June 2023

WORKING PAPER SERIES

2023-EQM-03

Does Productivity Change at All in Swedish District Courts? Empirical Analysis Focusing on Horizontal Mergers

Xiaoqing Chen
School of Management Science and Engineering, Nanjing University of Information Science and Techno- logy, Nanjing, Jiangsu, China, and IESEG School of Management, 3 rue de la Digue, F-59000 Lille, France, XQchen1001@163.com.

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

Mike Tsionas
Montpellier Business School, France & Lancaster University Management School, UK
m.tsionas@montpellier-bs.com.

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

Does Productivity Change at All in Swedish District Courts? Empirical Analysis Focusing on Horizontal Mergers*

Xiaoqing Chen† Kristiaan Kerstens‡ Mike Tsionas§

28th November 2022
This version: 28th June 2023

Abstract

This contribution is the first to compare the Malmquist and Hicks-Moorsteen productivity indices in the context of horizontal mergers of Swedish district courts during the period 2000-2017. It is also the first to calculate these productivity indices for convex and nonconvex nonparametric frontier specifications in courts under both constant and variable returns to scale. Moreover, a one-sample symmetric Wilcoxon test and a t-test are performed on the average productivity index to determine whether it is significantly different from unity. Also Li-test statistics examine the differences in productivity between these two indices or between convexity and nonconvexity for a given index. Furthermore, we compare these two productivity indices before and after the mergers to investigate the impact of the horizontal merger activity. The empirical results indicate that overall there is no significant technical change at all. Furthermore, horizontal mergers overall do neither result in technical change, nor in post-merger productivity gains.

*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 three referees and an editor for their most helpful comments. The usual disclaimer applies.

†School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China, and IESEG School of Management, 3 rue de la Digue, F-59000 Lille, France, XQchen1001@163.com.

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

§Montpellier Business School, France & Lancaster University Management School, UK m.tsionas@montpellier-bs.com.

Keywords: Horizontal mergers; Malmquist productivity index; Hick-Moorsteen productivity index; Convex and Nonconvex Nonparametric Technologies

Declarations of interest: none

1 Introduction

Mergers and acquisitions (M&As) are a prominent strategic alternative for organizational growth and expansion. Horizontal mergers and acquisitions (HM&As) occur between organizations in the same market and are currently a main pillar of competition regulations worldwide (Gaughan [1]). From a competition perspective (e.g., Capron [2]), since HM&As can allow the consolidation of resources necessary to improve the financial support of their merged entities, HM&As are considered a more direct measure to achieve corporate competitiveness. From a regulatory standpoint (see Belleflamme and Peitz [3] or Viscusi et al. [4]), since HM&As reduce the number of competitors, they increase the probability of generating market power: this implies a loss of social welfare. While HM&As can expand the market power of companies by integrating industries, resulting in a dead weight loss, HM&As can also achieve social welfare gains by cost reductions. The two main purposes of HM&As are thus to achieve economies of scale (advantages of production in higher volumes) and economies of scope (efficiencies or cost savings resulting from producing various goods or services together rather than separately) on the one hand, and to increase the degree of concentration in the industry on the other hand (e.g., De Loecker et al. [5] and the ensuing book of Eeckhout [6]).

The cost savings effect is widely known to be frequently overturned by the market power effect (e.g., Farrell and Shapiro [7]). Since greater cost savings contribute to HM&As being approved by regulators, HM&As participants have a clear incentive to overstate the cost savings potential. In addition, HM&As can also be an effective way to create overcapacity in operating entities. In the case of overcapacity, some entities may have an inefficient product mix or can be located within the production frontier (thus being technically inefficient).

For public sector organisations HM&As mainly aim to achieve economies of scale. Often concentration in the public sector is high by definition, especially if the state is the sole supplier. But, most public goods are produced and distributed for free, for prices below cost, or by implicit prices (queuing).

Effects of HM&As are empirically estimated adopting various approaches. In the industrial organization research, the main approach is the event study methodology, which is usually based on changes in stock market prices of public companies 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 [3, Section 15.4] for a broad overview or Budzinski and Ruhmer [8] for a survey on merger simulations). Moreover, this literature recognises that technical and cost inefficiencies contribute to cost savings of HM&As (see, e.g., Caves [9] or Viscusi et al. [4, p. 88-89] for a general argument and Akhavein et al. [10] for an empirical study).

In fact, our empirical evaluation tool for the impact of HM&As is based on applied production analysis. In particular, deterministic nonparametric frontier models (sometimes called Data Envelopment Analysis (DEA)) are adopted to provide an internal approximation of the production possibility set boundaries, subject to a set of minimal axioms about what is feasible (see Ray [11]). According to Ray [11], efficiency measures are employed to locate observations with respect to the boundary of this deterministic nonparametric production frontier: either the observation is part of the boundary and technically efficient, or it is situated below the boundary and within the technology and technically inefficient.

In this deterministic nonparametric production frontier literature, a large stream of research examines and measures the ex ante potential gains, the ex post merger efficiency, and productivity changes of HM&As. For this contribution, we give a selective review of this literature while focusing primarily on our own methodological choices, and distinguishing between non-court and court mergers. While quite some research exists on the efficiency of various types of courts as summarized in the somewhat dated survey by Voigt [12], productivity studies and the evaluation of HM&As are entirely absent in this survey.

Research using nonparametric frontier methods in non-court areas includes the following examples. For the ex ante gains in non-court areas, Bogetoft and Wang [13] propose a nonparametric frontier method to evaluate the ex ante potential gains of HM&As from the Danish agricultural extension agency, and decompose the ex ante gains into technical efficiency, scale, and harmony (mix) gains, indicating that HM&As indeed bring significant expected gains. Kristensen et al. [14] measure potential gains from planned mergers and perform a decomposition to Danish hospitals: they find that many hospitals are technically inefficient and experience decreasing returns to scale.

Turning to the courts sector, Mattsson and Tidanå [15] utilize this above frontier decomposition method to identify the potential ex ante merger gains of HM&As for Swedish district courts, showing that some mergers have no potential for efficiency gains, while others can yield significant merger gains.

In a similar vein, for the ex post efficiency in the non-court areas, Cummins et al. [16] employ a nonparametric frontier approach to calculate technical and scale efficiencies, and determine returns to scale of HM&As in the US life insurance industry. Their key results suggest that merged firms obtain greater efficiency gains and firms with increasing returns to scale are more probable acquisition targets. Harris et al. [17] adopt an intertemporal production frontier to evaluate the technical and scale efficiency of horizontal mergers in US hospitals, and demonstrate that mergers do increase efficiency, and that scale efficiency is a major contributor to this improved efficiency.

Similar studies of technical and scale efficiencies on courts can also be found. Gorman and Ruggiero [18] adopt the nonparametric frontier technique to perform the ex post technical efficiency evaluation of the US judicial district prosecutors, suggesting that technical inefficiency is the primary source of underachievement. Castro and Guccio [19] employ nonparametric frontier methods to measure the technical and scale efficiency in Italian judicial districts in 2006 and show that technical efficiency and scale efficiency are about equally important. Finally, Castro and Guccio [20] analyse 165 Italian judicial counties in 2011 and find that the courts' reorganization via horizontal mergers and some abolitions leads to an improved technical efficiency.

Other studies measure the effects of HM&As on productivity growth. For the non-court areas, for the US life insurance industry Cummins et al. [16] compute a Malmquist productivity index (MPI) to measure productivity change in a nonparametric way: they show that target firms experience significantly larger gains in technical efficiency change and in productivity over the sample period than do the non-M&A firms. Rezitis [21] adopts the MPI to examine the impact on productivity growth of HM&As in Greek banks: he shows that the impact of HM&As on productivity growth of Greek banks is rather negative. Monastyrenko [22] employs a nonparametric frontier method to compute a slight variation to the MPI to measure the eco-efficiency among European electricity producers during the period 2005–2013: he finds that the systematically regulated domestic horizontal M&As have no impact. Finally, Arocena et al. [23] propose a merger consistent decomposition of the MPI to analyze the effects on productivity change of Japanese water supply systems: these authors find that horizontal mergers contribute positively to productivity change.

Analogous productivity studies focusing on courts can be discussed. Falavigna et al. [24] use the MPI to examine the productivity growth in Italian first instance tax courts from 2009 to 2011, discovering that lowering the number of active departments has a minor beneficial influence on average productivity. Giacalone et al. [25] adopt the MPI to empirically examine Italian first instance courts during the years 2011-2016. Their results suggest that on average, there is only a little beneficial influence on productivity, despite the majority of provinces seeing a favorable technical shift. Mattsson et al. [26] measure productivity growth of HM&As in Swedish district courts calculating the MPI between 2012 and 2015, showing an average annual productivity decline of 1.7%. Finally, Blank and van Heezik [27] apply a nonfrontier-based parametric cost function model to time series data from 1980 to 2016 of the Dutch judiciary sector to measure productivity development and obtain a sharp decline in productivity throughout the period despite various policy measures and technological changes. Thus, overall there is mixed evidence of either slow productivity growth or decline in the judicial sector.

Turning to a discussion of the methodologies adopted in this contribution, Caves et al. [28] propose the MPI using distance functions and apply it for productivity growth analysis. This MPI is probably the best known and most extensively used productivity measure, especially when nonparametric specifications are applied to micro data. While the MPI has many advantages in measuring productivity growth, it also implies some drawbacks. First, some of the distance functions that constitute the MPI using general technologies may be infeasible (see Briec and Kerstens [29] for a demonstration). Second, O'Donnell [30] argues that input- or output-oriented Malmquist index is not multiplicatively complete and usually does not accurately measure changes in total factor productivity (TFP).

In the latter respect, Bjurek [31] proposes a Hicks-Moorsteen productivity index (hereafter HMPI) to address the above major issues with the MPI, which can further decompose the multiplicatively complete TFP indices into technical change and different measures of efficiency change. The HMPI is defined as a ratio of an aggregate output-quantity index evaluated in the output direction over an aggregate input-quantity index evaluated in the input direction. Therefore, this HMPI is well-defined under the general assumptions of variable returns to scale and strong disposability (Briec and Kerstens [32]). However, despite its attractive properties, the HMPI has been less empirically applied. Moreover, despite a growing literature associated with the application of the conventional MPI on mergers, we are unaware of the use of the HMPI in a merger context.

Farrell [33] points out that the convexity assumption maintained in almost all production models precludes the various reasons that may generate nonconvexities in technology: examples include the indivisibility of inputs and outputs, economies of scale and increasing returns to scale, economies of specialization, and negative and positive externalities (such as network externalities and non-rival inputs), etc. (see also Scarf [34]). Hence, a basic nonconvex (NC) deterministic nonparametric production frontier imposing variable returns to scale has been originally developed by Deprins et al. [35]. However, to the best of our knowledge, there are only a very small number of studies that have applied such NC frontier to measure productivity changes. For instance, Kerstens and Van de Woestyne [36] use the MPI and HMPI under convex and nonconvex technologies to analyze the degree of variation in the calculation of productivity indices. As another example Baležentis et al. [37] employ an additive version of the HMPI index (known as a Luenberger-Hicks-Moorsteen indicator) and report opposite signs between convex (C) and NC productivity for substantial parts of the sample.

While it is intuitively clear that convexity can have an effect on technologies, Cesaroni et al. [38] document that certain efficiency concepts (in particular, overall technical efficiency, technical efficiency, and scale efficiency) can be quite different under convexity and nonconvexity, and that the determination of returns to scale for individual observations is also affected. Seminal contributions to axiomatic production theory already reveal that the cost function is convex in the outputs if and only if technology is convex (e.g., Jacobsen [39, Corollary 5.5]). Kerstens and Van de Woestyne [40] systematically review evidence in the empirical literature and exemplify the substantial impact of convexity on cost function estimates and on the determination of scale economies. Therefore, in this contribution we will systematically add a nonconvex technology to the traditional convex technology to assess its potential impact.

In this contribution, we address three questions with regard to our empirical application based on a large unbalanced panel data of Swedish district courts. Firstly, are the average MPI and HMPI significantly different from unity or not? For this question, we calculate the average values of MPI and HMPI under a wide variety of specifications, and then perform a one-sample symmetric Wilcoxon test and t-test to assess whether the average MPI and HMPI are significantly different from unity. To the best of our knowledge, this contribution is the first to combine MPI and HMPI under such wide variety of technology specifications for courts.

The three existing MPI studies on courts employ much less flexible technology specifications. First, Falavigna et al. [24] implement an output-oriented MPI using C and constant returns to scale technologies (except for the MPI components) to examine Italian productivity growth from 2009 to 2011. Second, Giacalone et al. [25] use an output-oriented MPI to examine Italian courts during 2011-2016: these authors seemingly specify C technologies but do not indicate the maintained returns to scale assumption.[1] Third, Mattsson et al. [26] measure an output-oriented MPI using C and constant returns to scale technologies (except for the MPI components) of HM&As in Swedish district courts between 2012 and 2015. Thus, while these studies impose C and constant returns to scale, we also opt for more flexible NC and variable returns to scale technology specifications: this is new in the literature on court productivity.

[1]The fact that Giacalone et al. [25] do not report infeasibilities seems to indicate that these authors use constant returns to scale technologies.

While these three MPI studies on courts opt for an output-oriented measurement (implicitly assuming that courts pursue output maximization), we opt for the analysis of public sector production from the more natural assumption of input-oriented measurement based on the observation that demand for most public services (including public utilities) is exogenous. Studies taking a similar viewpoint include Giménez and Prior [41, p. 4] and Msann and Saad [42, p. 9], for instance, for government studies, and Viton [43, p. 34] and Tovar et al. [44, p. 5396], among others, for public utilities.

Moreover, we assume that inputs are managed at the decentralized level of the individual courts. In this context, our choice for input-orientation makes perfect sense. Of course, one could envision that inputs are managed at a centralized level: reallocations of inputs could indeed potentially enhance efficiency and productivity of a court system. But, while quite some theoretical progress has been made in frontier-based centralized resource allocation models (see, e.g., the survey by Afsharian et al. [45]), such resource allocation mechanisms are to our knowledge not operating in the Swedish courts under study. Furthermore, we are unaware of such centralized resource allocation models that aim to maximize some aggregation of court-base productivity indices.

Secondly, is there any significant difference in productivity changes obtained from the MPI and HMPI using convex and nonconvex technologies? For this question, we compute the two indices under convex and nonconvex technologies and test whether the two indices empirically agree with one another for our unbalanced panel data. In particular, following Kerstens and Van de Woestyne [36] we check whether MPI and HMPI lead to contradictory results, and following Baležentis et al. [37] we also report whether C and NC results of these indices yield contradictory results. To our knowledge we are the first to do so for courts.

Thirdly, is the difference in Malmquist and Hicks-Moorsteen productivity change between the pre-merger and post-merger observations significant under convex and nonconvex technologies for unbalanced panel data? For this problem, we test the productivity changes between the pre-merger and post-merger observations under convex and nonconvex technologies. Again, we are the first to report this for courts.

For these purposes, the remainder of this contribution is structured as follows. Section 2 introduces a brief presentation of technology frontiers and efficiency measures, MPI and HMPI. Then follows a Section 3 with presentation of the unbalanced sample data as well as a summary of research on Swedish district courts. Section 4 presents detailed empirical illustrations based on this secondary data sets of Swedish district courts. The final conclusions are given in Section 5.

2 Definitions of Primal Productivity Indices

We first introduce the definitions of technology frontiers and efficiency measures, and then provide the definitions for computing the input-oriented Malmquist and Hicks-Moorsteen productivity indices.

2.1 Technology Frontiers: Efficiency Measure

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 $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathbb{R}_+^n$ of the evaluating operating units. For each time period t, the production possibility set $\boldsymbol{S}^t$ summarizes the set of all feasible vectors of inputs and outputs, and it is defined as follows:

$$\boldsymbol{S}^t = \{(\boldsymbol{x}^t, \boldsymbol{y}^t) \in \mathbb{R}_+^{m+n} : \boldsymbol{x}^t \text{ can produce at least } \boldsymbol{y}^t\}. \tag{1}$$

Throughout this contribution, it is assumed that technology $\boldsymbol{S}^t$ satisfies some combinations of the following conventional assumptions:
($\boldsymbol{S}$.1) No free lunch and possibility of inaction, i.e., $(0, 0) \in \boldsymbol{S}^t$, and if $(0, \boldsymbol{y}^t) \in \boldsymbol{S}^t$, then $\boldsymbol{y}^t = 0$.
($\boldsymbol{S}$.2) Technology $\boldsymbol{S}^t$ is closed of $\mathbb{R}_+^n \times \mathbb{R}_+^m$.
($\boldsymbol{S}$.3) Strong disposability on inputs and outputs, i.e., $\forall(\boldsymbol{x}^t, \boldsymbol{y}^t) \in \boldsymbol{S}^t : (\boldsymbol{x}^t, -\boldsymbol{y}^t) \leq (\boldsymbol{u}^t, -\boldsymbol{v}^t)$ and $(\boldsymbol{u}^t, \boldsymbol{v}^t) \geq 0$, then implies that $(\boldsymbol{u}^t, \boldsymbol{v}^t) \in \boldsymbol{S}^t$.

The first axiom establishes the possibility of inaction while simultaneously demonstrating that there is no such thing as a free lunch. The second axiom assumes that technology is closed. The third axiom states that inputs and outputs enjoy strong disposability.

The next two additional axioms are sometimes coupled in a different way from the preceding ones:
($\boldsymbol{S}$.4) Technology $\boldsymbol{S}^t$ is convex.
($\boldsymbol{S}$.5) $\delta\boldsymbol{S}^t \subseteq \boldsymbol{S}^t, \forall\delta > 0$.

The fourth axiom of convexity of technology allows for the linear combinations of activities. The fifth axiom imposes constant returns to scale rather than the more flexible variable returns to scale hypothesis that is normally maintained. The latter two axioms are not always maintained in our empirical application.

Efficiency is estimated relative to technologies using radial efficiency measures. In the input-orientation, the radial efficiency measure $E_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t)$ is defined as follows:

$$E_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t) = \min\left\{\lambda \mid (\lambda\, \boldsymbol{x}^t, \boldsymbol{y}^t) \in \boldsymbol{S}^t, \lambda \geq 0\right\} \tag{2}$$

This radial input-oriented efficiency measure indicates the minimum contraction of an input vector by a scalar λ while still remaining in the input correspondence. Obviously, the resulting input combination is located at the boundary of this input correspondence. For our purpose, this radial input efficiency has two nice properties (see, e.g., Hackman [46]). First, it is smaller than or equal to unity ($0 < E_i^t\,(\boldsymbol{x}^t, \boldsymbol{y}^t) \leq 1$), whereby efficient production on the isoquant of $L(\boldsymbol{y}^t)$ is represented by unity and $1 - E_i^t\,(\boldsymbol{x}^t, \boldsymbol{y}^t)$ indicates the amount of inefficiency. Second, it has a cost interpretation.

In the output-orientation, the radial efficiency measure $E_o^t(\boldsymbol{x}^t, \boldsymbol{y}^t)$ is defined as the maximum expansion of the output vector by a scalar θ to the boundary of the technology, expressed as follows:

$$E_o^t(\boldsymbol{x}^t, \boldsymbol{y}^t) = \max\left\{\theta \mid (\boldsymbol{x}^t, \theta\, \boldsymbol{y}^t) \in \boldsymbol{S}^t, \theta \geq 1\right\} \tag{3}$$

This radial output-oriented efficiency measure indicates the maximum proportional expansion of an output vector by a scalar θ while still remaining in the output correspondence. The radial output efficiency also has two useful properties. First, it is larger than or equal to unity ($E_o^t\,(\boldsymbol{x}^t, \boldsymbol{y}^t) \geq 1$), whereby efficient production on the isoquant of the output set $P(\boldsymbol{x^t})$ related to technology $\boldsymbol{S}^t$ is represented by unity and $E_o^t\,(\boldsymbol{x}^t, \boldsymbol{y}^t) - 1$ indicates the amount of inefficiency. Second, it has a revenue interpretation.

In addition, for $(a, b) \in \{t, t+1\}$, the time-related versions of the radial input efficiency measure are given as follows:

$$E_i^a(\boldsymbol{x}^b, \boldsymbol{y}^b) = \min\left\{\lambda \mid (\lambda\, \boldsymbol{x}^b, \boldsymbol{y}^b) \in \boldsymbol{S}^a\right\} \tag{4}$$

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

Similarly, the time-related versions of the radial output efficiency measure are provided as follows:

$$E_o^a(\boldsymbol{x}^b, \boldsymbol{y}^b) = \max\left\{\theta \mid (\boldsymbol{x}^b, \theta\, \boldsymbol{y}^b) \in \boldsymbol{S}^a\right\} \tag{5}$$

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

Furthermore, following Briec et al. [47], the convex and nonconvex nonparametric technologies under constant and variable returns to scale assumptions for a sample of K observations is defined by the following productivity possibility sets:

$$\boldsymbol{S}^{t,\Lambda,\Gamma} = \left\{ (\boldsymbol{x}^t, \boldsymbol{y}^t) \in \mathbb{R}_+^{m+n} : \ \boldsymbol{x}^t \geq \sum_{k=1}^{K} \alpha \ z_k^t \ \boldsymbol{x}_k^t, \ \boldsymbol{y}^t \leq \sum_{k=1}^{K} \alpha \ z_k^t \ \boldsymbol{y}_k^t, \ \boldsymbol{z} \in \Lambda, \alpha \in \Gamma \right\}. \tag{6}$$

The sets Γ are $\Gamma \equiv \Gamma^{\text{CRS}} = \{\alpha : \ \alpha \geq 0\}$; and $\Gamma \equiv \Gamma^{\text{VRS}} = \{\alpha : \ \alpha = 1\}$. The sets Λ are $\Lambda \equiv \Lambda^{\text{C}} = \left\{ \boldsymbol{z} = (z_1^t, \ldots, z_k^t) | \sum_{k=1}^{K} z_k^t = 1 \text{ and } \forall k \in \{1, 2, ..., K\} : z_k^t \geq 0 \right\}$; and $\Lambda \equiv \Lambda^{\text{NC}} = \left\{ \boldsymbol{z} | \sum_{k=1}^{K} z_k^t = 1 \text{ and } \forall k \in \{1, 2, ..., K\} : z_k^t \in \{0, 1\} \right\}$. First, there is the activity vector ($\boldsymbol{z}$) operating subject to a convexity (C) or a 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 either fixed at unity under variable returns to scale (VRS), or simply non-negative under constant returns to scale (CRS).

In the nonconvex case, axioms ($\boldsymbol{S}$.1)–($\boldsymbol{S}$.3) are retained, whereas the convex case imposes ($\boldsymbol{S}$.1)–($\boldsymbol{S}$.4). Furthermore, each of these systems can impose CRS ($\boldsymbol{S}$.5) rather than VRS. This unified specification is nonlinear, although it can be easily linearized in the convex situation. In the nonconvex scenario, it entails solving either nonlinear mixed integer problems or scaled vector dominance methods (see Briec et al. [48] for details).

2.2 Input-oriented Malmquist Productivity Index

In line with Chen et al. [49] and as customary in public sector analysis (see Section 1), we prefer an input-oriented efficiency measure since the outputs are determined by the demand for justice of citizens. Thus, public sector bureaucrats can at most control inputs.

Using the radial input efficiency measures, the input-oriented Malmquist productivity index in base period t can be defined as follows:

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

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

In a similar vein, an input-oriented Malmquist productivity index with base period $t+1$

is also defined as:

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

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

Moreover, to avoid an arbitrary choice of base period, Färe et al. [50] follow Caves et al. [28, p. 1397-1398] and propose defining the input-oriented Malmquist productivity index as a geometric mean of a period t and a period $t+1$ index:

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

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

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

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

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

Remark that the above definitions deviate from the original ones in Caves et al. [28] in that the ratios have been inverted. This ensures that productivity indices above (below) unity reveal productivity growth (decline), which is in line with traditional productivity indices.

Moreover, the efficiency measures in (9) can be obtained by solving four different mathematical programming models (expressed as linear programming models or binary integer programming models): these four different models are described in Appendix A.

2.3 Hicks-Moorsteen Productivity Index

According to Bjurek [31], a HMPI with base period t is defined as a ratio of an output-oriented Malmquist index in period t over an input-oriented Malmquist index in period t as follows:

$$HM^t(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}) = \frac{M_o^t(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{y}^{t+1})}{M_i^t(\boldsymbol{x}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^t)} \tag{11}$$

where the output quantity index is defined as $M_o^t(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{y}^{t+1}) = \frac{E_o^t(\boldsymbol{x}^t, \boldsymbol{y}^t)}{E_o^t(\boldsymbol{x}^t, \boldsymbol{y}^{t+1})}$ and the input quantity index is $M_i^t(\boldsymbol{x}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^t) = \frac{E_i^t(\boldsymbol{x}^t, \boldsymbol{y}^t)}{E_i^t(\boldsymbol{x}^{t+1}, \boldsymbol{y}^t)}$. If the HMPI index is larger (smaller) than unity, then it reveals that there is an improvement (decline) in productivity.

In a similar way, a HMPI with base period $t+1$ is defined as a ratio of an output-oriented Malmquist index in period $t+1$ over an input-oriented Malmquist index in period $t+1$ as follows:

$$HM^{t+1}(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}) = \frac{M_o^{t+1}(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}, \boldsymbol{y}^t)}{M_i^{t+1}(\boldsymbol{x}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})} \tag{12}$$

where the output quantity index is defined as $M_o^{t+1}(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}, \boldsymbol{y}^t) = \frac{E_o^{t+1}(\boldsymbol{x}^{t+1}, \boldsymbol{y}^t)}{E_o^{t+1}(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})}$ and the input quantity index is $M_i^{t+1}(\boldsymbol{x}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}) = \frac{E_i^{t+1}(\boldsymbol{x}^t, \boldsymbol{y}^{t+1})}{E_i^{t+1}(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})}$. Once again, if the HMPI index is larger (smaller) than unity, it reveals a gain (loss) in productivity.

Moreover, to avoid a choice of base year, we can take a geometric mean of these two HMPI:

$$HM^{t,t+1}(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}) = \sqrt{HM^t(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1}) \cdot HM^{t+1}(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})} \tag{13}$$

Once more, if the HMPI is larger (smaller) than unity, it shows an increase (decrease) in productivity. Note that the HMPI defined above is always a TFP index compared with the MPI discussed previously.

Furthermore, the efficiency measures in (13) can be obtained by solving eight different models (which are also described as linear or binary integer programming models): these eight different models are also described in Appendix A.

Finally, we can find that in the base period t, the denominator of both the Malmquist output and input quantity indices compares a “hypothetical”observation consisting of inputs and outputs seen from various time periods to a technology in period t. In base period $t+1$, the same observation applies to the numerator for the appropriate Malmquist output and input quantity indices. Such "hypothetical" observations do not show up in the MPI, making its interpretation considerably simpler.

2.4 Primal Productivity Indices: A Comparison

Following Kerstens and Van de Woestyne [36, Section 2.4], we conclude this section with some comparative comments on the characteristics of these two primal productivity indices MPI and HMPI.

First, Bjurek et al. [51] argue that the MPI is not always a TFP index, although its TFP characteristics are maintained in the case of CRS. However, Grifell-Tatjé and Lovell [52] illustrate that these TFP characteristics are not preserved in the case of VRS. Following Grosskopf [53], the MPI only captures local technical change (that is, changes in the production frontier that take into account efficiency changes), but it cannot be a measure of TFP change in general. By contrast, Bjurek [31] is first to state that the HMPI has a TFP interpretation. Moreover, O'Donnell [30] mentions that many TFP indices are decomposable into measures of technical change and technical efficiency change (as initiated by Nishimizu and Page [54]), but furthermore into scale efficiency change and mix efficiency change components.

Second, for the MPI we stick to the basic decomposition between technical change and technical efficiency change (see Nishimizu and Page [54]) that is almost universally accepted. More evolved decompositions have been proposed in the literature (see, e.g., Zofío [55] for a survey), but these are not without controversy. We have not adopted any decomposition of the HMPI, since such decomposition is of recent date (see Diewert and Fox [56] for the initial proposal) and it may not be universally accepted. Furthermore, this decomposition of the HMPI is distinct from the decompositions of the MPI, which makes comparisons difficult if not impossible. Therefore, the MPI is decomposed into technical change and technical efficiency change, while the HMPI is not decomposed, and we focus on comparing both productivity indices as such.

Third, Färe et al. [57] prove that the MPI and HMPI coincide under the strong conditions of both CRS and inverse homotheticity. Balk [58] somewhat weakens these conditions and proves that MPI and HMPI coincide under CRS and either input homotheticity or output homotheticity. Mizobuchi [59] also proves that MPI and HMPI coincide under CRS and Hicks neutral technical change. Moreover, Bjurek et al. [51] also mention some additional relationships: in the situation of (i) a single input and many outputs, (ii) a single output and multiple inputs, and (iii) where all inputs and/or outputs of a unit change proportionally, both MPI and HMPI coincide under CRS. Lastly, O'Donnell [30, p. 258] states that in the absence of technological change both MPI and HMPI are equal under CRS.

Fourth, when estimated using generic methodologies, several of the distance functions that comprise the MPI are likely to be undefined or infeasible (see Färe et al. [50]). Empirical studies frequently fail to report on this issue of infeasibility of the MPI. Briec and Kerstens [29] show that infeasibility can also develop for a more general productivity indicator based on a more general distance function. Therefore, even this more general indicator fails to obey the determinateness property defined in index theory. By contrast, the HMPI meets determinateness as validated by Briec and Kerstens [32] for strong disposability of inputs and outputs.

Fifth, in the light of the above comment, it is crucial to distinguish between an infeasibility owing to unavailable data (e.g., due to the panel's unbalanced nature) and a computational infeasibility. The former case is arguably better described as a logical impossibility since the underlying adjacent period efficiency measures that are part of the MPI simply cannot be measured. Overall, the TFP character of the HMPI, as well as the ease with which it can be made transitive by selecting the appropriate basis, indicates that it deserves more consideration.

Finally, since we focus on comparing C and NC productivity, one should mention that Ang et al. [60] specify the conditions under which the C and NC HMPI coincide. In particular, when there is a single output and a single variable input, then the C and NC HMPI coincide and can be obtained via implicit enumeration algorithms.

3 Unbalanced Panel Data from Swedish District Courts and Focused Literature Review

3.1 Description of the Sample

Drawing on Mattsson et al. [26], Mattsson and Tidanå [15], and Agrell et al. [61], the sample is an unbalanced panel over the years 2000-2017 of Swedish district courts based on annual statistics. Due to the available data, we choose the same inputs and same outputs.[2] We first

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

discuss the specification of inputs and outputs.

There are three labor inputs ((i) judges, (ii) law clerks, and (iii) administrative employees measured as full-time equivalents (other personnel)), and a capital input (court area). Among these inputs, court area is selected as a proxy 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), but also to operational expenditures (e.g., heating, maintenance, and insurance). Mattsson et al. [26], Mattsson and Tidanå [15], and Agrell et al. [61] argue that including a measure of capital is important, since it is to some extent possible to substitute labor with capital in court production. An example is the incorporation of video conferences, which decreases the traveling time for judges according to the Swedish National Court Administration (SNCA).

In addition, there are three outputs: decided criminal cases, decided civil cases and decided matters. For more institutional details on the Swedish court system and the role of district courts, readers are referred to these three above existing studies.

While in our case the choice of inputs and outputs follows the earlier specifications of Mattsson et al. [26], Mattsson and Tidanå [15], and Agrell et al. [61], it should not be forgotten that the basic question as to the specification of inputs and outputs for courts is not without controversy. The surveys by Voigt [12, Section 6] and Falavigna and Ippoliti [62, Table 1] provide a detailed discussion with particular attention to the status of stock and flow variables. For instance, are pending cases inputs in the courts production process, or are pending cases intermediate inputs or outputs? It is surprising to see that even today there is yet no consensus on these very basic questions on court production.

The descriptive statistics of the average level and standard deviations are reported in Table 1. As it can be seen, all input and output indicators show an increasing trend from 2000 to 2017. For example, the number of criminal cases goes on average up from 674.67 to 2144.09 (3.18 times), and the number of full-time equivalent law clerks expands from 5.28 to 15.32 (2.90 times). In addition, note that the standard deviations of civil cases, criminal cases and law clerks are almost as large as their averages. However, the standard deviations of output matters, and the inputs of judges, other personnel and court area vary very little.

3.2 What is Known about the Performance Impact of Horizontal Mergers in Swedish District Courts?

Table 1: Descriptive Statistics of Inputs and Outputs over the Years

Years	Outputs			Inputs			
	Civil Cases	Matters	Criminal Cases	Judges	Law Clerk	Personnel	Court Area
2000	538.67 (996.86)	309.40 (554.18)	674.67 (937.82)	6.88 (12.54)	5.28 (7.28)	13.77 (22.85)	2545.44 (3759.61)
2001	545.11 (1033.60)	307.10 (583.11)	694.42 (1010.07)	6.77 (12.64)	5.88 (8.86)	12.06 (24.48)	2431.77 (3905.87)
2002	659.58 (1100.40)	381.90 (684.08)	871.41 (1279.75)	8.20 (12.49)	6.90 (9.07)	15.15 (25.83)	2826.34 (4277.40)
2003	709.96 (1149.60)	430.37 (781.78)	966.61 (1329.20)	8.64 (13.57)	7.43 (9.43)	16.61 (27.51)	2976.49 (4312.72)
2004	732.46 (1178.10)	431.49 (724.64)	994.76 (1363.65)	8.57 (13.04)	6.96 (8.71)	16.05 (27.08)	2947.66 (4355.62)
2005	756.08 (1163.37)	453.40 (807.05)	1062.12 (1414.87)	8.45 (12.88)	7.17 (9.00)	16.52 (27.48)	2978.64 (4589.40)
2006	900.63 (1214.52)	507.70 (766.84)	1320.42 (1525.47)	10.60 (14.36)	10.15 (12.52)	20.50 (30.04)	3587.35 (4847.75)
2007	901.79 (1031.68)	488.95 (599.25)	1369.44 (1240.41)	10.41 (12.04)	10.42 (10.76)	21.64 (26.21)	3428.62 (3795.91)
2008	1124.97 (1227.90)	532.12 (537.46)	1692.34 (1332.40)	11.94 (10.99)	12.01 (11.11)	24.54 (25.51)	3818.40 (3705.45)
2009	1239.33 (1312.78)	574.86 (546.29)	1764.74 (1444.52)	12.09 (10.59)	11.49 (10.76)	24.77 (25.16)	3925.16 (4220.51)
2010	1410.55 (1468.59)	645.91 (611.20)	1951.05 (1550.58)	12.86 (11.25)	13.63 (13.14)	25.97 (24.55)	4267.81 (4176.22)
2011	1402.48 (1456.66)	677.53 (669.73)	2045.46 (1606.21)	13.67 (12.19)	15.05 (14.34)	26.47 (24.35)	4380.81 (4182.80)
2012	1458.05 (1484.60)	499.21 (473.05)	2111.06 (1811.46)	14.42 (12.29)	15.47 (14.95)	26.42 (24.68)	4370.77 (4196.37)
2013	1500.93 (1530.99)	513.67 (519.15)	2051.45 (1823.05)	14.77 (13.09)	15.78 (15.00)	27.34 (25.69)	4464.98 (4218.27)
2014	1519.02 (1531.44)	512.18 (507.30)	2034.52 (1811.65)	15.52 (13.77)	16.06 (15.17)	27.73 (25.79)	4461.00 (4179.80)
2015	1428.08 (1513.24)	507.86 (513.45)	2042.89 (1806.69)	15.76 (13.74)	15.81 (15.26)	28.02 (26.83)	4510.96 (4174.75)
2016	1342.87 (1374.82)	481.88 (482.61)	2067.05 (1869.23)	16.01 (14.54)	15.86 (15.51)	27.58 (26.42)	4554.46 (4086.36)
2017	1395.79 (1414.38)	464.61 (423.44)	2144.09 (1903.39)	15.56 (13.17)	15.32 (14.39)	27.71 (25.48)	4509.27 (3702.75)

Standard deviation is displayed in parentheses.

Extending Chen and Kerstens [63, p. 221-222], the impact of horizontal mergers in Swedish district courts on performance from a variety of perspectives using deterministic nonparametric frontier methods has already lead to the following studies known to us.

First, Mattsson and Tidanå [15] utilize a nonparametric frontier decomposition method developed by Bogetoft and Wang [13] to identify the potential ex ante merger gains: these authors find that some mergers have little potential for efficiency gains, while others can generate significant merger gains.

Second, Agrell et al. [61] use a nonparametric frontier method to measure the ex post efficiency of horizontal mergers for the Swedish district courts: they report that the merged courts are more efficient than the non-merged ones.

Moreover, Chen et al. [49] are the first to combine traditional C with NC nonparametric frontier methods to calculate technical and scale efficiency before and after the HM&As of Swedish district courts. These authors suggest that HM&As improve efficiency mainly via scale efficiency under nonconvexity, but technical efficiency under convexity. In addition, under convexity most observations are decreasing returns to scale, while under nonconvexity one could have selected among increasing returns to scale observations.

Third, Mattsson et al. [26] include the potential heterogeneity between outputs by a weighting based on differences in spent resources between 14 categories. These same authors also use a super-efficiency model to eliminate outliers, and furthermore compute the output-oriented MPI by applying a nonparametric frontier technology to evaluate efficiency and productivity changes from 2011 to 2015: they report a 1.7% average productivity loss per year.

Fourth, Chen and Kerstens [63] are the first to apply input- and output-oriented plant capacity concepts to assess horizontal mergers. These authors empirically illustrate that horizontal mergers doubles input-oriented plant capacity utilisation, while output-oriented plant capacity improves somewhat.

Finally, Månsson et al. [64] is seemingly the first analysis of the cost efficiency of courts and this cost efficiency is decomposed into technical and allocative efficiencies. Under C this study finds substantial cost inefficiencies mainly due to allocative inefficiencies, which may be due to regional heterogeneity in input prices.

Our own study contributes to the knowledge about horizontal mergers in Swedish district courts by analysing MPI and HMPI under the angle of a wide variety of technology specifications.

4 Empirical Illustration

In this section, our empirical analysis proceeds in three steps. The first step is the calculation of the MPI and HMPI to obtain the information on whether there are substantial changes in performance over time. The second step is the comparison of the changes underlying both MPI and HMPI under convex and nonconvex specifications. Finally, the third step is a detailed comparison of the productivity changes between the pre-merger and post-merger observations.

4.1 Malmquist and Hicks-Moorsteen Productivity Indices: Any Substantial Progress?

The main purpose of this contribution is to focus on the productivity changes of HM&As among Swedish district courts. Hence, we are concerned with the extent to which productivity changes are substantial in the context of our unbalanced panel data. For this issue, we calculate the MPI and HMPI and conduct a one-sample symmetric Wilcoxon test and t-test to evaluate whether both average productivity indices are significantly different from unity or not.[3] The descriptive statistics of the average level of input-oriented MPI and HMPI during the period 2000-2017 are reported in Table 2 for eight different specifications of technology: convex versus nonconvex, and CRS versus VRS. Per productivity index, the first (last) two columns report convex (nonconvex) technologies under CRS and VRS. The rows list the years two by two.

Table 2: Productivity Indices under Different Technology Specifications

Years	Malmquist Productivity Index				Hicks-Moorsteen Productivity Index			
	$S^{C,CRS}$	$S^{C,VRS}$	$S^{NC,CRS}$	$S^{NC,VRS}$	$S^{C,CRS}$	$S^{C,VRS}$	$S^{NC,CRS}$	$S^{NC,VRS}$
2000-2001	1.0420	0.9130	0.9872	0.9284	1.0403	1.0426	1.0327	1.0350
2001-2002	1.0624	1.1209	1.0747	1.1068	1.0719	1.0916	1.0806	1.0820
2002-2003	1.0586	1.0120	1.0392	1.0131	1.0552	1.0612	1.0288	1.0445
2003-2004	1.0647	1.0904	1.0656	1.0707	1.0662	1.0661	1.0634	1.0665
2004-2005	1.1245	1.1874	1.1058	1.1773	1.1246	1.1290	1.1004	1.1042
2005-2006	0.9669	0.9236	0.9458	0.9507	0.9572	0.9481	0.9310	0.9526
2006-2007	1.0085	1.0110	1.0017	1.0546	1.0105	1.0023	1.0022	0.9955
2007-2008	1.0503	1.0566	1.0361	1.0641	1.0471	1.0414	1.0302	1.0203
2008-2009	1.0546	1.0679	1.0425	1.1108	1.0537	1.0501	1.0465	1.0748
2009-2010	1.0308	1.0141	1.0301	1.0897	1.0340	1.0134	1.0355	1.0355
2010-2011	0.9787	1.0049	0.9825	0.9967	0.9785	0.9854	0.9802	0.9732
2011-2012	0.9549	0.9467	0.9343	0.9028	0.9544	0.9196	0.9287	0.8964
2012-2013	0.9771	0.9784	0.9802	0.9922	0.9776	0.9795	0.9774	0.9830
2013-2014	0.9886	0.9901	0.9849	1.0053	0.9899	0.9875	0.9851	0.9819
2014-2015	0.9547	0.9595	0.9539	0.9460	0.9554	0.9544	0.9567	0.9582
2015-2016	0.9844	0.9788	0.9928	0.9766	0.9849	0.9857	0.9903	0.9785
2016-2017	1.0222	1.0199	1.0336	1.0583	1.0222	1.0216	1.0340	1.0226
Average	1.0191	1.0162	1.0112	1.0261	1.0190	1.0164	1.0120	1.0120
Stand. Dev	0.0475	0.0712	0.0475	0.0741	0.0485	0.0545	0.0490	0.0542
One-sample Wilcoxon test	109	90	93	105	109	99	99	97
p-value	0.132	0.548	0.459	0.190	0.132	0.306	0.306	0.353
One-sample t-test†	1.652	0.938	0.974	1.454	1.618	1.245	1.007	0.915
p-value	0.118	0.362	0.344	0.165	0.125	0.231	0.329	0.374

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

First, looking at the descriptive statistics over the years 2000 till 2017 at the bottom of Table 2, the following conclusions can be deduced. First, from the average values of the two productivity indices for all specifications both under convex and nonconvex technologies: these are slightly larger than unity, but rather close to unity. There is no contradictory result for MPI and HMPI for all specifications under both measures. Second, average values state that these two productivity indices are almost identical under a $S^{C,CRS}$ specification. Finally, from the standard deviations of the two productivity indices, productivity indices are seeming more stable for changes under a CRS specification both under convex and nonconvex

[3] While the t-test may be somewhat more powerful if the underlying assumption of normality is valid, it tends to be more sensitive to outliers than the non-parametric one-sample symmetric Wilcoxon signed rank test. For robustness sake, we simply report both test statistics.

technologies.

Second, we look at the slight differences between the average level of the two productivity indices over time. We can make the following pertinent observations. First, years with small productivity growth are followed with years of small productivity losses. Second, MPI and HMPI seem to agree on most of these changes over the years except that for some years there are contradictions between technology specifications and between both indices. We list now these contradictions between technology specifications for the MPI: (i) for the year 2000-2001 $S^{C,CRS}$ yields growth, while the other technologies indicate a decline; (ii) for the year 2010-2011 $S^{C,VRS}$ yields growth, while the other technologies indicate a decline; and (iii) for the year 2013-2014 $S^{NC,VRS}$ yields growth, while the other technologies indicate a decline. We list now these contradictions between technology specifications for the HMPI: for the single year 2006-2007 $S^{NC,VRS}$ yields decline, while the other technologies indicate growth. Thus, HMPI seems slightly more robust for changes in technology specifications. Turning to the contradictions between both indices we observe that all of the above four listed pairs of years leads to minor conflicts between MPI and HMPI. Examples include for (i) $S^{C,VRS}$ the years 2000-2001 and 2010-2011; (ii) $S^{NC,CRS}$ the year 200-2001; and $S^{NC,VRS}$ the years 2000-2001, 2006-2007, and 2013-2014. These conflicting results between MPI and HMPI are certainly in line with the ones reported in Kerstens and Van de Woestyne [36] for different data sets: French fruit producers and hydro-electric power plants..

Therefore, we also perform a one-sample symmetric Wilcoxon signed rank test and a t-test to evaluate whether the average MPI and HMPI are significantly different from unity (null hypothesis $H_0 : \mu = 1$) or not (alternative hypothesis $H_1 : \mu \neq 1$) and we report the corresponding p-values at the bottom of Table 2. If the p-value is larger than 0.05, then it means that we cannot reject the null hypothesis that the average productivity is equal to unity at the 5% significance level. If the p-value is smaller than 0.05, then we can rather safely reject the null hypothesis and consider that the average productivity differs from unity. From the last two lines of Table 2, the p-values obtained from the one-sample symmetric Wilcoxon test and t-test are all greater than 0.05. These results indicate that average levels for both MPI and HMPI are equal to unity and thus that there is no obvious improvement or deterioration in productivity among Swedish district courts over our period.

Mattsson et al. [26] compute an MPI between the shorter period from 2012 to 2015 and report an average annual productivity decline of 1.7%. These more limited results for Swedish district courts are thus not corroborated over our longer panel data. Though Mattsson et al. [26] opt for an output-orientation, one should realise that under C and CRS (maintained in their study) input-and output-oriented MPI coincide. The differences in results are then explained by the different years studied, the elimination of some outliers, and the fact that outputs in their study are weighted by hearing time.

4.2 Empirical Results for the Primal Productivity Indices

While Table 2 only provides average values per pair of years as well as limited descriptive statistics over all the years combined (average and standard deviation), Table 3 contains basic descriptive statistics for both the MPI and HMPI under C and NC technologies for the Swedish district courts over all observations in the sample combined. Given the number of periods covered, it is simply not possible to report the results for each period in great detail, apart from the yearly averages in the previous Table 2. Thus, only the overall descriptive statistics results are reported.

More specifically, the first horizontal part of Table 3 reports the relative presence of non-availability due to unavailable data (denoted "na"). Then, the top four lines of the second and third horizontal parts report descriptive statistics for the average, standard deviation, minimum and maximum values of the MPI and HMPI, respectively. In addition, the computational infeasibilities (denoted "Inf") for the Swedish district courts and the Li-test results between C and NC are reported in the fifth and sixth lines of the second and third horizontal parts.

Finally, in the fourth horizontal part the contradictory results at the individual observation level and the Li-test statistics between MPI and HMPI are given. Indeed, these efficiency measures are compared by performing nonparametric tests comparing two overall distributions as originally developed by Li [65] and refined by Fan and Ullah [66] and most recently by Li et al. [67]. The Li-test statistic tests for the eventual significance of differences between two kernel density estimates of some statistical distribution. This distributional test must be distinguished from the more common location test. A location test is a type of statistical hypothesis test that compares a statistical population's location parameter (e.g., mean, median, or mode) to a fixed constant (e.g., see the above one-sample symmetric t-test) or the location parameters of two statistical populations to one another.

The null hypothesis of the Li-test statistic states that both density functions are almost equal. The alternative hypothesis agrees that both density functions are different.[4] To get around the issue of spurious mass at the boundary, Simar and Zelenyuk [68] further refine

[4]This test is valid for both dependent and independent variables. Matlab code developed by P.J. Kerstens based on Li et al. [67] is found at: `https://github.com/kepiej/DEAUtils`.

this test statistic for nonparametric frontier estimators. Their Algorithm I ignores boundary estimates, and their Algorithm II smoothes boundary estimates by adding uniform noise of a smaller order of magnitude than the noise added by the particular estimator. Algorithm II may perform somewhat better overall, according to Monte Carlo evidence, however the strength of the test statistic declines with the dimensionality of the production specification. In short, we utilize the Algorithm II from Simar and Zelenyuk [68] version of this test modified by Li et al. [67]. All the kernel densities underlying these Li-test results are made available in Appendix B. All these results are commented upon below in a sequential way.

Table 3: Descriptive Statistics for Malmquist and Hicks-Moorsteen Productivity Indices

Productivity Indices		Convex		Nonconvex	
		$S^{C,CRS}$	$S^{C,VRS}$	$S^{NC,CRS}$	$S^{NC,VRS}$
	% Non-availability (%na)	0.0618	0.0618	0.0618	0.0618
Malmquist Productivity Index (MPI)	Average	1.0257	1.0205	1.0159	1.0290
	Stand.Dev	0.1596	0.2870	0.1343	0.2785
	Min	0.4043	0.1475	0.4317	0.1883
	Max	2.9305	4.9314	2.2963	3.5104
	% Infeasibilities (Inf)	0.0000	0.0314	0.0000	0.0334
	Simar & Zelenyuk Li-test† (p-value) (C vs. NC)	1.8703 (0.0294**) (CRS)		3.7541 (0.0016***) (VRS)	
	One-sample Wilcoxon test (p-value)	18713 (0.000***)	13385 (0.727)	16583 (0.009**)	14897 (0.323)
	One-sample t-test‡ (p-value)	4.839 (0.000***)	0.9585 (0.339)	3.3739 (0.001***)	1.2682 (0.206)
Hicks-Moorsteen Productivity Index (HMPI)	Average	1.0258	1.0246	1.0182	1.0192
	Stand.Dev	0.1526	0.1655	0.1332	0.1627
	Min	0.4211	0.4102	0.4035	0.3515
	Max	2.7970	2.8319	2.2787	2.7119
	% Infeasibilities (Inf)	0.0000	0.0000	0.0000	0.0000
	Simar & Zelenyuk Li-test (p-value) (C vs. NC)	1.9405 (0.0304**) (CRS)		0.5632 (0.2160) (VRS)	
	One-sample Wilcoxon test (p-value)	18857 (0.000***)	19508 (0.000***)	18594 (0.000***)	19128 (0.000***)
	One-sample t-test (p-value)	5.0993 (0.000***)	5.263 (0.000***)	4.9258 (0.000***)	5.0778 (0.000***)
% Contrad. Res. MPI/ HMPI		0.0255	0.1631	0.1306	0.3821
% Contrad. Res. C/ NC		0.1729 (MPI-CRS)	0.2456 (MPI-VRS)	0.1562 (HMPI-CRS)	0.1621 (HMPI-VRS)
Simar & Zelenyuk Li-test (p-value) (MPI/HMPI)		-1.8998 (1.000)	-0.4934 (0.685)	-0.9337 (0.926)	5.4613 (0.000***)

† Li-test: critical values at 1% level=2.33(***); 5% level=1.64(**); 10% level =1.28(*).
‡ t-test: critical values at 1% level=2.55 (***); 5% level=1.73 (**); 10% level=1.33(*).

First, the first horizontal part lists the unavailable data for the unbalanced panel: it is the same for both MPI and HMPI at 6.18%. Second, the average values for CRS MPI and HMPI under C are 1.0257 and 1.0258, and the average values for CRS MPI and HMPI under NC are 1.0159 and 1.0182. This shows that the two productivity results are all larger than unity and very close to one another for C, and the average result is slightly different under NC, with HMPI being a bit higher. Moreover, the average values for VRS MPI and HMPI under C are 1.0205 and 1.0246, and the average value for VRS MPI and HMPI under NC measure are 1.0290 and 1.0192. This indicates that the two productivity results are all larger than unity, while larger average values appear for HMPI under C and for MPI under NC. In addition, notice that the standard deviations for HMPI for CRS and VRS both under C and NC measures are smaller than that for MPI. Note that average values and standard deviations are slightly different in Tables 2 and 3: average values and standard deviations in Table 2 are computed over the yearly averages and standard deviations, while in Table 3 these are computed over all observations.

Next, we analyze the infeasibilities of MPI depending on the various technology specifications. The percentage of computational infeasibilities ("Inf") seems rather stable. More specifically, for the MPI with $S^{C,CRS}$ and $S^{NC,CRS}$ specifications, we find no computational infeasibilities. However, for the MPI with $S^{C,VRS}$ specification, it is 3.14%. While for the MPI with $S^{NC,VRS}$ specification, it is 3.34%. Thus, more infeasibilities appear under the more flexible $S^{NC,VRS}$ compared to less flexible $S^{C,VRS}$ specifications. Moreover, the HMPI does not have a single computational infeasibility and is always feasible for all the technology specifications over all periods. In addition, the bottom line containing the results of the Li-test statistic in the second and third horizontal parts under C and NC confirm that MPI under both CRS and VRS specifications and HMPI under the CRS specification solely differ significantly at the 5% significance level. Only HMPI under the VRS specification has no significant difference between C and NC.

We also report the results and p-values from a one-sample symmetric Wilcoxon test and a t-test similar to above of MPI and HMPI for CRS and VRS specifications under C and NC in the second and third horizontal parts. The following observations can be made. First, for the MPI results, the p-values under the $S^{C,CRS}$ and $S^{NC,CRS}$ specifications are smaller than 0.05, while under the $S^{C,VRS}$ and $S^{NC,VRS}$ specifications the p-values are larger than 0.05, implying that only under the latter more flexible specifications no significant productivity changes occur. Second, for the HMPI results the p-values under the four specifications ($S^{C,CRS}$, $S^{C,VRS}$, $S^{NC,CRS}$ and $S^{NC,VRS}$) are all smaller than 0.05 meaning that there exists obvious productivity improvement or deterioration.

Note that these one-sample symmetric Wilcoxon test and t-test results are slightly different in Tables 2 and 3: in Table 2 the one-sample symmetric Wilcoxon test and t-test statistic is computed by using the 17 average productivity values, while in Table 3 the one-sample symmetric Wilcoxon test and t-test statistic is performed by using all the observations over all years. This leads to slightly different results: mainly under C (in contrast to NC) one can reject the null hypothesis of no productivity growth.

Moreover, the above sample-level results may hide contradictory results at the individual observation level. In this regard, the fourth horizontal part in each specification reports the number of observations between the MPI and HMPI, and between C and NC that provide conflicting results: a single observation points to a decline in productivity under one index or under C(NC), while the same observation shows an increase in productivity under another index (denoted "% $Contrad.\ Res.\ MPI/HMPI$") or under NC(C) (denoted "% $Contrad.\ Res.\ C/NC$"), or vice versa.

On the one hand, between the MPI and HMPI, we find that for the CRS assumption about 2.55% under C and 13.06% under NC yield contradictory results. For the VRS specification, about 16.31% under C and 38.21% under NC obtain contradictory results. Hence, more contradictory results appear under VRS compared to CRS and under NC compared to C. This result is in line with results reported in Kerstens and Van de Woestyne [36, Tables 1 and 2]. On the other hand, between C and NC we can find an about 17.29% under CRS and 24.56% under VRS for MPI obtaining contradictory results, and an about 15.62% under CRS and 16.21% under VRS for HMPI yielding contradictory results. Therefore, more contradictory results appear under VRS compared to CRS and for MPI compared to HMPI. The latter result is in line with the results reported in Baležentis et al. [37, Table VI]: the latter authors employ a Luenberger-Hicks-Moorsteen productivity indicator (an additive counterpart of the HMPI) under VRS solely using an environmental by-production technology applied to a sample of Chinese textile companies..

In addition, the last line containing the results of the Li-test statistic reveal that the differences between MPI and HMPI are significantly different under the more flexible $S^{NC,VRS}$ specification at the 1% significance level. However, for the $S^{C,CRS}$, $S^{NC,CRS}$ and $S^{C,VRS}$ specifications, MPI and HMPI results are identical.

4.3 Did Horizontal Mergers Increase Productivity?

In addition to the empirical analysis at the sample level above, we now can dig deeper in detail by focusing on the comparison between pre-merger and post-merger observations solely. In this subsection, 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. [61] for details). In addition, although in general horizontal mergers consist of a relatively large district court taking over one or more smaller (adjacent) district courts, during this period some of the new courts are formed from parts of the original courts rather than merely two or more other courts. For instance, as mentioned in Agrell et al. [61, p. 673], there are in total seven earlier courts that are dissolved and merged into five new courts containing parts of earlier courts in 2007.[5] We do not consider these five merger scenarios described above. Hence, for the merger case of 76 (=83-7) courts into 31 (=36-5) courts, we conduct a comparative analysis and perform a Li-test statistic on the productivity change for all periods before the pre-merger and all periods after the merger.

[5]For more details about these mergers in parts, see Chen et al. [49, Section 3].

Table 4: Productivity between the Pre-merger and Post-merger Observations

	MPI		HMPI	
	Pre-merger obs.	Post-merger obs.	Pre-merger obs.	Post-merger obs.
$S^{C,CRS}$	1.1194	1.0628	1.1225	1.0609
$S^{C,VRS}$	0.9953	1.0665	1.1678	1.0592
$S^{NC,CRS}$	1.1028	1.0541	1.1057	1.0452
$S^{NC,VRS}$	0.9753	1.0991	1.1952	1.0689
Simar & Zelenyuk Li-test† (C vs. NC)	-0.7275 (CRS)	-0.5523 (CRS)	-0.3706 (CRS)	-0.2365 (CRS)
	(0.9444)	(0.9038)	(0.6458)	(0.5692)
(p-value)	-0.3302 (VRS)	0.4247 (VRS)	-0.6009 (VRS)	-0.7679 (VRS)
	(0.6162)	(0.1822)	(0.8142)	(0.9884)

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

Basic descriptive statistics on pre-merger and post-merger observations solely are reported in Table 4. This table is structured as follows. The columns two and three contain MPI results between the pre-merger and post-merger observations. The columns four and five display HMPI results between the pre-merger and post-merger observations. Horizontally, we report both the CRS and VRS assumptions imposed on a given C or NC technology in the upper part. The lower horizontal part reports the results of Li-test statistic and p-values for MPI and HMPI of pre-merger and post-merger observations between C and NC for given returns to scale.

From Table 4, we can infer the following observations. First, MPI under $S^{C,CRS}$ decreases after the merger, while it improves for $S^{C,VRS}$. Second, MPI under $S^{NC,CRS}$ also decreases after the merger, while it equally improves for $S^{NC,VRS}$. Hence, for both C and NC the MPI results increase after the mergers under the flexible assumption of VRS, and decrease after the mergers for the strong assumption of CRS. Third, for the CRS and VRS specifications both under C and NC, the HMPI results decrease after the mergers.

Finally, the last horizontal part reporting the results of the Li-test statistics confirm that for both MPI and HMPI for pre-merger and post-merger observations under C and NC all experience no significant differences at even the 10% significance level.

Furthermore, we explore the differences between these two productivity indices for a given sample (pre-merger observations or post-merger observations), and the differences in productivity growth between the pre-merger and post-merger observations under a certain index (MPI or HMPI). These specific Li-test statistics results and p-values are shown in Table 5. The first two columns test the differences between these two productivity indices under various specifications for a given sample data, i.e, pre-merger observations or post-merger observations. The last two columns test the difference of these same productivity indices under various specifications between pre-merger and post-merger observations for a certain productivity index (i.e., MPI or HMPI).

Table 5: Li-Test Statistics Results under Various Specifications

		MPI vs. HMPI		Pre-merger vs. Post-merger	
		Pre-merger	Post-merger	MPI	HMPI
Simar & Zelenyuk Li-test† (C vs. NC) p-value	$S^{C,CRS}$	-1.9351 (0.9990)	-0.7909 (0.9984)	4.0398 (0.000***)	3.7367 (0.000***)
	$S^{C,VRS}$	-0.3628 (0.6862)	-0.7817 (0.9932)	3.2661 (0.000***)	2.2834 (0.0106**)
	$S^{NC,CRS}$	-0.6973 (0.9308)	-0.7199 (0.9456)	2.0902 (0.0106**)	1.2945 (0.012*)
	$S^{NC,VRS}$	0.3192 (0.2788)	0.0327 (0.3784)	2.3158 (0.0026**)	1.6442 (0.0272**)

† Li-test: critical values at 1% level=2.33 (***); 5% level=1.64 (**); 10% level=1.28 (*).
p-value is displayed in the parentheses.

Table 5 allows to deduce two conclusions. First, the differences in Li-test statistics between MPI and HMPI results under all specifications for pre-merger observations are identically insignificant. Second, the differences in Li-test statistics between MPI and HMPI results under all specifications for post-merger observations experience the same distribution. Third, the differences in Li-test statistics for MPI and HMPI results between pre-merger versus post-merger observations are always significant under all specifications at least at the 10% significance level.

This approach of mixing up all years before and after the merger (from 76 courts to 31 courts) may lead to various types of asymmetries between the numbers of periods before and after the merger. Therefore, as a robustness check, we also use a symmetric number of periods before and after the merger: in particular, we experiment with (i) a single year, (ii) two years, and (iii) three years before and after the mergers. The relevant results in parallel to Tables 4 and 5 are reported in Appendix C. Apart from some minor details that change, the fundamental conclusions remain intact: for both MPI and HMPI for pre-merger and post-merger observations under C and NC all experience no significant differences at even the 10% significance level.

5 Conclusions

Inspired by Kerstens and Van de Woestyne [36], we perform an empirical analysis on productivity growth of the HM&As in Swedish district courts during the period 2000-2017 using both input-oriented MPI and HMPI under both C and NC. The empirical results allow for the following conclusions. First, based on the calculated average values of MPI and HMPI for various specifications under C and NC, a one-sample symmetric Wilcoxon test and a t-test are conducted to examine whether the average MPI and HMPI are significantly different from unity or not. This one-sample symmetric Wilcoxon test and t-test results indicate that average productivity changes of these two indices are negligible and that no obvious technical change is being generated. This result obtained for Swedish data covering about two decades differs somewhat from the small positive or small negative productivity growth reported for courts in the literature over shorter periods. It certainly contrasts with the long run negative productivity growth reported for the Netherlands in Blank and van Heezik [27].

Turning to the question whether MPI and HMPI are empirically differentiated or not for various specifications under C and NC, one observes similar to O'Donnell [30, p. 258] that both MPI and HMPI are almost equal under $S^{C,CRS}$ in the almost absence of technological change. Moreover, more infeasibility results of MPI occur under $S^{NC,VRS}$ compared to $S^{C,VRS}$. In addition, $S^{NC,VRS}$ produces the most contradictory results between the MPI and HMPI, and yields the most contradictory results between C and NC productivity indices.

Moreover, focusing on the analysis of these two productivity indices between the pre-merger and post-merger observations, we find that for VRS both under C and NC the MPI results increase after the mergers. For the CRS under C and NC the MPI results decrease after the mergers. Furthermore, for the CRS and VRS specifications both under C and NC the HMPI results decrease after the mergers. Last but not least, all results for MPI and HMPI indices are slightly larger than unity after the mergers (but close to unity), indicating that horizontal mergers do not contribute to post-merger productivity gains. This corroborates our earlier results on the absence of technical change.

The effects of uneven productivity growth on the economy as a whole and the health of various industries have been examined in the seminal article by Baumol [69] and further developed in Baumol et al. [70] and Baumol [71]. These authors postulate that industries with below-average productivity growth rates (stagnating industries) will typically see cost increases that are above normal. The ensuing cost disease may cause stagnant industries to see price rises that are above average, quality declines, and financial stress. The drag from stagnant sectors may also result in a decrease in the economy's overall productivity rate and real output growth. Thus, consumers' increasing demand for labor-intensive services, whose productivity growth is inherently constrained, may cause secular stagnation and a decline in real income growth. There have been many critical objections and qualifications to this framework: see Ferris and West [72], Oulton [73], and van der Ploeg [74], among others.

Within this Baumol framework, one question for future research is to see whether the salary growth of Swedish district courts is related to the observed productivity growth. Further open questions for future research are whether these findings on the almost absence of productivity growth in courts are corroborated for different courts in Sweden as well as for different countries in Europe with a similar or different judicial tradition. Especially studies with an even longer length of panel data are being called for. Furthermore, given the rather perplexing results on the almost absence of productivity growth, there is also a need for comparative productivity studies across countries: of course, the latter are hampered by data issues, but one could hope that these problems can be overcome. There is much to be learned from such international comparisons of court productivity. Finally, robust productivity indexes which are less sensitive to atypical and extreme values have been developed for the MPI under the form of, e.g., bootstrapping (e.g., Simar and Wilson [75]) or order-α estimators (e.g., Tzeremes and Tzeremes [76]). However, similar work for the HMPI is unknown to us. Therefore, exploring such robust estimations of MPI and HMPI remains to be done in future work.

References

[1] P. A. Gaughan, Mergers, Acquisitions, and Corporate Restructurings, John Wiley & Sons, 2010.

[2] L. Capron, The long-term performance of horizontal acquisitions, Strategic Management Journal 20 (1999) 987–1018.

[3] P. Belleflamme, M. Peitz, Industrial Organization: Markets and Strategies, Cambridge University Press, Cambridge, 2010.

[4] W. Viscusi, J. Harrington, J. Vernon, Economics of Regulation and Antitrust, 4th ed., MIT Press, Cambridge, 2005.

[5] J. De Loecker, J. Eeckhout, G. Unger, The rise of market power and the macroeconomic implications, Quarterly Journal of Economics 135 (2020) 561–644.

[6] J. Eeckhout, The Profit Paradox: How Thriving Firms Threaten the Future of Work, Princeton University Press, Princeton, 2021.

[7] J. Farrell, C. Shapiro, Horizontal mergers: An equilibrium analysis, American Economic Review 80 (1990) 107–126.

[8] O. Budzinski, I. Ruhmer, Merger simulation in competition policy: A survey, Journal of Competition Law & Economics 6 (2010) 277–319.

[9] R. Caves, In praise of the old I.O., International Journal of Industrial Organization 25 (2007) 1–12.

[10] J. Akhavein, A. Berger, D. Humphrey, The effects of metamergers on efficiency and prices: Evidence from a bank profit function, Review of Industrial Organization 12 (1997) 95–139.

[11] S. Ray, Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research, Cambridge University Press, Cambridge, 2004.

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

[13] P. Bogetoft, D. Wang, Estimating the potential gains from mergers, Journal of Productivity Analysis 23 (2005) 145–171.

[14] T. Kristensen, P. Bogetoft, K. M. Pedersen, Potential gains from hospital mergers in Denmark, Health Care Management Science 13 (2010) 334–345.

[15] P. Mattsson, C. Tidanå, Potential efficiency effects of merging the Swedish district courts, Socio-Economic Planning Sciences 67 (2019) 58–68.

[16] J. D. Cummins, S. Tennyson, M. A. Weiss, Consolidation and efficiency in the US life insurance industry, Journal of Banking & Finance 23 (1999) 325–357.

[17] J. Harris, H. Ozgen, Y. Ozcan, Do mergers enhance the performance of hospital efficiency?, Journal of the Operational Research Society 51 (2000) 801–811.

[18] M. F. Gorman, J. Ruggiero, Evaluating US judicial district prosecutor performance using dea: are disadvantaged counties more inefficient?, European Journal of Law and Economics 27 (2009) 275–283.

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

[20] M. F. Castro, C. Guccio, Measuring potential efficiency gains from mergers of Italian first instance courts through nonparametric model, Public Finance Review 46 (2018) 83–116.

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

[22] E. Monastyrenko, Eco-efficiency outcomes of mergers and acquisitions in the European electricity industry, Energy Policy 107 (2017) 258–277.

[23] P. Arocena, D. S. Saal, T. Urakami, M. Zschille, Measuring and decomposing productivity change in the presence of mergers, European Journal of Operational Research 282 (2020) 319–333.

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

[25] M. Giacalone, E. Nissi, C. Cusatelli, Dynamic efficiency evaluation of Italian judicial system using DEA based Malmquist productivity indexes, Socio-Economic Planning Sciences 72 (2020) 100952.

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

[27] J. Blank, A. van Heezik, Policy reforms and productivity change in the judiciary system: A cost function approach applied to time series of the dutch judiciary system between 1980 and 2016, International Transactions in Operational Research 27 (2020) 2002–2020.

[28] D. W. Caves, L. R. Christensen, W. E. Diewert, The economic theory of index numbers and the measurement of input, output, and productivity, Econometrica 50 (1982) 1393–1414.

[29] W. Briec, K. Kerstens, Infeasibility and directional distance functions with application to the determinateness of the Luenberger productivity indicator, Journal of Optimization Theory and Applications 141 (2009) 55.

[30] C. J. O'Donnell, An aggregate quantity framework for measuring and decomposing productivity change, Journal of Productivity Analysis 38 (2012) 255–272.

[31] H. Bjurek, The Malmquist total factor productivity index, The Scandinavian Journal of Economics (1996) 303–313.

[32] W. Briec, K. Kerstens, The Hicks–Moorsteen productivity index satisfies the determinateness axiom, Manchester School 79 (2011) 765–775.

[33] M. J. Farrell, The convexity assumption in the theory of competitive markets, Journal of Political Economy 67 (1959) 377–391.

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

[35] D. Deprins, L. Simar, H. Tulkens, Measuring labor efficiency in post offices, in: M. Marchand, P. Pestieau, H. Tulkens (Eds.), The Performance of Public Enterprises: Concepts and Measurements, North Holland, Amsterdam, 1984, pp. 243–268.

[36] K. Kerstens, I. Van de Woestyne, Comparing Malmquist and Hicks–Moorsteen productivity indices: Exploring the impact of unbalanced vs. balanced panel data, European Journal of Operational Research 233 (2014) 749–758.

[37] T. Baležentis, K. Kerstens, Z. Shen, 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 forthcoming (2022) forthcoming.

[38] G. Cesaroni, K. Kerstens, I. Van de Woestyne, Global and local scale characteristics in convex and nonconvex nonparametric technologies: A first empirical exploration, European Journal of Operational Research 259 (2017) 576–586.

[39] S. E. Jacobsen, Production correspondences, Econometrica 38 (1970) 754–771.

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

[41] V. Giménez, D. Prior, Long- and short-term cost efficiency frontier evaluation: Evidence from spanish local governments, Fiscal Studies 28 (2007) 121–139.

[42] G. Msann, W. Saad, Assessment of public sector performance in the MENA region: Data envelopment approach, International Review of Public Administration 25 (2020) 1–21.

[43] P. Viton, Technical efficiency in multi-mode bus tansit: A production frontier analysis, Transportation Research B31 (1997) 23–39.

[44] B. Tovar, F. Ramos-Real, E. De Almeida, Efficiency and performance in gas distribution. evidence from brazil, Applied Economics 47 (2015) 5390–5406.

[45] M. Afsharian, H. Ahn, S. Harms, A review of DEA approaches applying a common set of weights: The perspective of centralized management, European Journal of Operational Research 294 (2021) 3–15.

[46] S. Hackman, Production Economics: Integrating the Microeconomic and Engineering Perspectives, Springer, Berlin, 2008.

[47] W. Briec, K. Kerstens, P. Vanden Eeckaut, Non-convex technologies and cost functions: Definitions, duality and nonparametric tests of convexity, Journal of Economics 81 (2004) 155–192.

[48] W. Briec, K. Kerstens, I. Van de Woestyne, Nonconvexity in production and cost functions: An exploratory and selective review, in: S. Ray, R. Chambers, S. Kumbhakar (Eds.), Handbook of Production Economics, volume 2, Springer, Singapore, 2022, pp. 721–754.

[49] X. Chen, K. Kerstens, Q. Zhu, Exploring Horizontal Mergers in Swedish District Courts Using Convex and Nonconvex Technologies: Usefulness of a Conservative Approach, Working paper 2021-EQM-05, IESEG School of Management, Lille, 2021.

[50] R. Färe, S. Grosskopf, M. Norris, Z. Zhang, Productivity growth, technical progress, and efficiency change in industrialized countries, American Economic Review (1994) 66–83.

[51] H. Bjurek, F. Førsund, L. Hjalmarsson, Malmquist productivity indices: An empirical investigation, in: R. Färe, S. Grosskopf, R. Russell (Eds.), Index Numbers: Essays in Honour of Sten Malmquist, Kluwer, Boston, 1998, pp. 217–239.

[52] E. Grifell-Tatjé, C. Lovell, A note on the Malmquist productivity index, Economics Letters 47 (1995) 169–175.

[53] S. Grosskopf, Some remarks on productivity and its decompositions, Journal of Productivity Analysis 20 (2003) 459–474.

[54] M. Nishimizu, J. M. Page, Total factor productivity growth, technological progress and technical efficiency change: dimensions of productivity change in Yugoslavia, 1965-78, Economic Journal 92 (1982) 920–936.

[55] J. L. Zofío, Malmquist productivity index decompositions: A unifying framework, Applied Economics 39 (2007) 2371–2387.

[56] W. Diewert, K. Fox, Decomposing productivity indexes into explanatory factors, European Journal of Operational Research 256 (2017) 275–291.

[57] R. Färe, S. Grosskopf, P. Roos, On two definitions of productivity, Economics Letters 53 (1996) 269–274.

[58] B. Balk, Industrial Price, Quantity, and Productivity Indices : The Micro-Economic Theory and an Application, Springer, New York, 1998.

[59] H. Mizobuchi, Productivity indexes under hicks neutral technical change, Journal of Productivity Analysis 48 (2017) 63–68.

[60] F. Ang, K. Kerstens, J. Sadeghi, Energy productivity and greenhouse gas emission intensity in dutch dairy farms: A Hicks-Moorsteen by-production approach under non-convexity and convexity with equivalence results, Journal of Agricultural Economics 74 (2023) 492–509.

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

[62] G. Falavigna, R. Ippoliti, Institutional efficiency and budget constraints: A Directional Distance Function approach to lead a key policy reform, Technical Report CNR-IRCrES Working Paper 6/2021, CNR-IRCrES, Istituto di Ricerca sulla Crescita Economica Sostenibile, 2021.

[63] X. Chen, K. Kerstens, Evaluating horizontal mergers in Swedish district courts using plant capacity concepts: With a focus on nonconvexity, RAIRO - Operations Research 57 (2023) 219–236.

[64] J. Månsson, C. Andersson, F. Bonander, What lessons can be learned from cost efficiency? the case of Swedish district courts, European Journal of Law and Economics 54 (2022) 1–21.

[65] Q. Li, Nonparametric testing of closeness between two unknown distribution functions, Econometric Reviews 15 (1996) 261–274.

[66] Y. Fan, A. Ullah, On goodness-of-fit tests for weakly dependent processes using kernel method, Journal of Nonparametric Statistics 11 (1999) 337–360.

[67] Q. Li, E. Maasoumi, J. S. Racine, A nonparametric test for equality of distributions with mixed categorical and continuous data, Journal of Econometrics 148 (2009) 186–200.

[68] L. Simar, V. Zelenyuk, On testing equality of distributions of technical efficiency scores, Econometric Reviews 25 (2006) 497–522.

[69] W. Baumol, Macroeconomics of unbalanced growth: The anatomy of urban crisis, American Economic Review 57 (1967) 415–426.

[70] W. Baumol, S. Blackman, E. Wolff, Unbalanced growth revisited: Asymptotic stagnancy and new evidence, American Economic Review 75 (1985) 806–817.

[71] W. Baumol, The Cost Disease: Why Computers Get Cheaper and Health Care Doesn't, Yale University Press, New Haven & London, 2012.

[72] J. Ferris, E. West, The cost disease and government growth: Qualifications to Baumol, Public Choice 89 (1996) 35–52.

[73] N. Oulton, Must the growth rate decline? Baumol's unbalanced growth revisited, Oxford Economic Papers 53 (2007) 605–627.

[74] F. van der Ploeg, Sustainable social spending and stagnant public services: Baumol's cost disease revisited, FinanzArchiv 63 (2007) 519–547.

[75] L. Simar, P. Wilson, Estimating and bootstrapping Malmquist indices, European Journal of Operational Research 115 (1999) 459–471.

[76] P. Tzeremes, N. Tzeremes, Productivity in the hotel industry: An order-α Malmquist productivity indicator, Journal of Hospitality & Tourism Research 45 (2021) 133–150.

Appendices: Supplementary Material

Appendix A Efficiency Models to Compute Productivity

To calculate productivity change through the Malmquist index, we need to compute the four different efficiency measures by solving linear programming models or binary integer programs. Assuming that K is the number of observations, and $\boldsymbol{x}$ and $\boldsymbol{y}$ are input and output vectors. Let $\boldsymbol{z}$ be an activity vector and λ be the objective function, and the input and output values in the period of t and $t+1$ respectively are expressed as $(\boldsymbol{x}^t, \boldsymbol{y}^t)$ and $(\boldsymbol{x}^{t+1}, \boldsymbol{y}^{t+1})$.

In what follows, if we regard the evaluated operating unit as Decision Making Unit p (denoted DMU_p), and the corresponding input and output values in the periods of t and $t+1$ are defined as $(\boldsymbol{x}_p^t, \boldsymbol{y}_p^t)$ and $(\boldsymbol{x}_p^{t+1}, \boldsymbol{y}_p^{t+1})$, then the input-oriented Malmquist productivity index is obtained by solving the following four mathematical programming models:

$$\begin{aligned} E_i^t(\boldsymbol{x}_p^t, \boldsymbol{y}_p^t) &= \min\ \lambda \\ s.t.\ \sum_{k=1}^{K} \alpha\ z_k^t \boldsymbol{x}_{km}^t &\leq \lambda \boldsymbol{x}_p^t, \forall m \\ \sum_{k=1}^{K} \alpha\ z_k^t \boldsymbol{y}_{kn}^t &\geq \boldsymbol{y}_p^t, \forall n \\ \boldsymbol{z} \in \Lambda&, \alpha \in \Gamma. \end{aligned} \tag{A1}$$

$$\begin{aligned} E_i^{t+1}(\boldsymbol{x}_p^{t+1}, \boldsymbol{y}_p^{t+1}) &= \min\ \lambda \\ s.t.\ \sum_{k=1}^{K} \alpha\ z_k^{t+1} \boldsymbol{x}_{km}^{t+1} &\leq \lambda \boldsymbol{x}_p^{t+1}, \forall m \\ \sum_{k=1}^{K} \alpha\ z_k^{t+1} \boldsymbol{y}_{kn}^{t+1} &\geq \boldsymbol{y}_p^{t+1}, \forall n \\ \boldsymbol{z} \in \Lambda&, \alpha \in \Gamma. \end{aligned} \tag{A2}$$

$$
\begin{aligned}
E_i^t(\boldsymbol{x}_p^{t+1}, y_p^{t+1}) &= \min\ \lambda \\
s.t.\ \sum_{k=1}^{K} \alpha\ z_k^t \boldsymbol{x}_{km}^t &\leq \lambda \boldsymbol{x}_p^{t+1}, \forall m \\
\sum_{k=1}^{K} \alpha\ z_k^t \boldsymbol{y}_{kn}^t &\geq \boldsymbol{y}_p^{t+1}, \forall n \\
\boldsymbol{z} \in \Lambda, \alpha &\in \Gamma.
\end{aligned}
\tag{A3}
$$

$$
\begin{aligned}
E_i^{t+1}(\boldsymbol{x}_p^t, \boldsymbol{y}_p^t) &= \min\ \lambda \\
s.t.\ \sum_{k=1}^{K} \alpha\ z_k^{t+1} \boldsymbol{x}_{km}^{t+1} &\leq \lambda \boldsymbol{x}_p^t, \forall m \\
\sum_{k=1}^{K} \alpha\ z_k^{t+1} \boldsymbol{y}_{kn}^{t+1} &\geq \boldsymbol{y}_p^t, \forall n \\
\boldsymbol{z} \in \Lambda, \alpha &\in \Gamma.
\end{aligned}
\tag{A4}
$$

Note that the sets $\Gamma \in \{\Gamma^{CRS}, \Gamma^{VRS}\}$ and $\Lambda \in \{\Lambda^C, \Lambda^{NC}\}$ are as defined in the main text. Moreover, there is a scaling parameter (α) allowing for a particular scaling of all K observations spanning the technology. Finally, the scaling parameter is fixed at unity under variable returns to scale (VRS) and non-negative under constant returns to scale (CRS).

Models (A1) and (A2) are the ones where the technology and the observation under evaluation belong to the same period and therefore the solution value is less than or equal to unity. Models (A3) and (A4) occur when the reference technology is constructed from data in one period while the observation to be evaluated belongs to an adjacent period.

To calculate now the Hicks-Moorsteen productivity index, it is required to determine the values of eight different efficiency measures by solving the corresponding mathematical programs. The four output-oriented efficiency measures are obtained by solving the following four mathematical programming models:

$$
\begin{aligned}
E_o^t(\boldsymbol{x}_p^t, \boldsymbol{y}_p^t) &= \max\ \theta \\
s.t.\ \sum_{k=1}^{K} \alpha\ z_k^t \boldsymbol{x}_{km}^t &\leq \boldsymbol{x}_p^t, \forall m \\
\sum_{k=1}^{K} \alpha\ z_k^t \boldsymbol{y}_{kn}^t &\geq \theta \boldsymbol{y}_p^t, \forall n \\
\boldsymbol{z} \in \Lambda, \alpha &\in \Gamma.
\end{aligned}
\tag{A5}
$$

$$
\begin{aligned}
E_o^t(\boldsymbol{x}_p^t, \boldsymbol{y}_p^{t+1}) &= \max\ \theta \\
s.t.\ \sum_{k=1}^{K} \alpha\ z_k^t \boldsymbol{x}_{km}^t &\leq \boldsymbol{x}_p^t, \forall m \\
\sum_{k=1}^{K} \alpha\ z_k^t \boldsymbol{y}_{kn}^t &\geq \theta \boldsymbol{y}_p^{t+1}, \forall n \\
\boldsymbol{z} \in \Lambda, \alpha &\in \Gamma.
\end{aligned}
\tag{A6}
$$

$$
\begin{aligned}
E_o^{t+1}(\boldsymbol{x}_p^{t+1}, \boldsymbol{y}_p^t) &= \max\ \theta \\
s.t.\ \sum_{k=1}^{K} \alpha\ z_k^{t+1} \boldsymbol{x}_{km}^{t+1} &\leq \boldsymbol{x}_p^{t+1}, \forall m \\
\sum_{k=1}^{K} \alpha\ z_k^{t+1} \boldsymbol{y}_{kn}^{t+1} &\geq \theta \boldsymbol{y}_p^t, \forall n \\
\boldsymbol{z} \in \Lambda, \alpha &\in \Gamma.
\end{aligned}
\tag{A7}
$$

$$
\begin{aligned}
E_o^{t+1}(\boldsymbol{x}_p^{t+1}, \boldsymbol{y}_p^{t+1}) &= \max\ \theta \\
s.t.\ \sum_{k=1}^{K} \alpha\ z_k^{t+1} \boldsymbol{x}_{km}^{t+1} &\leq \boldsymbol{x}_p^{t+1}, \forall m \\
\sum_{k=1}^{K} \alpha\ z_k^{t+1} \boldsymbol{y}_{kn}^{t+1} &\geq \theta \boldsymbol{y}_p^{t+1}, \forall n \\
\boldsymbol{z} \in \Lambda, \alpha &\in \Gamma.
\end{aligned}
\tag{A8}
$$

For the four input-oriented efficiency measures, the models (A1) and (A2) for calculating the MPI are the same as the ones for computing the HMPI. We just need another two mathematical programming models for the input-oriented efficiency measures as follows:

$$\begin{aligned} E_i^{t+1}(\boldsymbol{x}_p^t, \boldsymbol{y}_p^{t+1}) &= \min\ \lambda \\ s.t.\ \sum_{k=1}^{K} \alpha\ z_k^{t+1} \boldsymbol{x}_{km}^{t+1} &\leq \lambda \boldsymbol{x}_p^t, \forall m \\ \sum_{k=1}^{K} \alpha\ z_k^{t+1} \boldsymbol{y}_{kn}^{t+1} &\geq \boldsymbol{y}_p^{t+1}, \forall n \\ \boldsymbol{z} \in \Lambda, \alpha &\in \Gamma. \end{aligned} \tag{A9}$$

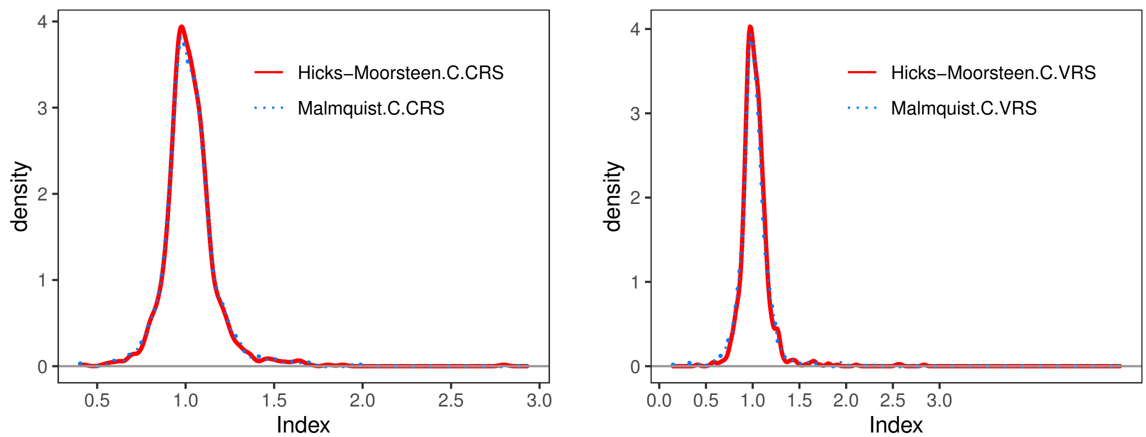
$$\begin{aligned} E_i^{t}(\boldsymbol{x}_p^{t+1}, \boldsymbol{y}_p^{t}) &= \min\ \lambda \\ s.t.\ \sum_{k=1}^{K} \alpha\ z_k^{t} \boldsymbol{x}_{km}^{t} &\leq \lambda \boldsymbol{x}_p^{t+1}, \forall m \\ \sum_{k=1}^{K} \alpha\ z_k^{t} \boldsymbol{y}_{kn}^{t} &\geq \boldsymbol{y}_p^{t}, \forall n \\ \boldsymbol{z} \in \Lambda, \alpha &\in \Gamma. \end{aligned} \tag{A10}$$

Appendix B Kernel Densities

To understand the observed differences in more detail, we now plot the kernel densities of the productivity indices under various frontier specifications for the Swedish district courts. Notice that for comparison purposes, the densities on each figure are estimated using a common bandwidth. These kernel densities are the basis for the Li-test statistics discussed in the main text.

Figure B.1 shows the densities between MPI and HMPI under the $\boldsymbol{S}^{C,\Gamma}$ specification. Figure B.2 shows the densities between MPI and HMPI under the $\boldsymbol{S}^{NC,\Gamma}$ specification. In each case, the CRS specification is displayed in the first sub-figure (a), while the VRS specification is shown in the second sub-figure (b).

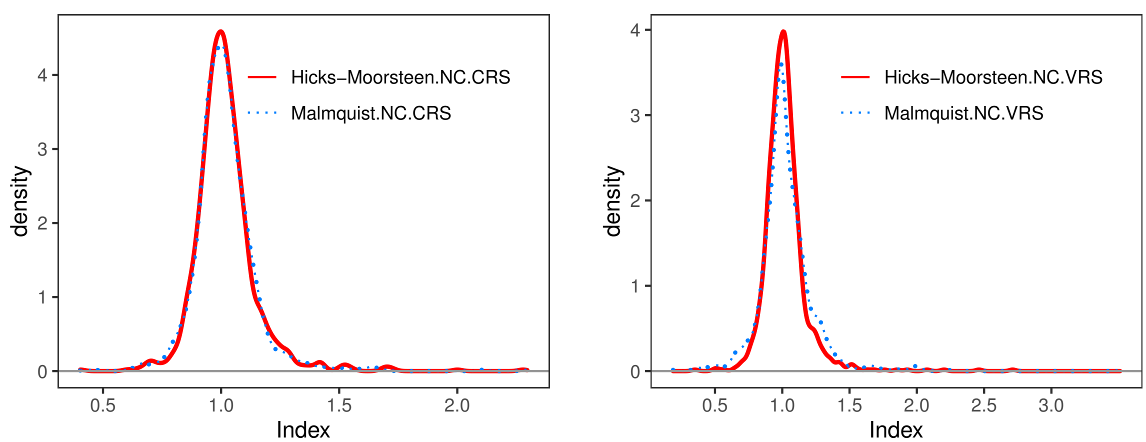
Figure B.3 shows the densities of the Malmquist productivity index between C and NC under the $\boldsymbol{S}^{\Lambda,\Gamma}$ specification. Figure B.4 shows the densities of the Hicks-Moorsteen productivity index between C and NC under the $\boldsymbol{S}^{\Lambda,\Gamma}$ specification. In each case, again the CRS specification is displayed in the first sub-figure (a), while the VRS specification is shown in the second sub-figure (b).



(a) MPI vs. HMPI under $\boldsymbol{S}^{C,CRS}$

(b) MPI vs. HMPI under $\boldsymbol{S}^{C,VRS}$

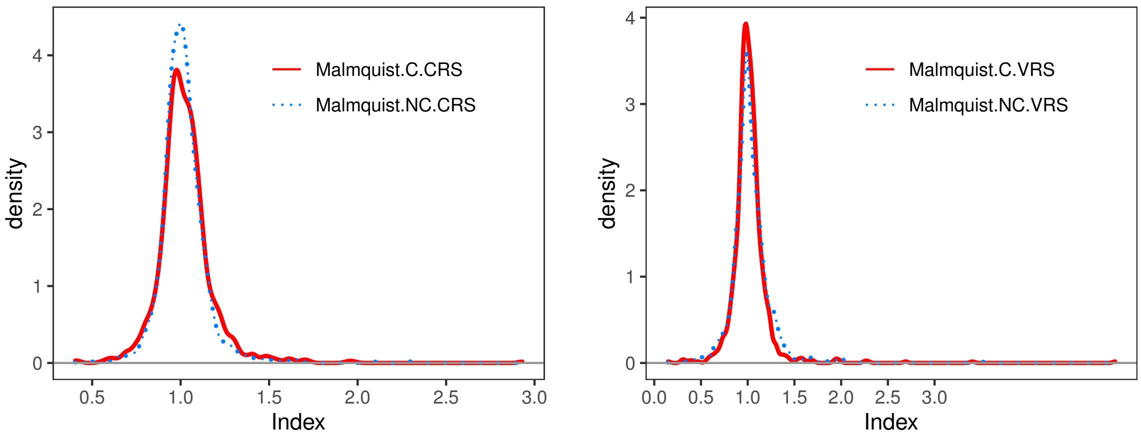
Figure B.1: Kernel densities between MPI and HMPI under the $\boldsymbol{S}^{C,\Gamma}$ specification



(a) MPI vs. HMPI under $\boldsymbol{S}^{NC,CRS}$

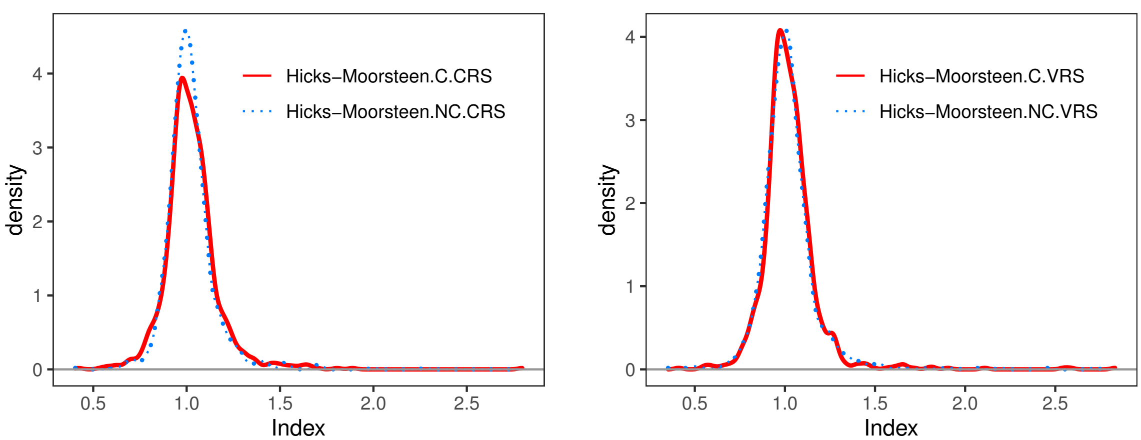
(b) MPI vs. HMPI under $\boldsymbol{S}^{NC,VRS}$

Figure B.2: Kernel densities between MPI and HMPI under the $\boldsymbol{S}^{NC,\Gamma}$ specification



(a) Malmquist under $\boldsymbol{S}^{\Lambda,CRS}$ (b) Malmquist under $\boldsymbol{S}^{\Lambda,VRS}$

Figure B.3: Kernel densities of Malmquist productivity index between C and NC under the $\boldsymbol{S}^{\Lambda,\Gamma}$ specification



(a) Hicks-Moorsteen under $\boldsymbol{S}^{\Lambda,CRS}$ (b) Hicks-Moorsteen under $\boldsymbol{S}^{\Lambda,VRS}$

Figure B.4: Kernel densities of Hicks-Moorsteen productivity index between C and NC under the $\boldsymbol{S}^{\Lambda,\Gamma}$ specification

Appendix C Did Horizontal Mergers Increase Productivity? Results for Different Time Horizons

Notice that the Tables C.1 and C.2 are similar in conclusions to Tables 4 and 5 in the main text. As can be seen from the results in Table C.1, the same conclusions can be obtained according to Table 4. Moreover, in Table C.1, the differences in Li-test statistics between

Table C.1: Productivity Comparisons of the Single Year Before and After the Merger

	MPI		HMPI	
	Pre-merger obs.	Post-merger obs.	Pre-merger obs.	Post-merger obs.
$S^{C,CRS}$	1.1164	1.0654	1.1184	1.0642
$S^{C,VRS}$	1.0068	1.0693	1.1572	1.0635
$S^{NC,CRS}$	1.0958	1.0528	1.0993	1.0442
$S^{NC,VRS}$	0.9812	1.0953	1.1801	1.0700
Simar & Zelenyuk Li-test† (C vs. NC) (p-value)	-0.7128 (CRS) (0.9538)	-0.5182 (CRS) (0.8878)	-0.4369 (CRS) (0.6952)	-0.1218 (CRS) (0.4656)
	-0.3113 (VRS) (0.5998)	0.4227 (VRS) (0.1834)	-0.4369 (VRS) (0.6970)	-0.7593 (VRS) (0.9870)

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

Table C.2: Li-Test Statistics Results under Various Specifications for the Single Year

		MPI vs. HMPI		Pre-merger vs. Post-merger	
		Pre-merger	Post-merger	MPI	HMPI
Simar & Zelenyuk Li-test† (C vs. NC) p-value	$S^{C,CRS}$	-0.1090 (0.9988)	-0.8119 (0.9958)	4.3906 (0.000***)	4.2795 (0.000***)
	$S^{C,VRS}$	-0.2810 (0.6110)	-0.7814 (0.9944)	3.3739 (0.000***)	1.8728 (0.0178**)
	$S^{NC,CRS}$	-0.8048 (0.9622)	-0.6756 (0.9466)	2.5898 (0.0048***)	1.7863 (0.0224**)
	$S^{NC,VRS}$	0.0320 (0.3984)	0.0327 (0.3724)	0.3558 (0.2494)	2.0228 (0.0138**)

† Li-test: critical values at 1% level=2.33 (***); 5% level=1.64 (**); 10% level=1.28 (*).
p-value is displayed in the parentheses.

MPI and HMPI results under all specifications for pre-merger observations are the same distribution, and the same conclusion is reached for the post-merger observations. This is consistent with the results in the first two columns of Table 5. Nevertheless, the difference in Li-test statistics for MPI results between pre-merger versus post-merger observations are always significant under $S^{C,CRS}$, $S^{C,VRS}$ and $S^{NC,CRS}$ specifications at least at the 1% significant level, except under the $S^{NC,VRS}$ specification, which has an identical distribution. Moreover, the difference in Li-test statistics for HMPI results between pre-merger versus post-merger observations are always significant under four specifications ($S^{C,CRS}$, $S^{C,VRS}$, $S^{NC,CRS}$ and $S^{NC,VRS}$) at least at the 5% significant level.

Notice that the Tables C.3 and C.4 are also similar in conclusions to Tables 4 and 5. As can be seen from the results in Table C.3, minor differences exist compared to Table 4. In Table C.3, MPI under $S^{C,CRS}$, $S^{C,VRS}$ and $S^{NC,CRS}$ specifications decrease after merger, while it improves only for the $S^{NC,VRS}$ specification. Moreover, HMPI under four specifications ($S^{C,CRS}$, $S^{C,VRS}$, $S^{NC,CRS}$ and $S^{NC,VRS}$) all decrease after the merger, which is the same result as in Table 5. Furthermore, in Table C.3, the differences in Li-test statistics between MPI and HMPI results under all specifications for pre-merger observations are the same distribution, and the same conclusion is reached for the post-merger observations. This is consistent with the results in the first two columns of Table 5. In addition, the difference

Table C.3: Productivity Comparisons of the Two Years Before and After the Merger

	MPI		HMPI	
	Pre-merger obs.	Post-merger obs.	Pre-merger obs.	Post-merger obs.
$S^{C,CRS}$	1.0927	1.0492	1.0937	1.0477
$S^{C,VRS}$	1.0769	1.0531	1.1200	1.0464
$S^{NC,CRS}$	1.0601	1.0315	1.0808	1.0262
$S^{NC,VRS}$	1.0408	1.0887	1.1124	1.0360
Simar & Zelenyuk Li-test† (C vs. NC) (p-value)	-0.1416 (CRS) (0.4994) -0.6185 (VRS) (0.7524)	-0.8089(CRS) (0.9594) 0.8029 (VRS) (0.1140)	-0.3668 (CRS) (0.6700) -0.9039 (VRS) (0.9520)	-0.7899 (CRS) (0.8296) -1.1051 (VRS) (0.9330)

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

Table C.4: Li-Test Statistics Results under Various Specifications for the Two Years

		MPI vs. HMPI		Pre-merger vs. Post-merger	
		Pre-merger	Post-merger	MPI	HMPI
Simar & Zelenyuk Li-test† (C vs. NC) p-value	$S^{C,CRS}$	-1.0151 (0.9970)	-1.7408 (0.9996)	1.6768 (0.0228**)	3.2552 (0.002***)
	$S^{C,VRS}$	0.9009 (0.1086)	-0.6146 (0.8996)	1.5268 (0.0362*)	1.9354 (0.0120**)
	$S^{NC,CRS}$	-0.3082 (0.6168)	-0.8359 (0.9438)	1.5673 (0.0276*)	2.0414 (0.019**)
	$S^{NC,VRS}$	2.3897 (0.007***)	-0.9706 (0.9764)	1.1521 (0.0702*)	2.3213 (0.0112**)

† Li-test: critical values at 1% level=2.33 (***); 5% level=1.64 (**); 10% level=1.28 (*).
p-value is displayed in the parentheses.

in Li-test statistics for pre-merger observations between MPI versus HMPI are always the same under $S^{C,CRS}$, $S^{C,VRS}$ and $S^{NC,CRS}$ specifications, except that it is significant under the $S^{NC,VRS}$ specification. The differences in Li-test statistics for post-merger observations between MPI versus HMPI are always the same under the four specifications. Finally, the differences in Li-test statistics for MPI and HMPI results between pre-merger versus post-merger observations are always significant under the four specifications ($S^{C,CRS}$, $S^{C,VRS}$, $S^{NC,CRS}$ and $S^{NC,VRS}$) at least at the 10% significant level.

Table C.5: Productivity Comparisons of the Three Years Before and After the Merger

	MPI		HMPI	
	Pre-merger obs.	Post-merger obs.	Pre-merger obs.	Post-merger obs.
$S^{C,CRS}$	1.0856	1.0265	1.0866	1.0254
$S^{C,VRS}$	1.0736	1.0240	1.1080	1.0164
$S^{NC,CRS}$	1.0558	1.0087	1.0707	1.0043
$S^{NC,VRS}$	1.0422	1.0711	1.0981	1.0085
Simar & Zelenyuk Li-test† (C vs. NC) (p-value)	-0.2491 (CRS) (0.5800) -0.4037 (VRS) (0.7384)	1.1261 (CRS) (0.0590) 1.1094 (VRS) (0.3700)	-0.4471 (CRS) (0.6744) -0.5734 (VRS) (0.8238)	0.4586 (CRS) (0.1696) -0.5736 (VRS) (0.8686)

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

Notice that the Tables C.5 and C.6 are also similar in conclusions to Tables 4 and 5.

Table C.6: Li-Test Statistics Results under Various Specifications for the Three Years

		MPI vs. HMPI		Pre-merger vs. Post-merger	
		Pre-merger	Post-merger	MPI	HMPI
Simar & Zelenyuk Li-test[†] (C vs. NC) p-value	$S^{C,CRS}$	-0.8051 (0.9564)	-1.3758 (0.9972)	2.0207 (0.0138**)	2.1596 (0.0134**)
	$S^{C,VRS}$	0.2737 (0.2610)	-0.7007 (0.9486)	2.5061 (0.0054***)	2.0514 (0.0100**)
	$S^{NC,CRS}$	-0.5610 (0.7746)	-0.6152 (0.8802)	3.0899 (0.0026***)	2.4632 (0.0070***)
	$S^{NC,VRS}$	2.3532 (0.008***)	0.0412 (0.3714)	0.3774 (0.2440)	2.5139 (0.0054***)

[†] Li-test: critical values at 1% level=2.33 (***); 5% level=1.64 (**); 10% level=1.28 (*).
p-value is displayed in the parentheses.

As can be seen from the results in Table C.5, minor differences exist compared with Table 4. In Table C.5, MPI under the $S^{C,CRS}$, $S^{C,VRS}$ and $S^{NC,CRS}$ specifications decrease after merger, while it improves only for the $S^{NC,VRS}$ specification. Moreover, HMPI under the four specifications ($S^{C,CRS}$, $S^{C,VRS}$, $S^{NC,CRS}$ and $S^{NC,VRS}$) all decrease after the merger, which is the same as those in Table 5. Furthermore, in Table C.5, the differences in Li-test statistics between MPI and HMPI results under the $S^{C,CRS}$, $S^{C,VRS}$ and $S^{NC,CRS}$ specifications for pre-merger observations are the same distribution, while it is significantly different under the $S^{NC,VRS}$ specification. The differences in Li-test statistics between MPI and HMPI results under all specifications for post-merger observations are the same distribution. This is consistent with the results in the first two columns of Table 5. In addition, the difference in Li-test statistics for pre-merger observations between MPI versus HMPI are always the same under $S^{C,CRS}$, $S^{C,VRS}$ and $S^{NC,CRS}$ specifications, except that it is significant under the $S^{NC,VRS}$ specification. The differences in Li-test statistics for post-merger observations between MPI versus HMPI are always the same under the four specifications. Finally, the differences in Li-test statistics for MPI and HMPI results between pre-merger versus post-merger observations are always significant under the four specifications ($S^{C,CRS}$, $S^{C,VRS}$, $S^{NC,CRS}$ and $S^{NC,VRS}$) at least at the 5% significance level.