imply that courts themselves implement input changes: it may well be that their administrative authority continuously adds and/or subtracts (re-shuffles) resources. More details on these computations are provided in Table A.1 in Appendix A. We also provide an empirical example regarding the court in Stockholm in Table A.2 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, the resulting mergers, and the newly emerging courts from the third to the fifth columns. In particular, the second column presents the number of courts in each year. The third column shows the number of courts in which mergers are occurring in each year, and the fourth column shows the number of new courts acquired as a result of the mergers. Mergers occur when the location of the merged court coincides with the location of one of its constituent merging courts. Finally, the fifth column indicates the number of newly emerging courts not resulting from the previously described merger operations: here we have courts that are situated in a new location distinct from the constituent parts. For instance, in the year 2000 there are 95 courts in total: four courts are merged, resulting in two new courts. Furthermore, two new courts have been created: ***Blekinge*** and ***Västmanland***. In brief, 95 courts minus 4 merging, plus 2 resulting from the merger operation, and plus 2 new courts yield again 95 courts in 2001. For lack of space, more details on these merger operations on a year by year basis are provided in Appendix C. In particular, a summary of the detailed information on the mergers of Swedish district courts is found in Table C.1 in Appendix C. In addition, there are only 48 courts in each of the years 2010–2017: there are no mergers between courts. All changes that do not produce a change in the number of courts have been grouped into one row for recording purposes. Finally, we sum the number of courts, the number of courts in which mergers occurred, the number of courts after mergers per year, and the number of newly emerging courts, respectively: these results are shown in the last row.

In total we have 1085 observations. There are 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 et al. (2020) for details). In total 7 newly emerging courts appear. Observe that most mergers have taken place in the two years 2001 and 2005 with no less than 42 (=24+18) merged courts resulting

[11] Mattsson et al. (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.

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.[12] For instance, as mentioned in Agrell et al. (2020, p. 673), there were five such merger scenarios in 2007: (1) Sollentuna and parts of Södra Roslagen are merged into Attunda; (2) parts of Handen, Huddinge, and parts of Stockholm are merged into Södertrön; (3) Nacka, parts of Handen, and parts of Stockholm are merged into a new court in Nacka; (4) parts of Stockholm and parts of Södra Roslagen are merged into Solna; and (5) Solna and parts of Stockholm are merged into Stockholm. In addition, given that Stockholm is a very large court that merged in 2007, we have added two rows at the end of Table 1 to capture the differences in means and variances of the inputs and outputs for Stockholm versus the other courts. These results show that the mean and standard deviation of each input and output for the Stockholm court before the merger are greater than the corresponding values for the other courts. Furthermore, the means of three labor inputs and all outputs after the Stockholm court merger remain larger than their counterparts in the other courts, while the mean value of the court area input is smaller than the counterpart in the other courts. Furthermore, for all inputs, civil cases and matters outputs after the Stockholm court merger, the standard deviations are smaller than their counterparts in the other courts, while the standard deviations of criminal cases are larger than their counterparts in the other courts.

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

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 merger observations (based on simple addition of observations), and the post-merger observations in the merging years are all reported in Table 3. In particular, the first two lines present the results for mean and standard deviation of all inputs and outputs for all courts (n=1085) from 2000 to 2017. The third and fourth lines list the results for mean and standard deviation of inputs and outputs for all courts (n=701) during the period in which the merger occurs (2000–2009). The fifth and sixth lines show results for mean and standard deviation of inputs and outputs for all courts (n=384) during the period without merger (2010–2017). Furthermore, the seventh and eighth lines display the results of the means and standard deviations of inputs and outputs for the 83 courts in Table 2 for the year of the merger (here called pre-merger observations). In the ninth and tenth lines, we sum the values of the inputs and outputs of the merged courts to produce a hypothetical court after the merger (n=36), and analyse the mean and standard deviation of the values for these hypothetical courts. Finally, the last two lines report an analysis of the means and standard deviations of the values of each indicator for the 36 courts in Table 2 that appear after the merger.

[12] M&As normally involve the combination of two or more companies or organizations into one. But, in our sample we are actually looking at more complex restructuring activities. To keep the terminology as simple as possible, we keep referring to these restructuring activities as basic M&A activities.

Table 2 Summary of mergers of Swedish district courts

Years	# Courts	# Merging courts	# Merged courts	# Newly emerging courts
2000	95	4	2	2 (***Blekinge*** & ***Västmanland***)
2001	95	24	9	1 (***Ångermanland***)
2002	78	9	3	0
2003	72	0	0	0
2004	72	8	4	1 (***Värmland***)
2005	68	18	7	0
2006	57	4	2	2 (***Attunda*** & ***Södertörn***)
2007	57	7	5	0
2008	53	0	0	1 (***Skaraborg***)
2009	54	9	4	0
2010–2017	48	0	0	0
All	1085	83	36	7

In contrast to Table 1, Table 3 reports only the means and standard deviations of the inputs and outputs for all courts in each year. Moreover, Table 2 simply reports the number of all courts in each year, the number of courts where mergers occurred in each year, and the number of courts after mergers in each year. Hence, in Table 3 we make a detailed distinction between the samples so as to analyse the means and standard deviations of the inputs and outputs for the courts.

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. A numerical example clarifying the concept of hypothetical mergers is developed in Appendix B: this example also serves to illustrate that hypothetical mergers can lead to computational infeasibilities for which in the literature no solution strategy has been proposed.

This phenomenon reveals that the Swedish administrators did not just blindly combine pre-merger observations into hypothetical mergers using addition, 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.

This at least implicit rejection of additivity by the Swedish administrators can be interpreted as casting doubt on convexity of technology. While convexity is sometimes considered a primitive axiom, it can also be derived from other axioms. Arrow and Hahn (1971, p. 59–60) prove that both additivity and divisibility imply convexity and constant returns to scale together. Rejecting additivity then undermines one condition to obtain convexity: technology may then well be nonconvex.[13] Testing the plausibility of convexity is focus in our empirical analysis developed in the next Sect. 5.

5 Empirical illustration

Our empirical analysis proceeds in two steps. The first step is the determination of convex and nonconvex technical and scale efficiency scores from the static efficiency decomposition (4). The second step is the detailed comparison between pre-merger and post-merger observations. In both steps we make use of state of the art statistical tests to assess whether the traditional convexity assumption can be maintained or not.

5.1 *OTE* decomposition under C and NC: a first analysis

In the recent study of Chen et al. (2024) a one-sample symmetric Wilcoxon test and a t-test of the Malmquist and Hicks-Moorsteen productivity indices under various specifications on exactly the same data set indicate that average productivity changes of these two indices are negligible. Since no obvious technical change is being generated, this justifies the use of an intertemporal frontier approach that basically ignores technical change. Hence, we use a pooled frontier for the whole period as a benchmark when measuring the OTE based on CRS, TE based on VRS, and SCE as a ratio of both previous concepts under C and NC technologies. With 1085 observations, this is among the biggest samples analysed in court efficiency studies (see the Voigt (2016) survey and Peyrache and Zago (2016) use 990 observations).

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 4. 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 lines list the results for two state of the art test statistics: on the one hand the Li-test, and on the other hand $KSW-test\#1$ and $KSW-test\#2$. Table 4 reports these descriptive statistics for both the nonconvex and convex efficiency estimates in the columns 3–5 and the columns 6–8, respectively. The final three columns report the difference in terms of the nonconvex estimates (e.g., $TE_i^{\Delta} = (TE_i^{NC} - TE_i^{C}))/TE_i^{NC})$. 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.

[13] If production is divisible, but not additive, then the technology may well be nonconvex. Suppose y^1 and y^2 are two activities, then by divisibility λy^1 and $(1-\lambda)y^2$ are activities for $0 \leq \lambda \leq 1$. But, $\lambda y^1 + (1-\lambda)y^2$ may not be an activity by the lack of additivity. Recall that production is additive whenever λy^1 and $(1-\lambda)y^2$ are activities, then $\lambda y^1 + (1-\lambda)y^2$ is an activity. Hence, technology may well be nonconvex by the lack of additivity.

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	994.63	1406.94	460.26	10.90	10.49	20.74	3518.22
(n=1085)	(1298.11)	(1555.33)	(620.03)	(13.03)	(12.23)	(26.15)	(4173.64)
Merging years	754.57	1050.92	417.69	8.75	7.81	17.22	3012.55
(n=701)	(1131.15)	(1300.07)	(660.56)	(12.56)	(9.71)	(25.97)	(4140.14)
Non-merging years	1432.22	2055.94	537.86	14.82	15.37	27.15	4440.01
(n=384)	(1460.26)	(1761.16)	(530.44)	(12.96)	(14.62)	(25.26)	(4080.91)
Pre-merger Obs	809.29	1080.58	477.44	10.06	8.97	19.82	3252.79
(n=83)	(1587.10)	(1903.17)	(1035.97)	(19.33)	(17.30)	(41.30)	(6174.47)
Hypothetical Obs	1582.00	2125.01	933.21	19.75	17.84	38.99	6654.03
(n=36)	(2260.99)	(2703.56)	(1442.36)	(27.29)	(25.07)	(57.25)	(8425.49)
Post-merger Obs	1341.98	1890.71	669.19	14.50	14.74	28.11	4964.80
(n=36)	(1412.14)	(1462.69)	(623.87)	(11.71)	(12.60)	(27.03)	(4109.15)

Average on first line. Standard deviation is displayed in parentheses

Table 4 Nonconvex and convex efficiency estimates: descriptive statistics

Sample		Nonconvex			Convex			Δw.r.t NC		
		OTE_i^{NC}	TE_i^{NC}	SCE_i^{NC}	OTE_i^{C}	TE_i^{C}	SCE_i^{C}	OTE_i^{Δ}	TE_i^{Δ}	SCE_i^{Δ}
All years (n=1085)	#Eff. Obs	247	722	247	14	53	7	17.64	13.62	35.29
	Average	0.872	0.976	0.892	0.622	0.749	0.827	0.287	0.232	0.073
	Stand. Dev	0.131	0.056	0.114	0.151	0.142	0.105	−0.154	−1.552	0.082
	Min	0.161	0.487	0.225	0.108	0.225	0217	0.327	0.539	0.038
	Max	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
	Simar & Zelenyuk Li-test†	179.74*** (OTE)			459.17*** (TE)			113.89*** (SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.0000 (SCE)		
	KSW-test#1 (p-value)				2.3446 (0.0010)					
	KSW-test#2 (p-value)				0.7173 (0.0410)					
Merging years (n=701)	#Eff. Obs	141	460	141	9	30	5	15.67	15.33	28.20
	Average	0.849	0.972	0.871	0.586	0.718	0.813	0.309	0.261	0.067
	Stand. Dev	0.143	0.063	0.126	0.151	0.147	0.110	−0.055	−1.346	0.124
	Min	0.161	0.487	0.225	0.108	0.225	0.217	0.327	0.539	0.038
	Max	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
	Simar & Zelenyuk Li-test†	104.70*** (OTE)			298.40*** (TE)			55.49*** (SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.0000(SCE)		
	KSW-test#1 (p-value)				1.6267 (0.0150)					
	KSW-test#2 (p-value)				0.8361 (0.0040)					

Table 4 continued

Sample		Nonconvex			Convex			Δw.r.t NC		
		OTE_i^{NC}	TE_i^{NC}	SCE_i^{NC}	OTE_i^{C}	TE_i^{C}	SCE_i^{C}	OTE_i^{Δ}	TE_i^{Δ}	SCE_i^{Δ}
Non-merging years (n=384)	#Eff. Obs	106	262	106	5	23	2	21.20	11.39	53.00
	Average	0.914	0.983	0.930	0.687	0.806	0.852	0.248	0.180	0.084
	Stand. Dev	0.090	0.040	0.076	0.127	0.114	0.089	−0.405	−1.867	−0.174
	Min	0.535	0.764	0.575	0.316	0.458	0.498	0.409	0.401	0.134
	Max	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
	Simar & Zelenyuk Li-test†	81.69*** (OTE)			162.94*** (TE)			56.87*** (SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.0000 (SCE)		
	KSW-test#1 (p-value)				2.3140 (0.0040)					
	KSW-test#2 (p-value)				0.9311 (0.0010)					
Hypothetical Obs. (n=36)	Average	0.798	1.072	0.750	0.537	0.716	0.759	0.327	0.332	−0.012
	Stand. Dev	0.162	0.199	0.123	0.107	0.149	0.125	0.337	0.251	−0.015
	Min	0.300	0.683	0.393	0.199	0.417	0.406	0.336	0.390	−0.034
	Max	1.046	1.687	0.920	0.733	0.986	0.999	0.300	0.416	−0.085
	Simar & Zelenyuk Li-test†	23.42*** (OTE)			22.57*** (TE)			22.16***(SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.0000 (SCE)		
	KSW-test#1 (p-value)				1.6182 (0.0140)					
	KSW-test#2 (p-value)				0.7819 (0.0020)					

† Li-test: critical values at 1% level=2.33 (***); 5% level=1.64(**); 10% level=1.28(*)
KSW-test#1 is based on averaging the Kneip et al. (2016) test statistic over multiple sample-splits
KSW-test#2 is based on a Kolmogorov-Smirnov test: see Simar and Wilson (2020)

This empirical analysis at the sample level generates the following conclusions. First, among all 1085 observations, the number of efficient observations is 247 under CRS and 722 under VRS under NC, while the number of efficient observations is just 14 under CRS and 53 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 0.872 and for TE it is 0.976 under NC, while the average value of OTE is 0.622 and TE is 0.749 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 28.7% lower in the CRS case and 23.2% lower in the VRS case.

Thirdly, the Li-test statistic has a null hypothesis stating that there exists no difference between the C and NC efficiency distributions for a given returns 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. Finally, *p*-values of the two convexity tests KSW-test#1 and KSW-test#2 are both smaller than 0.05, implying that the production possibility set is not convex for all observations.[14]

Furthermore, the above mentioned lack of average productivity change also makes it possible to empirically analyze between merging and non-merging years: this comparison generates the following conclusions.[15] First, among the 701 merging year observations and the 384 non-merging year observations, the number of efficient observations is 141 under CRS and 460 under VRS under NC and just 9 under CRS and 30 under VRS under C in the merging years, and 106 under CRS and 262 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 15 times higher under NC than that under C in the merging years, and more

[14] We ignore the serial correlation that may be present in the pooled data over time: this may potentially affect our convexity test results in a variety of ways.

[15] A referee points out the potential bias when comparing merging and non-merging years in the case of input and output growth given the many non-merging years at the end of the studied period. The growth of inputs and outputs observed in Table 1 is indeed likely due to: on the one hand, the addition of resources and an increasing demand for justice, and on the other hand, the merger operations may somehow have mitigated or magnified this growth phenomenon. A previous study of Chen et al. (2024) on exactly the same data has found no evidence at all of any productivity change using a Malmquist productivity index as well as a Hicks-Moorsteen total factor productivity index under various conditions. Thus, the observed growth in inputs and outputs seems to imply that on average courts grow bigger and change position with respect to an in essence almost stationary technology that experiences no significant average shifts of the technological frontier itself. Our empirical analysis employs technical efficiency, overall technical efficiency, and scale efficiency. The observed growth in inputs and outputs with regard to an almost stationary technology may or may not have an effect when we measure technical efficiency with regard to a VRS technology. Similarly, it may or may not have an effect when we measure overall technical efficiency with regard to a constant returns to scale technology. Finally, the observed growth in inputs and outputs with regard to an almost stationary technology may or may not have an effect when we measure scale efficiency, since the latter is simply the ratio of the above two models. Thus, under these circumstances merging and non-merging years can be meaningfully compared since no productivity growth is at stake during this period. There may potentially be a bias when we measure technical efficiency with regard to a variable returns to scale technology, with respect to a constant returns to scale technology, or when measuring scale efficiency as a ratio of the two above efficiencies. In short, we do not expect any systematic bias in our analysis due to the general tendency of growth in our sample. However, in our interpretation, the fear of this reviewer for a potential bias when a general tendency of growth is observed is not borne out in the empirical analysis.

than 11 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 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, *p*-values of the two convexity tests KSW-test#1 and KSW-test#2 or both merging and non-merging years are also smaller than 0.05. These results suggest that the production possibility set is unlikely convex for merging and non-merging years.

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 TE is 1.072, which is larger than unity. However, the mean value of OTE and SCE are only 0.798 and 0.750, which are both smaller than unity. Under C, the mean values of OTE, TE and SCE are 0.537, 0.716 and 0.759, which are all smaller than unity. Thus, the hypothetical mergers are situated in front of the NC VRS frontier and therefore generate a technological progress, which is absent under C. Second, the C estimates are on average among 33.2% lower in the VRS case and 32.7% lower in the CRS case. Third, 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. Finally, *p*-values of the two convexity tests KSW-test#1 and KSW-test#2 are again smaller than 0.05: we can also reject convexity in favour of nonconvexity for the hypothetical mergers.

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. Gorman and Ruggiero (2009) analyse US prosecutor offices and find that the average SCE of 0.88 is larger than the average TE of 0.68. Thus, TE is clearly the main source of under achievement. 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 et al. (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 5.

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 significantly due to the merger under NC, while this amount increases slightly 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 one 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 detail in Cesaroni et al. (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. Gorman and Ruggiero (2009) discover that the majority of prosecution offices have mainly decreasing returns to scale in their sample.

5.2 *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.

Descriptive statistics are reported in Table 6. 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 22 for TE under NC for pre-merger observations, and this number increases after the mergers to 7 for OTE and SCE, and 22 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 21% 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 21.5% and 29.6% higher than their C counterparts, and this is now also valid for the SCE component (9.9%). 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. Last but not least, for the pre-merger observations, p-values of two convexity tests KSW-test#1 and KSW-test#2 both indicate that we can safely reject convexity. For the post-merger observations, we safely reject convexity under KSW-test#2 and only marginally reject it under KSW-test#1.

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.

Table 5 RTS classification over all years, merging and non-merging years

	Sample	#CRS	#NDRS (IRS)	#NIRS (DRS)	Total #observations
Nonconvexity	All obs	249(22.95%)	447(41.20%)	389(35.85%)	1085
	Merging years	141(20.11%)	302(43.08%)	258(36.80%)	701
	Non-merging years	108(28.13%)	145(37.76%)	131(34.11%)	384
Convexity	All obs	16(1.47%)	4(0.37%)	1065(98.16%)	1085
	Merging years	9(1.28%)	4(0.57%)	688(98.15%)	701
	Non-merging years	7(1.82%)	0(0.00%)	377(98.18%)	384

Table 6 Pre- and post-merger observations under C and NC: descriptive statistics

	Sample	Nonconvexity			Convexity			Δw.r.t NC		
		OTE_i^{NC}	TE_i^{NC}	SCE_i^{NC}	OTE_i^{C}	TE_i^{C}	SCE_i^{C}	OTE_i^{Δ}	TE_i^{Δ}	SCE_i^{Δ}
Pre-merger Obs.	# Eff. Obs	2	22	2	0	0	0	1.000	1.000	1.000
	Average	0.800	0.978	0.816	0.571	0.685	0.836	0.283	0.300	−0.024
	Stand. Dev	0.126	0.042	0.121	0.129	0.135	0.093	−0.026	−2.230	0.227
	Min	0.510	0.833	0.518	0.317	0.405	0.606	0.379	0.514	−0.170
	Max	1.000	1.000	1.000	0.867	0.933	0.997	0.133	0.067	0.003
	Simar & Zelenyuk Li-test†	8.768***(OTE)			14.487***(TE)			1.823**(SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.0310 (SCE)		
	KSW-test#1 (p-value)				0.7933 (0.0180)					
	KSW-test#2 (p-value)				0.3395 (0.0180)					
Post-merger Obs.	# Eff. Obs	7	22	7	0	1	0	1.000	0.962	1.000
	Average	0.883	0.975	0.904	0.622	0.765	0.815	0.296	0.215	0.099
	Stand. Dev	0.091	0.046	0.073	0.119	0.108	0.119	−0.302	−1.371	−0.619
	Min	0.668	0.827	0.714	0.382	0.593	0.560	0.429	0.283	0.216
	Max	1.000	1.000	1.000	0.888	1.000	0.999	0.112	0.000	0.001
	Simar & Zelenyuk Li-test†	16.525***(OTE)			15.212***(TE)			2.085***(SCE)		
	p-value	0.0000 (OTE)			0.0000 (TE)			0.0164 (SCE)		
	KSW-test#1 (p-value)				1.1127 (0.0780)					
	KSW-test#2 (p-value)				0.5000 (0.0110)					

† Li-test: critical values at 1% level=2.33 (***); 5% level=1.64(**); 10% level=1.28(*)
KSW-test#1 is based on averaging the Kneip et al. (2016) test statistic over multiple sample-splits
KSW-test#2 is based on a Kolmogorov-Smirnov test: see Simar and Wilson (2020)

Table 7 Number of observations with improved or deteriorated performance

	Nonconvexity			Convexity		
	OTE_i^{NC}	TE_i^{NC}	SCE_i^{NC}	OTE_i^{C}	TE_i^{C}	SCE_i^{C}
# Obs. with improved performance (without case weights)	29	29	29	22	25	18
# Obs. with decreased performance (without case weights)	7	7	7	14	11	18
# Obs. with improved performance (with case weights)	26	8	31	23	26	16
# Obs. with decreased performance (with case weights)	10	28	5	13	10	20

Definition 5.1 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 5.1, 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 7.

Analysing Table 7 we can infer the following conclusions. On the one hand, if we perform the analysis without case weighted efficiency measures, then the findings are obtained as follows. First, for the three efficiency results of OTE, TE and SCE the large majority of the 36 observations improve under NC. 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 and improves for half of the cases. On the other hand, if we conduct the analysis with case weighted efficiency measures, then the findings are generated as follows. First, for the two efficiency results of OTE and SCE the large majority of the 36 observations improve, while the efficiency result of TE in the large majority of the 36 observations decrease under NC. Second, for the same three efficiency results under C only OTE and TE improve in the majority of cases, while for SCE performance deteriorates and improves in nearly half of the cases, which is similar to the result without considering case weights.

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 8.

For the numbers of the pre-merger observations under different returns to scale, we can make the following conclusions. First, among the 83 pre-merger observations 5 observations experience CRS under C and 14 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 6 observations experience IRS, while the largest group of observations (72)

Table 8 RTS classification between pre- and post-merger observations

	Sample	#CRS	#NDRS (IRS)	#NIRS (DRS)	Total # observations
Nonconvexity	Pre-merger Obs	14(16.87%)	35(42.17%)	34(40.96%)	83
	Post-merger Obs	6(16.67%)	11(30.56%)	19(52.78%)	36
Convexity	Pre-merger Obs	5(6.02%)	6(7.23%)	72(86.75%)	83
	Post-merger Obs	0(0.000%)	0(0.000%)	36(100.0%)	36

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, 35 observations experience IRS, while a slightly larger group of 34 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 0 observation experiences IRS, while the remaining group of 36 observations is characterised by DRS: thus, all observations have actually become too big. Under NC, 11 observations experience IRS, while a slightly larger group of 19 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.

One remark on the particular case of the Stockholm court. While it is decreasing returns to scale before the merger under both convexity and nonconvexity, it also remains decreasing returns to scale after the merger in both cases.

6 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 well-known methodology to a large unbalanced panel of Swedish district courts observed over the years 2000–2017. 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 3.1, 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.

The recent, most comprehensive court productivity study of Chen et al. (2024) known to us finds no productivity growth in Swedish district courts. In particular, using an input-oriented Malmquist productivity index and a Hicks-Moorsteen total factor productivity index Chen et al. (2024) find that there is no significant productivity growth at all. This finding is robust to variations in specifications both under variable and constant returns to scale, and under convexity and nonconvexity. This serves to justify the use of an intertemporal or pooled frontier approach over all years that basically ignores any technical change in our 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. Our results are consistent with, e.g., Gorman and Ruggiero (2009) and Frantz (2015) in that mergers improve efficiency mainly via SCE, and that there is no evidence of a facilitating role of TE in this merger process.

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 obviously due to the merger under NC, while it increases slightly 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 is only one 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 5.1 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 all post-merger observations are DRS, while under NC the number of CRS, IRS, and DRS cases are more balanced.

Therefore, our main contributions 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. Furthermore, we have in all these steps conducted a battery of modern statistical tests: the Li-test and the KSW-test#1 and KSW-test#2 of convexity. The relevant results indicate the massive rejection of convexity of technology (a marginal rejection occurs in just one instance). Third, our results are complementary to the three existing studies analysing this merger wave among Swedish district courts. In short, it is clear that the traditional convexity assumption cannot make sense of this historical horizontal merger wave of Swedish district courts, while its nonconvex counterpart can make sense of it.

Several questions for future research arise. First, it is obviously highly desirable to study horizontal M&As in other countries and sectors using the basic efficiency decomposition of overall technical efficiency, technical efficiency and scale efficiency as we have done to see whether convexity has a similar impact. Second, since Saastamoinen et al. (2017) illustrate that merger gains can depend on whether the production frontier is estimated in a deterministic (e.g., DEA) or stochastic (e.g., StoNED) way, it may be important to duplicate our research efforts with also a stochastic frontier estimation method (e.g., StoNED may be a good candidate). Third, another aspect of this robustness question is related to the orientation of measurement. Since some articles opt for it, it is useful to also use an output-oriented measurement to see whether the basic efficiency decomposition results reported here remain true. Fourth, as suggested by a referee, one could compare the hypothetical merger observations with the post-merger observations and check whether convex or nonconvex efficiency estimates provide a better prediction. Some recent work suggest that nonconvex estimators provide a better job at predicting (see, e.g., Jin et al. (2024) or Delnava et al. (2024)). Fifth, we are unaware of meaningful nonconvex production models with weight restrictions in the current state of the literature and a comparison of convex and nonconvex approaches with weight restrictions is certainly a promising avenue for research.

However, it is important that for replication purposes and for robustness sake more horizontal mergers are investigated and wider topics in economics and in the DEA literature in operations research are being investigated for the impact of the convexity axiom. Only if massive and robust evidence can be assembled as to the theoretical redundancy of convexity and its empirical impact on the most important research questions in economics and operations research alike, then the research community may consider it time to ditch convexity. Furthermore, it should be remembered that nonconvex production frontier methods are always based on some scaled vector dominance that allows to compare some inefficient observation with an observation that actually dominates it (see Kerstens and Van de Woestyne (2014)): this makes validation easy and greatly facilitates learning.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10479-025-06508-9.

Declarations

Conflict of interest No potential Conflict of interest to disclose.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Agrell, P., Mattsson, P., & Månsson, J. (2020). Impacts on efficiency of merging the Swedish district courts. *Annals of Operations Research, 288*(2), 653–679.
Aiello, F., Bonanno, G., & Foglia, F. (2024). A meta-regression analysis on judicial efficiency literature: The role of methodological and courts diversity. *Journal of Applied Economics, 27*(1), 2284010.
Akhavein, J., Berger, A., & Humphrey, D. (1997). The effects of metamergers on efficiency and prices: Evidence from a bank profit function. *Review of Industrial Organization, 12*(1), 95–139.
Arrow, K., & Hahn, F. (1971). *General Competitive Analysis*. North Holland Publishing Company.
Atkinson, S., & Dorfman, J. (2009). Feasible estimation of firm-specific allocative inefficiency through Bayesian numerical methods. *Journal of Applied Econometrics, 24*(4), 675–697.
Baležentis, T., Kerstens, K., & Shen, Z. (2024). Economic and environmental decomposition of Luenberger-Hicks-Moorsteen total factor productivity indicator: Empirical analysis of Chinese textile firms with a focus on reporting infeasibilities and questioning convexity. *IEEE Transactions on Engineering Management, 71*, 2772–2785.
Banker, R., Charnes, A., & Cooper, W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*(9), 1078–1092.

Belleflamme, P., & Peitz, M. (2010). *Industrial Organization: Markets and Strategies*. Cambridge University Press.
Bogetoft, P., & Wang, D. (2005). Estimating the potential gains from mergers. *Journal of Productivity Analysis, 23*(2), 145–171.
Bogetoft, P., & Wittrup, J. (2021). Benefit-of-the-doubt approach to workload indicators: Simplifying the use of case weights in court evaluations. *Omega, 103*, 102375.
Briec, W., Kerstens, K., & Vanden Eeckaut, P. (2004). Non-convex technologies and cost functions: Definitions, duality and nonparametric tests of convexity. *Journal of Economics, 81*(2), 155–192.
Budzinski, O., & Ruhmer, I. (2010). Merger simulation in competition policy: A survey. *Journal of Competition Law & Economics, 6*(2), 277–319.
Castro, M., & Guccio, C. (2014). Searching for the source of technical inefficiency in Italian judicial districts: An empirical investigation. *European Journal of Law and Economics, 38*(3), 369–391.
Castro, M. F., & Guccio, C. (2018). Measuring potential efficiency gains from mergers of Italian first instance courts through nonparametric model. *Public Finance Review, 46*(1), 83–116.
Caves, R. (2007). In praise of the old I.O. *International Journal of Industrial Organization, 25*(1), 1–12.
Cesaroni, G., Kerstens, K., & Van de Woestyne, I. (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.
Chen, X., & Kerstens, K. (2023). Evaluating horizontal mergers in Swedish district courts using plant capacity concepts: With a focus on nonconvexity. *RAIRO - Operations Research, 57*(1), 219–236.
Chen, X., Kerstens, K., & Tsionas, M. (2024). Does productivity change at all in Swedish district courts? Empirical analysis focusing on horizontal mergers. *Socio-Economic Planning Sciences, 91*, 101787.
Cummins, J., Tennyson, S., & Weiss, M. (1999). Consolidation and efficiency in the US life insurance industry. *Journal of Banking & Finance, 23*(2–4), 325–357.
Dang, J. (2022): "Machine Learning Estimation of Nonparametric Econometric Models and Marginal Effects," Ph.D. thesis, University of California, Riverside.
Daraio, C., Kerstens, K., Nepomuceno, T., & Sickles, R. (2019). Productivity and efficiency analysis software: An exploratory bibliographical survey of the options. *Journal of Economic Surveys, 33*(1), 85–100.
Daraio, C., Kerstens, K., Nepomuceno, T., & Sickles, R. (2020). Empirical surveys of frontier applications: A meta-review. *International Transactions in Operational Research, 27*(2), 709–738.
Delnava, H., Kerstens, K., Kuosmanen, T., & Shen, Z. (2024): "Semi-parametric Estimation of Convex and Nonconvex By-Production Technologies," Working Paper 2024-EQM-02, IESEG School of Management, Lille.
Deprins, D., Simar, L., & Tulkens, H. (1984). Measuring labor efficiency in post offices. In M. Marchand, P. Pestieau, & H. Tulkens (Eds.), *The performance of public enterprises: Concepts and measurements* (pp. 243–268). North Holland.
Fan, Y., & Ullah, A. (1999). On goodness-of-fit tests for weakly dependent processes using Kernel method. *Journal of Nonparametric Statistics, 11*(1), 337–360.
Färe, R., Grosskopf, S., & Lovell, C. (1983). The structure of technical efficiency. *Scandinavian Journal of Economics, 85*(2), 181–190.
Färe, R., Grosskopf, S., & Lovell, C. (1994). *Production Frontiers*. Cambridge: Cambridge University Press.
Farrell, J., & Shapiro, C. (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, 4th edn.
Giacalone, M., Nissi, E., & Cusatelli, C. (2020). Dynamic efficiency evaluation of Italian judicial system using DEA based Malmquist productivity indexes. *Socio-Economic Planning Sciences, 72*, 100952.
Gorman, M., & Ruggiero, J. (2009). Evaluating US judicial district prosecutor performance using DEA: Are disadvantaged counties more inefficient? *European Journal of Law and Economics, 27*(3), 275–283.
Hackman, S. (2008). *Production Economics: Integrating the Microeconomic and Engineering Perspectives*. Berlin: Springer.
Harris, J., Ozgen, H., & Ozcan, Y. (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.
Jin, Q., Kerstens, K., & Van de Woestyne, I. (2024). Convex and nonconvex nonparametric frontier-based classification methods for anomaly detection. *OR Spectrum, 46*(4), 1213–1239.

Kerstens, K., Sadeghi, J., & Van de Woestyne, I. (2019). Convex and nonconvex input-oriented technical and economic capacity measures: An empirical comparison. *European Journal of Operational Research, 276*(2), 699–709.

Kerstens, K., & Van de Woestyne, I. (2014). Solution methods for nonconvex free disposal hull models: A review and some critical comments. *Asia-Pacific Journal of Operational Research, 31*(1), 1450010.

Kerstens, K., & Van de Woestyne, I. (2021). Cost functions are nonconvex in the outputs when the technology is nonconvex: Convexification is not harmless. *Annals of Operations Research, 305*(1–2), 81–106.

Kerstens, K., & Vanden Eeckaut, P. (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.

Kneip, A., Simar, L., & Wilson, P. (2016). Testing hypotheses in nonparametric models of production. *Journal of Business & Economic Statistics, 34*(3), 435–456.

Kristensen, T., Bogetoft, P., & Pedersen, K. M. (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., Maasoumi, E., & Racine, J. (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. *Journal of Econometrics, 148*(2), 186–200.

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

Mattsson, P., & Tidanå, C. (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.

O'Loughlin, C., & Wilson, P. (2021). Benchmarking the performance of US municipalities. *Empirical Economics, 60*(6), 2665–2700.

Peyrache, A., & Zago, A. (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.

Rezitis, A. (2008). Efficiency and productivity effects of bank mergers: Evidence from the Greek banking industry. *Economic Modelling, 25*(2), 236–254.

Saastamoinen, A., Bjørndal, E., & Bjørndal, M. (2017). Specification of merger gains in the Norwegian electricity distribution industry. *Energy Policy, 102*, 96–107.

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.

Scarf, H. (1994). The allocation of resources in the presence of indivisibilities. *Journal of Economic Perspectives, 8*(4), 111–128.

Seiford, M., & Zhu, J. (1999). An investigation of returns to scale in data envelopment analysis. *Omega, 27*(1), 1–11.

Shephard, R. (1970). *Theory of Cost and Production Functions*. Princeton University Press.

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

Simar, L., & Wilson, P. W. (2020). Hypothesis testing in nonparametric models of production using multiple sample splits. *Journal of Productivity Analysis, 53*(3), 287–303.

Simar, L., & Zelenyuk, V. (2006). On testing equality of distributions of technical efficiency scores. *Econometric Reviews, 25*(4), 497–522.

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

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

Weinberg, M. (2008). The price effects of horizontal mergers. *Journal of Competition Law & Economics, 4*(2), 433–447.

Wilson, P. W., & Zhao, S. (2023). Investigating the performance of Chinese banks over 2007–2014. *Annals of Operations Research, 321*(1–2), 663–692.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.