European Journal of Operational Research 329 (2026) 1052–1062

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/eor

Innovative Applications of O.R.

An exploratory analysis of learning from peers: Radial vs. nonradial efficiency measures and convex vs. nonconvex technologies☆

Kristiaan Kerstens [a,*], Bart Roets [b,c], Ignace Van de Woestyne [d], Shirong Zhao [e]

[a] Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Économie Management, 3 rue de la Digue, F-59000 Lille, France
[b] Traffic Management & Services Department, Infrabel, Fonsnylaan 13, B-1060 Brussels, Belgium
[c] Faculty of Economics and Business Administration, Ghent University, Sint-Pietersplein 7, B-9000 Ghent, Belgium
[d] KU Leuven, Research Centre for Operations Research and Statistics (ORSTAT), Brussels Campus, Warmoesberg 26, B-1000 Brussels, Belgium
[e] School of Finance, Dongbei University of Finance and Economics, Dalian 116025, Liaoning, China

ARTICLE INFO

Keywords:
Data Envelopment Analysis
Free Disposal Hull
Peer analysis
Technology
Learning

ABSTRACT

This work investigates to which extent the known substantial differences between technical efficiencies on convex and nonconvex technologies translate into different learning possibilities. We also study whether radial and nonradial efficiency measures lead to a different learning experience. To our knowledge, these questions have never been investigated. Our empirical research is guided by three working hypotheses regarding how the analysis of peers facilitates learning by comparing on the one hand convex versus nonconvex technologies, and on the other hand radial versus nonradial efficiency measures. These working hypotheses are investigated using three distinct metrics: peer count, peer similarity, and peer dominance. We employ five existing secondary data sets and one large sample of more than 10,000 observations on Belgian traffic control centres in an effort to refute our three working hypotheses using these three metrics. Anticipating our conclusion, the combination of the logical, the statistical, and the managerial arguments against convexity is rather overwhelming in our data and we think that convexity is an axiom that should be scrutinized in all these three respects in all future methodological innovations as well as in empirical applications.

1. Introduction

The estimation of efficiency has meanwhile a long history in economics and operations research (see Farrell (1957) and Boles (1966) for early contributions and Emrouznejad et al. (2008), Emrouznejad and Yang (2018) for surveys). Almost all if not all traditional parametric, semi-parametric, and nonparametric specifications of technologies maintain the convexity axiom on the technology.

Nonconvexities in technology are known to be possible for a rather wide variety of reasons (see Mas-Colell (1987) for an overview). First, inputs and/or outputs are indivisible. In other words, most inputs and outputs are only imperfectly divisible and cannot be varied continuously as standardly assumed. Second, there is the possibility of non-negligible setup-times that translate into positive setup costs due to indivisibilities in starting up production. Third, increasing returns to scale are possible due to indivisibilities in inputs, learning effects, or organizational advantages in the internal structure of production. Fourth, negative externalities due to interference between productive activities may induce nonconvexities. Furthermore, economies of specialization can also violate the traditional convexity axiom: see, e.g., Romer (1990) on nonrival inputs in the new growth theory. Finally, aggregating distinct convex technologies -in the sense of blueprint books- may yield some local nonconvex range (see Hung et al. (2009)). Nonconvexities complicate the role of prices in both equilibrium and welfare theories (e.g., Mas-Colell (1987)).

In empirical production analysis these reasons for nonconvexities have most often been ignored and the convexity axiom has been maintained because of the -often implicit- assumption of time divisibility (e.g., Shephard (1967, p. 214–215) or Shephard (1970, p. 15)), or simply because of analytical convenience.[1] But, if time is only imperfectly divisible (e.g., because of setup times and the associated costs), then nonconvexities may well affect both production and cost analysis. This is a logical argument questioning convexity.

☆ All authors contributed equally. We acknowledge the most constructive comments of three referees. The usual disclaimer applies.
* Corresponding author.
E-mail addresses: k.kerstens@ieseg.fr (K. Kerstens), bart.roets@infrabel.be, bart.roets@ugent.be (B. Roets), ignace.vandewoestyne@kuleuven.be (I. Van de Woestyne), shironz@163.com (S. Zhao).
[1] Shephard (1967, p. 214–215) and Shephard (1970, p. 15) state that convexity is only valid for "time divisibly-operable technologies".

https://doi.org/10.1016/j.ejor.2025.07.062
Received 14 August 2023; Accepted 29 July 2025
Available online 7 August 2025

This negligence of nonconvexities in production and cost analysis can -implicitly or explicitly- be interpreted in terms of a belief that there is no impact of nonconvexities on the parameters of interest in production and cost approaches alike. Few people will deny that there is no difference between convex and nonconvex production technologies. Briec, Kerstens, and Van de Woestyne (2022, p. 738–740) survey studies documenting differences in convex and nonconvex efficiency decompositions (focusing on technical and scale efficiencies), differences in productivity growth (Malmquist and Hicks-Moorsteen indices on the one hand and Luenberger and Luenberger-Hicks-Moorsteen indicators on the other hand are discussed), and differences in capacity utilization (differences in output-oriented plant capacity, attainable output-oriented plant capacity, input-oriented plant capacity, and economic cost-based capacity notions are listed).

Only rather recently, specific tests for convexity of the production technology have been presented by Kneip et al. (2016) and augmented by Simar and Wilson (2020). Empirical applications of this test are, among others, found in Apon et al. (2015) on US universities, López-Torres et al. (2021) on UK schooling, O'Loughlin and Wilson (2021) on US local governments, and Wilson (2021) on US banks. All of these studies reject convexity, except the first study where for several departments the hypothesis of convexity cannot be rejected. This may provide a statistical argument questioning convexity.

But, we maintain that the eventual impact of convexity on especially cost estimates has rarely if ever been explicitly tested empirically. This neglect is disturbing given that there exists a property of the cost function that is convex/nonconvex in the outputs when convexity of technology is imposed/rejected. Hence, one can only stick to convexity if there is substantial evidence that its impact on the majority of empirical cost function applications is negligible. It is impossible just to assume that the impact of convexity on cost function estimates is negligible since information on how well convex cost functions approximate nonconvex ones is almost absent. Briec, Kerstens, and Van de Woestyne (2022, p. 737–738) list a small selection of studies that report the results of convex and nonconvex frontier cost estimates: these seem to differ anywhere between 2% and 50%. Thus, these cost differences may well be very substantial. This impact of convexity on cost estimates serves to further motivate an interest in investigating the effect of convexity: due to space constraints, this contribution focuses on the effect of convexity on technical efficiency.

The main topic of this research is to investigate to which extent these seemingly substantial differences between technical efficiencies on convex (C) and nonconvex (NC) technologies translate into different peers and learning possibilities. We start out with the conjecture that the traditional radial efficiency measure may behave quite different from its nonradial alternatives. While the radial efficiency measure projects onto the isoquant of technology, the historically first nonradial alternative proposed in the literature, i.e., the Färe and Lovell (1978) efficiency measure, projects onto the efficient subset. To our knowledge, it has never been investigated to which extent radial and nonradial efficiency measures lead to different learning experiences. Learning is critical for frontier applications in management and policy.

To guide our empirical research, we develop three working hypotheses regarding to how the analysis of peers in frontier technologies facilitates learning by comparing C versus NC basic technologies on the one hand, and radial versus nonradial efficiency measures on the other hand. The first working hypothesis states that learning is easier when the number of peers involved is low. The second working hypothesis reads that learning is easier when peer similarity over model variations is high. The final working hypothesis is that learning is easier when more peers dominate the evaluated observation. To empirically investigate these three working hypotheses we develop three metrics that can quantify the empirical impacts. We start our empirical analysis by selecting five existing secondary data sets with a variety of specifications and sample sizes. We supplement these existing data sets with a very large sample of more than 10 000 observations on Belgian traffic control centres: we create subsamples from very small to very big. We try to verify our three working hypotheses on these secondary data sets as well on this original data set using these three metrics. This is a managerial argument questioning convexity.

This contribution is structured as follows. Section 2 defines the production technologies as well as the radial and nonradial efficiency measures employed in our empirical analysis. Section 3 provides a framework for the analysis of peers. It focuses mainly on the formulation of the three distinct working hypotheses that we want to be put to a test using three separate metrics. Section 4 discusses the selection of data sets: existing secondary data sets, and a new large data set. Section 5 focuses on the details of the empirical results. The final Section 6 concludes.

2. Technologies and efficiency measures: Basic definitions

This section introduces some basic notation and defines the production technologies utilized in this contribution. We start from a given production process that turns an N-dimensional input vector $x \in \mathbb{R}^N_+$ into an M-dimensional output vector $y \in \mathbb{R}^M_+$. The production possibility set or production technology T is defined as $T = \{(x, y) \mid x \text{ can produce at least y}\}$. Associated with this technology T, the input set $L(y) = \{x \mid (x, y) \in T\}$ denotes all input vectors x capable of producing at least a given output vector y.[2]

In this contribution, the technology T is assumed to satisfy a combination of the following standard assumptions:

(T.1) Possibility of inaction and no free lunch, i.e., $(0, 0) \in T$ and if $(0, y) \in T$, then $y = 0$.
(T.2) T is a closed subset of $\mathbb{R}^N_+ \times \mathbb{R}^M_+$.
(T.3) Strong input and output disposability, i.e., if $(x, y) \in T$ and $(x', y') \in \mathbb{R}^N_+ \times \mathbb{R}^M_+$, then $(x', -y') \geq (x, -y) \Rightarrow (x', y') \in T$.
(T.4) T is convex.

We comment briefly on these traditional assumptions by recalling the following elements (see, e.g., Färe et al. (1994) or Hackman (2008) for details). Inaction is feasible, and there is no free lunch. Technology is closed. We assume strong or free disposability of inputs and outputs: inputs can be wasted, and outputs can be discarded without any opportunity costs. Finally, technology is C. In our empirical analysis not all these axioms are maintained simultaneously.[3] In particular, C is not always maintained in the empirical analysis.

When discussing input-oriented efficiency measures, in our context it is important to distinguish between two subsets of the input set. First, we can define the isoquant of an input set as:

$$Isoq L(y) = \{x \in L(y) \mid \forall \lambda \in [0, 1) : \lambda x \notin L(y)\}. \quad (1)$$

Finally, the strongly efficient subset of the input set is defined as:

$$Eff L(y) = \{x \in L(y) \mid \forall u \in \mathbb{R}^N_+ : u \leq x \text{ and } u \neq x \Rightarrow u \notin L(y)\}. \quad (2)$$

Obviously, these two subsets of the input set are embedded: $Eff L(y) \subseteq Isoq L(y) \subseteq L(y)$.

The radial input efficiency measure characterizes the input set $L(y)$ completely. The formal definition of the Debreu (1951) and Farrell

[2] It is common to assume that the input and output data satisfy a series of conditions (see, e.g., Färe et al. (1994, p. 44–45)): (i) each producer employs non-negative amounts of each input to produce non-negative amounts of each output; (ii) there is an aggregate production of positive amounts of every output as well as an aggregate utilization of positive amounts of every input; and (iii) each producer employs a positive amount of at least one input to produce a positive amount of at least one output.

[3] For example, note that the C variable returns to scale technology need not satisfy inaction.

(1957) radial input-oriented efficiency measure is as follows:

$$DF : \mathbb{R}_+^N \times \mathbb{R}_+^M \to \mathbb{R}_+ \cup \{+\infty\} : (x, y) \mapsto$$
$$DF(x, y) = \inf_{\delta \in \mathbb{R}_+} \{\delta \mid \delta x \in L(y)\}. \tag{3}$$

This radial input efficiency measure indicates the maximal equiproportionate reduction in all inputs which still allows production of the given output vector. This radial input efficiency measure has the main property that it is smaller than or equal to unity for all feasible input–output combinations ($\forall x \in L(y) : DF(x, y) \leq 1$), with efficient production on the isoquant of $L(y)$ represented by unity. Furthermore, the radial input efficiency measure has a cost interpretation (see, e.g., Hackman (2008)).

The nonradial Färe and Lovell (1978) input efficiency measure can be defined as follows:

$$FL : \mathbb{R}_+^N \setminus \{0\} \times \mathbb{R}_+^M \to \mathbb{R}_+ \cup \{+\infty\} : (x, y) \mapsto$$
$$FL(x, y) = \inf_{\beta \in \mathbb{R}_+^N} \left\{ \frac{1}{n(I(x))} \sum_{i \in I(x)} \beta_i \middle| \beta \odot x \in L(y), \beta_i \in [0, 1] \right\}, \tag{4}$$

where $\odot$ denotes the Hadamard (element by element) product of two vectors, and for all $x \in \mathbb{R}_+^N$ the support of x is defined as $I(x) = \{i \in \{1, \ldots, N\} \mid x_i > 0\}$ and $n(I(x))$ denotes the cardinality of the set $I(x)$. This Färe and Lovell (1978) input efficiency measure indicates the minimum average sum of dimension-wise reductions in each input dimension which maintains production of given outputs on the efficient subset of the input set.[4]

In our static production context the above input efficiency measures are always well-defined. More details on the axiomatic properties of both input-oriented efficiency measures is found in Russell and Schworm (2009). It suffices to point out that the radial input efficiency measure does not comply with the indication property of the efficient subset, while the Färe and Lovell (1978) input efficiency measure does. Both input efficiency measures comply with another and perhaps more fundamental property of independence of units of measurement (for generalized commensurability: see Briec, Dumas, et al. (2022)).

It can be noted that we have limited the discussion to some of the most popular input-oriented efficiency measures, but it is also possible to define more general graph efficiency measures that operate in the full space of inputs and outputs (see Russell and Schworm (2011) for a survey). One example is the directional distance function, a generalized efficiency measure that is compatible with reductions in inputs and expansions in outputs and that has a normalized profit interpretation.

For the empirical application in Section 4, we assume a nonparametric frontier technology under the flexible or variable returns to scale assumption (VRS) and impose either C or NC.[5] Based on K observations consisting of input–output combinations $(x_k, y_k) \in \mathbb{R}_+^N \times \mathbb{R}_+^M$, $(k = 1, \ldots, K)$, a unified algebraic representation of C and NC nonparametric frontier estimators of the technologies under the flexible or variable returns to scale assumption is possible as follows:

$$T^{\Lambda, VRS} = \left\{ (x, y) \mid x \geq \sum_{k=1}^K z_k x_k, y \leq \sum_{k=1}^K z_k y_k, z \in \Lambda \right\}, \tag{5}$$

where

(i) $\Lambda \equiv \Lambda^{\mathrm{C}} = \left\{ z \mid \sum_{k=1}^K z_k = 1 \text{ and } z_k \geq 0 \right\}$;

(ii) $\Lambda \equiv \Lambda^{\mathrm{NC}} = \left\{ z \mid \sum_{k=1}^K z_k = 1 \text{ and } z_k \in \{0, 1\} \right\}$.

The activity vector z of real numbers summing to unity represents the C axiom. This same sum constraint with each vector component being a binary integer represents NC. The estimator of the C technology satisfies axioms (T.1) (except inaction) to (T.4), while the estimator of the NC technology adheres to axioms (T.1) to (T.3).

It is now straightforward to consider the input efficiency estimates, obtained by plugging the nonparametric frontier estimators of the technologies determined by (5) into the input efficiency measures (3) and (4). Both input efficiency estimates of (3) and (4) involve solving a simple linear program (e.g., Ferrier et al. (1994, Appendix A)) in the C case, and an implicit enumeration algorithms (e.g., Briec, Kerstens, and Van de Woestyne (2022)) in the NC case: see also Appendix B for further details. Let z_k^* denote the optimal solution for the activity vector components from these mathematical programming problems, then the reference unit is a weighted average of the existing units $(\sum_{k=1}^K z_k^* x_k, \sum_{k=1}^K z_k^* y_k)$ under C, and it is (x_k, y_k) corresponding to the sole non-zero optimal activity component $z_k^* = 1$ under NC.

3. Managerial framework for the analysis of peers

3.1. Literature review: A selection on peers and learning

Radial and nonradial efficiency measures have been compared to one another, but rarely on C and NC technologies simultaneously. Silva Portela et al. (2003) is such a study, but the authors use non-oriented efficiency measures that modify inputs and outputs simultaneously, and do not analyse the underlying peers. Ferrier et al. (1994) and De Borger et al. (1998) analyse input-oriented radial and nonradial (including the Färe and Lovell (1978)) efficiency measures on the same banking data set using C and NC technologies, respectively, but these articles do not compare C and NC results and ignore the analysis of the underlying peers. Thus, it seems that our empirical analysis is the first to compare input-oriented radial and Färe and Lovell (1978) efficiency measures on C and NC technologies and assess their underlying peers.

Månsson (2003) is the first to mitigate the impact of C on the selection of peers by suggesting to compute a sphere with minimal Euclidean distance between an evaluated observation and the set of efficient observations in the sample. This method is disconnected of the use of traditional efficiency measures, and the author does not further analyse the resulting peers using C and NC technologies and using radial and nonradial efficiency measures. Chavas and Kim (2015) are the first to create a technology that combines in a sense endogenously both the C and the NC technologies by defining a mixture technology relative to a neighbourhood of firms within a given distance of one another. But, these authors ignore the analysis of the underlying peers as well.

Ruiz and Sirvent (2022a) develop some C benchmarking models that seek to find the closest targets (projections onto the production frontier), while at the same time identifying reference sets consisting of peers (optimal activity variables) having the most similar performances to that of the unit under evaluation. Ruiz and Sirvent (2022b) propose further C benchmarking models that identify peers serving as benchmarks showing the way to achieve the targets that have been set. Thus, targets and peers set the same direction for improvement determined by the targets, while the peers identified are real benchmarks showing the way towards those targets.

Krüger (2018) restricts the identification of peers to the efficient subset of a basic NC technology. In particular, he identifies the strongly efficient subset of a NC technology and then computes optimal distances towards these points using a minimal or maximal directional distance function. In line with some earlier work, Ghahraman and Prior (2016) draw on social network analysis to transform the benchmarking information from C efficiency analysis into a network of possible efficiency improvements. These authors calculate optimal stepwise benchmarking paths, detect possible outliers, cluster units, and highlight specialized decision making units. Daraio and Simar (2016)

[4] Ruggiero and Bretschneider (1998) and Zhu (1996), for instance, define a weighted Färe and Lovell (1978) input efficiency measure.

[5] It cannot be excluded that the distinction between isoquant and efficient subset may also be relevant for parametric frontier analysis. We focus on nonparametric frontier technologies where this distinction has been the subject of substantial research as cited in this contribution.

propose a data-driven approach based on nonparametric local constant regression to objectively select a direction vector for the directional distance function. A problem of eventually abandoning a proportional version of the direction vector is that generalized commensurability is no longer respected (see Briec, Dumas, et al. (2022)).

But, in all these contributions these peers are not further analysed in terms of similarity between C and NC technologies and between radial and nonradial efficiency measures.

3.2. Managerial concerns

Decision makers seem to have difficulties understanding the efficiency results under C. This is evidenced in remarks, scattered in the literature, on the problems encountered in communicating the results of traditional efficiency measurement assuming C to decision makers. We provide some examples of quotes reflecting this doubt of managers on the axiom of C.

In a study applying C nonparametric frontier methods to measure bank branch efficiency, Parkan (1987, p. 242) notes:

> The comparison of a branch which was declared relatively efficient, to a hypothetical composite branch, did not allow for convincing practical arguments as to where the inefficiencies lay.

Epstein and Henderson (1989, p. 105) report similar experiences in that managers simply question the feasibility of the hypothetical projection points resulting from C nonparametric frontiers when discussing an application to a large public-sector organization:

> The algorithm for construction of the frontier was also discussed. The frontier segment connecting A and B was considered unattainable. It was suggested that either (1) these two DMUs (Decision Making Units, red.) should be viewed as abnormal and dropped from the model, (2) certain key variables have been excluded, or (3) the assumption of linearity was inappropriate in this organization. It appears that each of these factors was present to some degree.

In a very similar vein, Bouhnik et al. (2001, p. 243) state:

> Equally as important, it is our experience that managers often question the meaning of C combinations that involve what they perceive to be irrelevant DMUs.

These quotes point to the fact that C may well in practice combine observations that are too far apart in terms of input mix, output mix, and/or scale of operations. While one hopes for a rather uniformly dense rather well-spaced cloud of points that avoids the combination of extreme points of production, such extreme combinations apparently occur and are puzzling for managers.

In addition, some researchers admit that NC analysis of production facilitates the practical use of efficiency results. For instance, Bogetoft et al. (2000, p. 859) declare in this context:

> In general, allowing the possibility set to be nonconvex facilitates the practical use of productivity analysis in benchmarking. In particular, fictitious production possibilities, generated as convex combinations of those actually observed, are usually less convincing as benchmarks, or reference units, than actually observed production possibilities.

This experience is confirmed by Halme et al. (2014, p. 10):

> During our long experience of DEA applications we repeatedly encountered the phenomenon that DMs (Decision Maker) are reluctant to evaluate other than existing units.

While these quotes can be interpreted as questioning the practical usefulness of C, we are in favour of a more benign interpretation. These quotes reveal that managers have difficulties learning from peers when these do not dominate the DMU being evaluated, while they are capable to learn from peers when these do dominate the DMU being evaluated.

3.3. Comparison framework

Our empirical learning from peers analysis is based on the following three working hypotheses:

1 WH_1: Learning is facilitated when **peer count** is low.
2 WH_2: Learning is facilitated when **peer similarity** over model variations is high.
3 WH_3: Learning is facilitated when **peer dominance** is high.

It is important to highlight that these three working hypotheses are not directly subjected to statistical tests. Instead, these working hypotheses serve as a preliminary framework (expectation) that provides a starting point for further research. The working hypothesis is often associated with deductive and exploratory research goals within empirical investigations, frequently serving as a foundational framework in qualitative and quantitative research endeavors (see, e.g., Casula et al. (2021)).

The first working hypothesis WH_1 indicates that learning is simplest when the number of peers is low. Imagining that an inefficient observation learns via either visiting their peers or via some meetings regrouping their peers, it is intuitively clear that the time cost and complexity of such learning process is simplest when the number of peers is low. For instance, Krüger (2018) in fact restricts peers for a C technology to be selected among the unique units in the NC efficient subset: just as in the NC technology, this always leads to a single peer. This can be interpreted as an attempt to keep the number of peers low to facilitate learning.

The second working hypothesis WH_2 hopes for some robustness of the peers across model variations. The more peers are similar across model variations the easier it is to learn from peers and to implement changes that hopefully ameliorate performance. This idea is somewhat similar to articles investigating peer similarity across different efficiency measures (e.g., De Borger et al. (1998), Ferrier et al. (1994) or Silva Portela et al. (2003)).

Finally, the third working hypothesis WH_3 distinguishes between peers that are dominating the evaluated observation and those that do not dominate the evaluated observation: it is simply stated that learning is easier when the relative amount of peers dominating the evaluated observation is high. The above quotes in Section 3.2 seem to support this hypothesis. Indeed, a benign interpretation is that these quotes do not as such question C, but reveal that managers have difficulties learning from peers when these do not dominate the DMU being evaluated, while they are capable to learn from peers when these do dominate the DMU being evaluated. Månsson (2003) uses vector dominance to assess peers in a C frontier model, but does not formulate any hypothesis as to its impact on learning.

To systematically and coherently analyse how learning from radial and nonradial measures and learning from C and NC technologies can differ, we operationalize our hypotheses through the comparison framework depicted in Fig. 1. We perform four calculations of efficiencies and efficient peers, which combine the radial/nonradial and C/NC approaches: Convex-Radial (denoted as C&DF and Ⓐ throughout this paper), Convex-Nonradial (C&FL and Ⓑ), Nonconvex-Radial (NC&DF and Ⓒ), and Nonconvex-Nonradial (NC&FL and Ⓓ). We examine the impact of radial versus nonradial measures in comparisons 1 and 2, and the impact of C versus NC in comparisons 3 and 4.

To evaluate our hypothesis WH_1, we simply perform a **peer count**. For each observation $j_0 \in \{1, \ldots, K\}$ under evaluation, the set of efficient peer units is defined as $EP^{T\&E}(j_0) = \{k \in \{1, \ldots, K\} \mid z_{k,j_0} > 0\}$, with z_{k,j_0} representing the activity (intensity) variables for observation j_0, determined assuming technology T and applying efficiency measure E. The peer count for observation j_0 is then the number of elements (or cardinal) in the set of peers, or $n(EP^{T\&E}(j_0))$. The cardinality of this set $n(EP^{T\&E}(j_0))$ is at least unity under C and normally unity under NC. This set of efficient peer units must be distinguished from the reference

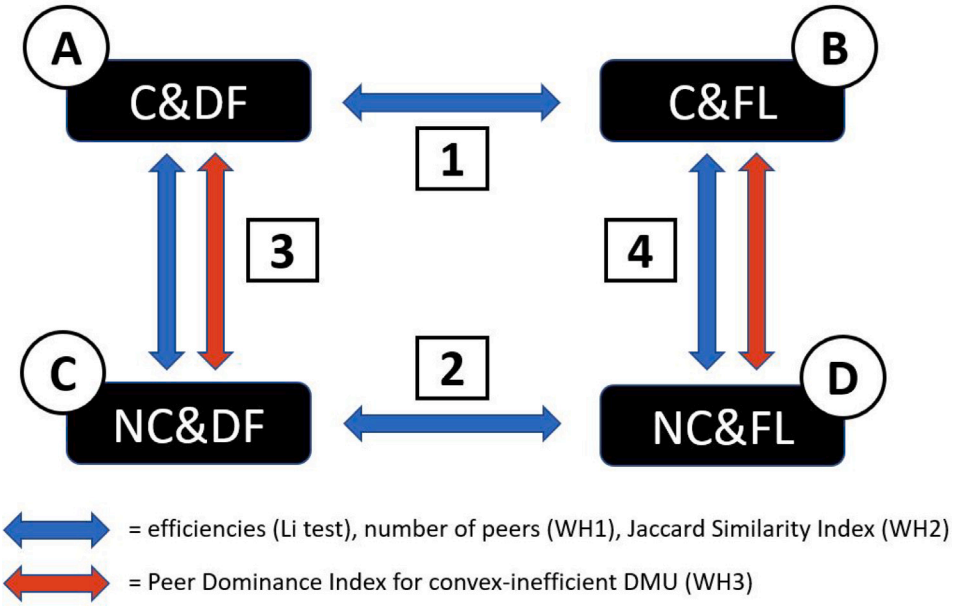

Fig. 1. Comparison framework.

unit defined at the end of Section 2. We develop a numerical example of the peer count based on one of our data sets in Appendix C.

Peers have certainly been analysed in the Data Envelopment Analysis (DEA) literature, but to our knowledge little is known about the amount of peers under radial efficiency measures projecting on the isoquant versus nonradial efficiency measures projecting on the efficient subset. Also, while counting may seem trivial for FDH (there is always one peer), we explore systematically the potential non-unique optimal solutions of the underlying implicit enumeration problems (see Appendix B). Indeed, it is not excluded that multiple dominating observations with an identical efficiency measure can be identified. This search for alternative optima under NC has to the best of our knowledge never been reported before.

We apply two similarity metrics to support the analysis of our hypotheses WH_2 and WH_3. To examine **peer similarity** we apply the Jaccard Similarity Index to quantify the overlap between the efficient peers identified by two different approaches. The Jaccard Index or Jaccard Similarity Index (JSI) is an intuitively simple yet powerful similarity metric, commonly used in domains such as biostatistics, fraud detection, and image recognition to assess the overlap between two sample sets. It is calculated by dividing the number of common peers (i.e., peers identified by the two approaches) by the total number of peers (i.e., peers identified by at least one approach). As such, it gauges the amount of overlap between two sets of efficient peers. More formally, the Jaccard Similarity Index for observation j_0, comparing approach with technology 1 and efficiency measure 1 ($T_1\&E_1$) with approach with technology 2 and efficiency measure 2 ($T_2\&E_2$) in terms of peer similarity, is expressed as:

$$JSI(j_0) = 100 \cdot \frac{n(EP^{T_1\&E_1}(j_0) \cap EP^{T_2\&E_2}(j_0))}{n(EP^{T_1\&E_1}(j_0) \cup EP^{T_2\&E_2}(j_0))} \tag{6}$$

where in the numerator we have the intersection of the cardinality of the peer set $n(EP^{T_1\&E_1}(j_0))$ with technology 1 and efficiency measure 1 and the cardinality of the peer set $n(EP^{T_2\&E_2}(j_0))$ with technology 2 and efficiency measure 2, and in the denominator we have their union. Obviously, the JSI is not the only resemblance measure available in the literature (see, e.g., Batagelj and Bren (1995) for an early review), but it is among the oldest and most well-known such indices. A numerical example of the JSI index based on one of our data sets is found in Appendix C.

Finally, for the C-NC comparison we complement the JSI with an additional, purpose-built similarity metric that focuses on **peer dominance**: the Peer Dominance Index (PDI). We define the Peer Dominance Index as the number of vector-dominating C DEA peers, divided by the number of C DEA peers. It is only calculated for non-efficient observations, or in other words for observations with learning potential. More formally, with technology T being C, the Peer Dominance Index for observation j_0 is defined as:

$$PDI(j_0) = 100 \cdot \frac{n(\{p \in EP^{C\&E}(j_0) \mid x_p \leq x_{j_0}, y_p \geq y_{j_0}\})}{n(EP^{C\&E}(j_0))} \tag{7}$$

with $(x_{j_0}, y_{j_0}) \in \mathbb{R}_+^N \times \mathbb{R}_+^M$ representing the input–output combination for observation j_0, and $(x_p, y_p) \in \mathbb{R}_+^N \times \mathbb{R}_+^M$ the input–output combinations for its peers. In the numerator, the subset of $EP^{C\&E}(j_0)$ contains the peers that vector-dominate the evaluated observation j_0 in the C technology with efficiency measure E: they produce at least as much of each output, with no more of any input.[6] The denominator $n(EP^{C\&E}(j_0))$ denotes simply the cardinality of the peer set in the C technology with efficiency measure E. For a numerical example of the PDI index based on one of our data sets, one can consult Appendix C.

The PDI conceptually broadens the JSI perspective and provides additional information for the C-NC comparison. Although the PDI does not explicitly compare DEA with FDH approaches, it does consider the (NC) dominance aspect of the peers identified by DEA, and therefore we position it in Fig. 1 next to the DEA-FDH comparison. Both the JSI and the PDI vary between 0 and 100 (percent), with higher values indicating a higher similarity.

We expect these three metrics to vary depending on the approach (Convex vs Non-Convex, Radial vs Nonradial). However, before applying real datasets and conducting statistical tests, the full scope of differences and similarities between these approaches remains unclear. Our study provides an exploratory analysis of learning from peers by thoroughly examining these three metrics across different approaches.

4. Selection of data sets

We perform our empirical analysis on both five existing (published) data sets and one new (unpublished) data set. Ordered according to

[6] The vector inequality convention is as follows: $a \leq b$ if and only if $a_i \leq b_i$ for all i.

Table 1
Sources of empirical data.

Article	Size (K)	# Inp. (N)	# Outp. (M)	Sector	Remarks
Färe et al. (1985)	32	3	1	Electricity	
Haag et al. (1992)	41	4	2	Agriculture	
Färe et al. (1983)	86	3	1	Electricity	Unbalanced (N=20 & T=5)
Cesaroni et al. (2011)	92	2	5	Car registration	
Fan et al. (1996)	471	3	1	Agriculture	
Traffic control centre (TCC)	10 198	2	3	Railways	

Table 2
Input and output variables for the traffic control centre (TCC) Models.

Inputs	
OPER	Number operators present during this specific hour
SURV	Numbers surveillance staff present during this specific hour
Outputs	
MOVE	Number of train movements
ADAPT	Traffic interventions by TCC staff (weighted by time needed)
SAFETY	Safety interventions by TCC staff (weighted by time needed)

Table 3
Average efficiencies.

Data set	K	C&DF Ⓐ	C&FL Ⓑ	NC&DF Ⓒ	NC&FL Ⓓ
Färe et al. (1985)	32	0.951	0.824	0.998	0.982
Haag et al. (1992)	41	0.880	0.715	1.000	1.000
Färe et al (1983)	86	0.929	0.831	0.986	0.952
Cesaroni et al. (2011)	92	0.729	0.645	0.911	0.858
Fan et al. (1996)	471	0.811	0.752	0.913	0.823
Average		**0.860**	**0.753**	**0.962**	**0.923**
TCC subsample 8	40	0.780	0.736	0.901	0.836
TCC subsample 7	80	0.779	0.738	0.909	0.835
TCC subsample 6	160	0.738	0.675	0.907	0.826
TCC subsample 5	319	0.745	0.680	0.885	0.801
TCC subsample 4	638	0.648	0.629	0.875	0.778
TCC subsample 3	1275	0.612	0.597	0.844	0.736
TCC subsample 2	2550	0.573	0.538	0.834	0.721
TCC subsample 1	5099	0.555	0.516	0.810	0.699
TCC full sample	10198	0.528	0.493	0.765	0.650
Average		**0.662**	**0.622**	**0.859**	**0.765**

ascending sample size, Table 1 presents key features of the six data sets employed in this study. This same ordering is maintained in other tables.

We first briefly discuss the five existing studies (earlier used in Cesaroni et al. (2017)). One article contains a small unbalanced panel (Färe et al. (1983)) and four articles use cross section data: Cesaroni (2011), Fan et al. (1996), Färe et al. (1985), and Haag et al. (1992). Key points to note are the following: (i) there are three single output articles, and two multiple-output articles; and (ii) the sample sizes vary from quite small to rather big. Finally, note that the time dimension in the panel data set is ignored: this assumes the absence of technical change over the five time periods.

The new empirical data are obtained from Belgium's national railway infrastructure operator Infrabel. In particular, data is gathered from Infrabel's staff scheduling tool and the real-time signalling and safety systems. The resulting large-scale data set captures the resources and activities of 11 Belgian traffic control centres (TCC), for each individual hour of the first 6 months of 2016. We first generated a random subsample of 11 772 DMUs, and then eliminated the observations that simultaneously exhibit zero outputs in all dimensions. The final dataset contains 10 198 DMUs.

Notice that from a statistical perspective the first four data sets displayed in Table 1 are very small given the slow convergence rates of C and NC technical efficiency estimators, which are $K^{-2/(N+M+1)}$ and $K^{-1/(N+M)}$, respectively: see Kneip et al. (2015) for more details. Therefore, one should take care when interpreting the results of these smallest data sets.

The model specification is the same as in Roets et al. (2018) and is displayed in Table 2. The input-oriented calculations are based on two inputs, capturing the number of operators (OPER) and the number of surveillance staff (SURV) present in the TCC during the hour in question. The three outputs gauge the activity in the TCC during this hour. The output MOVE quantifies the traffic volumes and therefore the traffic monitoring effort by the TCC, the second output ADAPT captures the non-safety critical interventions by the TCC staff (e.g., modifying train priorities or train itineraries), and the final output SAFETY measures the amount of safety critical interventions (e.g., protecting track workers in case of maintenance works). For more details on the model specifications we refer to Roets et al. (2018). For additional insights into the railway traffic control process we refer to Topcu et al. (2019).

The use of this TCC data set provides three particular advantages for our empirical analysis. First, the data is used intra-company to support decision-making, and given its real-world nature it is therefore particularly relevant for our research on "learning from peers". Second, its high level of temporal disaggregation produces inputs of a discrete nature (OPER and SURV), which should allow to identify multiple peers even under NC. Third, its large scale allows to specifically examine the impact of sample size. Starting from the original large-scale TCC data set, we sequentially generate random subsets, consecutively consisting of half the sample of the previous data set. In doing so, we generate eight subsets of data, containing 5099, 2550, 1275, 638, 319, 160, 80 and 40 DMUs, respectively. Note that the lower end of this range of sample sizes overlaps with the sample size range of the five existing publicly available data sets under investigation.

5. Empirical results

5.1. Efficiency distributions

Before reporting and analysing the metrics related to our three learning from peers working hypotheses (operationalized by the number of peers, the Jaccard Similarity Index, and the Peer Dominance Index), we take a closer look at the similarities and dissimilarities between the four different efficiency results. Table 3 displays the average efficiency scores for the five public data sets and the nine TCC data sets, as well as the average of these averages. The last four column headers of this table correspond with the labels used in Fig. 1. We can draw the following conclusions. First, for both the public and the TCC data sets, the radial efficiency scores (DF) are higher than the nonradial scores (FL), and the NC scores exhibit a higher efficiency than the C scores. Second, as expected, efficiency levels have a tendency to decrease as sample size K increases. Third, for one small data set (in Haag et al. (1992)), there are no NC-inefficient units reported. Such a situation is sometimes interpreted as a lack of discriminatory power of the NC technology. But, another way to look at the matter is to realize that the 12% to 28.5% inefficiencies found under the C technology depend entirely on the validity of the C hypothesis (see our discussion supra).

For a formal assessment of this difference in empirical densities, we employ a nonparametric test initially proposed by Li (1996). This test has been refined by Fan and Ullah (1999) and others: the most recent development is by Li et al. (2009). This nonparametric test analyses

the differences between entire distributions instead of focusing on, for instance, first moments (as, e.g., the Wilcoxon signed-ranks test). It tests the statistical significance of differences between two kernel-based estimates of density functions, f and g, of a random variable x. The null hypothesis states the equality of both density functions almost everywhere ($H_0 : f(x) = g(x)$ for all x). The alternative hypothesis negates the equality of both density functions ($H_1 : f(x) \neq g(x)$ for some x). This test is valid for both dependent and independent variables: observe that dependency is a characteristic of frontier estimators (i.e., efficiency levels depend on sample size, among others).

To avoid the issue of spurious mass at the boundary of the technical efficiency measures, Simar and Zelenyuk (2006) offer a further refinement of this Li-test statistic for nonparametric frontier estimators. Their Algorithm II adds uniform noise with one order of magnitude less than the noise added by the specific estimator to smooth the boundary estimates, while their Algorithm I ignores the boundary estimates. Their Monte Carlo simulations suggest that algorithm II performs somewhat better overall, even if the strength of the test statistic decreases with increasing dimension in the production specification. In summary, we employ the Li et al. (2009) version of this test, which has been adjusted by the application of Simar and Zelenyuk (2006) Algorithm II.[7]

The results of the Li test adjusted by Simar and Zelenyuk (2006) are displayed in Table 4 (with the p-values and significance levels in brackets underneath). The last four column headers of the table correspond with the labels used in Fig. 1. We first consider the radial versus nonradial comparison (see C&DF versus C&FL in column 1, and NC&DF versus NC&FL in column 2). Afterwards, we examine the C versus NC differences.

For the public datasets, all indicate significantly different distributions for radial versus nonradial measures, except for the NC efficiencies (column 2) in the Haag et al. (1992) data set. This result is confirmed by the TCC data sets: both for the C and NC efficiencies, the distributions become significantly different at sample sizes comparable to the pattern exhibited by the public data sets, except for the C efficiencies (column 1) in TCC subsample 8. The C versus NC comparison (C&DF versus NC&DF in column 3, and C&FL versus NC&FL in column 4) paints a different picture. Almost all efficiency distributions are significantly different, except for a few small TCC subsamples. The null hypothesis of equal efficiency distributions for C&DF versus NC&DF cannot be rejected in TCC subsamples 5, 6, and 8. The null hypothesis of equal efficiency distributions for C&FL versus NC&FL cannot be rejected in TCC subsamples 7 and 8.

Taken together, our empirical results quantify and test the impact of radial versus nonradial efficiency measures, and C versus NC technologies on the efficiency distributions. The reported divergences in efficiency are: (i) more pronounced for the radial versus nonradial efficiency comparison compared to the C versus NC efficiency measures, and (ii) highly dependent on sample size mainly for the C versus NC efficiency measures.

In Table 5 we report the specific test for C developed by Kneip et al. (2016) and augmented by Simar and Wilson (2020) for the production technology. The null hypothesis is that the technology is C. The alternative hypothesis is that the technology is NC. In fact, there are two tests, denoted as KSW#1 and KSW#2.[8] KSW#1 computes the average of the Kneip et al. (2016) test statistic across several sample splits. KSW#2 conducts a Kolmogorov–Smirnov test to evaluate the uniformity of the distribution of p-values across multiple sample splits. Both tests generally obtain similar results, except for Fan et al. (1996), and TCC subsamples 3, 6 and 8, in which only one test rejects C. All in all, Table 5 shows that we can always find evidence against C, except for Cesaroni et al. (2011), and TCC subsamples 5 and 7. This result is consistent with the result in Table 4, where we find that the C versus NC efficiency distributions are significantly different, except for a few small samples. Combining these results, it is clear that for most samples the production technology is NC.

5.2. Number of peers (working hypothesis $W H_1$)

Turning to the focus of our research, the learning from peers perspective, we now examine how the radial versus nonradial and the C versus NC modelling decisions influence the number of peers. Table 6 displays the average number of DMU peers for the four different combinations. Again, the last four columns headers correspond with the labels used in Fig. 1.

For the public data sets, we can infer the following conclusions. First, as expected, the number of peers is higher for the C than for the NC efficiency measures. The grand averages are 2.84 and 2.12 on the one hand, and 1.02 and 1.00 on the other hand, respectively. Second, the number of peers is higher for the radial than for the nonradial approaches. Grand averages amount to 2.84 versus 2.12 for the C case, and 1.02 versus 1.00 for the NC case, respectively. Third, the number of NC peers is consistently equal to one 1, with the exception of the NC&DF results for the Färe et al. (1983) and the Cesaroni et al. (2011) data sets. Indeed, our code explicitly considers all NC peers: there is always the possibility of multiple optimal solutions in that several dominating observations obtain an identical efficiency measure. In particular, for the latter two data sets we identify an average number of peers of 1.06 (4 DMUs having 2 peers) respectively 1.04 (5 DMUs having 2 peers).

These patterns observed in the public data sets are somewhat confirmed by the TCC data sets. First, the number of peers is higher for the C than for the NC efficiency measures for the smallest samples, but the reverse is true for the largest samples. Indeed, the NC results reveal a remarkable increase in average number of peers for the radial approach: with an average number of 15.40, it clearly and rapidly increases with sample size, and climbs from an average of 1.43 ($K = 40$, smallest subsample) to 41.99 ($K = 10\,198$, the highest subsample size). This is partially due to the discrete nature of the input variables for the TCC model (number of hours of operators and surveillance staff), but also confirms and at the same time magnifies the slumbering and slightly emerging effect found in the public data sets. The NC&FL results display a much more moderate but still pertinent increase in number of peers, and with an average of 3.76 it is comparable to (and very slightly exceeds) the C&DF and C&FL results. Second, the number of peers is again higher for the radial than for the nonradial approaches.

In conclusion, radial efficiency measures have on average more peers than nonradial efficiency measures on C but especially so on NC technologies. Thus, the choice of radial versus nonradial efficiency measures seems to have a much higher impact on the number of peers - and therefore on the learning process - in NC technologies depending on sample size. If our hypothesis H_1 is true, then it may be better to opt for a nonradial rather than a radial efficiency measure to simplify the learning process.

5.3. Peer similarity (working hypothesis $W H_2$)

We now consider the impact of the radial versus nonradial and C versus NC approaches on peer similarity as assessed through the Jaccard Similarity Index. Table 7 displays the Jaccard Similarity Index (JSI) for the five published and the nine TCC data sets. For each data set, the average JSI is calculated. The overall average is the average of these average values (as such it is not influenced by the respective

[7] Note that the added uniform noise depends on the convergence rates, which are $K^{-2/(N+M+1)}$ and $K^{-1/(N+M)}$ for C and NC estimators, respectively: see Kneip et al. (2015) for more details. Moreover, we use the `npdeneqtest` command from `np` package in R. This R-code is available upon simple request.

[8] These tests are computed using the FEAR package in R that is available on the web site: https://pww.people.clemson.edu/Software/FEAR/fear.html. Moreover, in the `test.convexity` command we set the following parameters: NSPLIT=20, NREP=1000, and NBCR=100.

Table 4
Results of Li tests adjusted by Simar and Zelenyuk (2006).

Data set	K	Radial vs nonradial		Convex vs nonconvex	
		C&DF C&FL 1	NC&DF NC&FL 2	C&DF NC&DF 3	C&FL NC&FL 4
Färe et al. (1985)	32	3.44 (0.002 ***)	16.76 (<0.001 ***)	16.65 (<0.001 ***)	5.36 (<0.001 ***)
Haag et al. (1992)	41	6.08 (<0.001 ***)	-0.89 (0.971)	1.94 (0.031 **)	3.88 (<0.002 ***)
Färe et al (1983)	86	6.30 (<0.001 ***)	-6.96 (0.001 ***)	28.33 (<0.001 ***)	18.24 (<0.001 ***)
Cesaroni et al. (2011)	92	1.06 (0.038 **)	-5.06 (0.086 *)	25.72 (<0.001 ***)	28.43 (<0.001 ***)
Fan et al. (1996)	471	14.51 (<0.001 ***)	75.06 (<0.001 ***)	85.32 (<0.001 ***)	16.08 (<0.001 ***)
TCC subsample 8	40	-0.18 (0.653)	2.24 (0.012 **)	3.50 (0.163)	2.48 (0.310)
TCC subsample 7	80	1.38 (0.033 **)	3.56 (0.002 ***)	5.25 (<0.001 ***)	0.26 (0.194)
TCC subsample 6	160	10.12 (<0.001 ***)	-0.20 (0.001 ***)	14.93 (0.187)	20.11 (<0.001 ***)
TCC subsample 5	319	24.67 (<0.001 ***)	14.79 (<0.001 ***)	35.58 (0.552)	41.50 (<0.001 ***)
TCC subsample 4	638	-0.92 (0.066 *)	5.33 (<0.001 ***)	69.36 (<0.001 ***)	88.43 (<0.001 ***)
TCC subsample 3	1275	14.72 (<0.001 ***)	123.77 (<0.001 ***)	134.77 (<0.001 ***)	154.05 (<0.001 ***)
TCC subsample 2	2550	159.52 (<0.001 ***)	-79.96 (<0.001 ***)	346.13 (<0.001 ***)	403.35 (<0.001 ***)
TCC subsample 1	5099	361.59 (<0.001 ***)	127.58 (<0.001 ***)	738.19 (<0.001 ***)	780.49 (<0.001 ***)
TCC full sample	10198	1391.52 (<0.001 ***)	232.05 (<0.001 ***)	1744.80 (<0.001 ***)	1533.16 (<0.001 ***)

Table 5
Results of KSW convexity tests.

Data set	K	KSW#1 1	KSW#2 2
Färe et al. (1985)	32	2.13 (0.044 **)	0.79 (0.011 **)
Haag et al. (1992)	41	3.58 (0.001 ***)	0.93 (0.001 ***)
Färe et al (1983)	86	2.93 (0.006 ***)	0.88 (0.002 ***)
Cesaroni et al. (2011)	92	-0.56 (0.760)	0.25 (0.412)
Fan et al. (1996)	471	-0.52 (0.760)	0.31 (0.097 *)
TCC subsample 8	40	-2.28 (0.987)	0.59 (0.021 **)
TCC subsample 7	80	-0.56 (0.743)	0.22 (0.506)
TCC subsample 6	160	-1.03 (0.978)	0.42 (0.015 **)
TCC subsample 5	319	-0.19 (0.736)	0.09 (0.963)
TCC subsample 4	638	1.27 (<0.001 ***)	0.48 (0.002 ***)
TCC subsample 3	1275	0.39 (0.007 ***)	0.28 (0.237)
TCC subsample 2	2550	0.85 (<0.001 ***)	0.43 (0.004 ***)
TCC subsample 1	5099	0.66 (0.011 **)	0.30 (0.083 *)
TCC full sample	10198	1.05 (0.004 ***)	0.50 (<0.001 ***)

Table 6
Average number of peers.

Data set	K	C&DF Ⓐ	C&FL Ⓑ	NC&DF Ⓒ	NC&FL Ⓓ
Färe et al. (1985)	32	2.13	1.75	1.00	1.00
Haag et al. (1992)	41	2.88	2.54	1.00	1.00
Färe et al (1983)	86	1.99	1.91	1.06	1.00
Cesaroni et al. (2011)	92	3.26	2.21	1.04	1.00
Fan et al. (1996)	471	3.96	2.21	1.00	1.00
Average		**2.84**	**2.12**	**1.02**	**1.00**
TCC subsample 8	40	2.17	1.95	1.43	1.18
TCC subsample 7	80	2.31	1.82	2.38	1.49
TCC subsample 6	160	2.97	2.71	3.52	1.99
TCC subsample 5	319	3.08	2.84	6.29	3.33
TCC subsample 4	638	3.58	2.75	10.34	4.16
TCC subsample 3	1275	3.40	2.76	13.05	3.14
TCC subsample 2	2550	3.85	3.16	22.64	5.29
TCC subsample 1	5099	4.05	3.39	36.99	7.16
TCC full sample	10198	3.80	3.27	41.99	6.13
Average		**3.26**	**2.74**	**15.40**	**3.76**

sample sizes). Since all numbers vary between 0 and 100, a heat map is applied. The heat map colours range from green (JSI = 100) to red (JSI = 0).

For the radial versus nonradial comparison, and for the public data sets, the similarity is twice as high in the NC compared to the C technology (an average JSI of 43 for C&DF versus C&FL, and 90 for NC&DF versus NC&FL). This difference in similarity is less pronounced in the TCC data sets, where we see an average of 59 in C technologies (column 1) against 68 in the NC technologies (column 2). There seems to be no clear impact of sample size in the C results (column 1), for both the public and the TCC data. However, there is a noticeable sample size effect in the NC results (column 2): the JSI gradually lowers from 89 to 56. Observe that for the two largest samples the similarity under NC is even lower than under C.

Peer similarity is much lower for the C versus NC comparison: although it is slightly better for the FL measurements (an average of 26 for the public data sets and 21 for the TCC data sets, see column 4), it is still markedly lower than the similarities found in the radial versus nonradial comparison. Here as well, sample size seems to play

Table 7
Average Jaccard Similarity Index (JSI).

Data set	K	Radial vs nonradial		Convex vs nonconvex	
		C&DF C&FL [1]	NC&DF NC&FL [2]	C&DF NC&DF [3]	C&FL NC&FL [4]
Färe et al. (1985)	32	44	100	28	31
Haag et al. (1992)	41	51	100	24	24
Färe et al (1983)	86	30	95	22	21
Cesaroni et al. (2011)	92	53	79	20	25
Fan et al. (1996)	471	37	75	18	27
Average		**43**	**90**	**22**	**26**
TCC subsample 8	40	66	89	44	52
TCC subsample 7	80	57	84	30	47
TCC subsample 6	160	62	75	20	25
TCC subsample 5	319	59	71	17	23
TCC subsample 4	638	63	66	8	15
TCC subsample 3	1275	62	60	8	15
TCC subsample 2	2550	50	58	6	6
TCC subsample 1	5099	53	54	4	4
TCC full sample	10198	58	56	4	6
Average		**59**	**68**	**16**	**21**

an important role. For the C&DF and NC&DF comparison in the column [3], this could be a (partial) consequence of the increasing number of peers found in the NC&DF approach. However, for the C&FL and NC&FL comparison (column [4]), where the number of peers is found to be of a comparable magnitude, the decrease in JSI follows an almost analogous pattern. For the largest TCC sample, the JSI reaches in both cases a remarkably low value: a JSI of 4 in the column [3], and a JSI of 6 in the column [4].

Clearly, from a learning from peers perspective, the similarity between C and NC approaches is substantially lower than the similarity between radial versus nonradial approaches (approximately two to three times lower). This discrepancy in learning between C and NC approaches is even further exacerbated with increasing sample size and can reach dramatically poor levels of similarity. Overall, and especially for large data sets, the choice of technology seems to be crucial for the peer identification process, though the use of nonradial measures seems to slightly alleviate this concern.

5.4. Vector-dominance of peers (working hypothesis WH_3)

Given the weak peer similarity found for the C versus NC comparison, it is relevant and useful to add an additional metric that provides further insight in the learning from peers divergences. The Peer Dominance Index (recall that the PDI is calculated as the number of dominating C DEA peers divided by the number of C DEA peers, and only for non-efficient DMUs (see Section 3.3)) conceptually broadens the JSI approach and provides additional information for the C versus NC comparison. The PDI varies between 0 and 100 percent, and can therefore also be reported in the form of a heat map in Table 8. The column headers [3] and [4] relate to the C-NC comparisons depicted in Fig. 1. Since there is (in contrast with the previous comparisons) only one type of efficient peers, we only consider the DMUs with C learning potential, i.e. the observations with C efficiency scores less than 1. The number of C-inefficient DMUs is displayed in the column "L (learning)". Note that this PDI is always (by definition) 100 percent for the NC method: therefore, it is not reported.

Results show that the proportion of vector-dominating peers is clearly limited and can be labelled as ranging from almost non-existing to rather weak. In the case of the five published data sets, the PDI is on average 6 for the DF measure, and 9 for the FL measure. For the Haag et al. (1992) data set, not a single dominating peer is found. There may seem to be an increase of the PDI with increasing sample size, but since the model specifications for the public data sets are substantially different (see Table 1), it is difficult to draw meaningful conclusions from this.

In the TCC samples, the PDI is on average 20 for the DF measure and 26 for the FL measure. For both measures, there is a clear and consistent decline in peer dominance with an increasing sample size. The PDI drops from 31 to 18 for the C&DF results (meaning that, on average, only 18% of the C peers is also dominating the DMU under observation) and from 42 to 19 for the FL measure. Furthermore, the FL measure clearly outperforms the DF measure in terms of peer dominance. However, as sample size increases, the FL measure seems to converge to similar PDI values.

In sum, our empirical analysis reveals that the large majority of best practices do not outperform the inefficient DMUs in terms of vector-dominance. For the TCC data with the most favourable overall average, about only 1/5 of the C&DF peers, identified with the objective to guide the DMUs towards better performance, are actually dominating these DMUs. The issue of non-dominating peers seems to be less prominent for the nonradial FL measure: about 1/4 of the peers are vector-dominating. For both measures, there seems to be an exacerbating impact of decreasing sample size.

Månsson (2003) is the only study known to us reporting vector-dominance results. For a small sample of 30 observations and 8 input–output dimensions, 17 observations are inefficient: only 2 among these are dominated by a single observation. Thus, our results are not exceptional in any sense.

5.5. Learning classes

In Appendix A we zoom in further on the obtained results by introducing the concept of learning classes. For the inefficient DMU, the divergences in the learning process can be classified according to the extent in which the identified peers are common. For each of the comparisons [1] till [4], we identify 3 learning classes: (i) "all peers are common" (JSI = 100 for all observations), (ii) "some peers are common" (0 <JSI <100 for all observations), and (iii) "no common peers" (JSI = 0 for all observations). Table A1 provides an explanatory overview of these learning classes, while their detailed reporting is found in Tables A2 till A5.

6. Conclusions

The main topic of this contribution is to verify how substantial differences between technical efficiencies on C and NC technologies translate themselves into different peers and learning possibilities.

Table 8
Average Peer Dominance Index (PDI).

Data set	K	L (learning) C&DF	L (learning) C&FL	C&DF 3	C&FL 4
Färe et al. (1985)	32	23	23	2	4
Haag et al. (1992)	41	31	31	0	0
Färe et al. (1983)	86	68	70	5	4
Cesaroni et al. (2011)	92	79	79	8	14
Fan et al. (1996)	471	421	422	13	21
Average				**6**	**9**
TCC subsample 8	40	30	31	31	42
TCC subsample 7	80	59	64	27	46
TCC subsample 6	160	124	148	19	21
TCC subsample 5	319	243	299	22	24
TCC subsample 4	638	609	621	14	20
TCC subsample 3	1275	1213	1244	18	23
TCC subsample 2	2550	2427	2534	17	19
TCC subsample 1	5099	4866	5079	17	16
TCC full sample	10198	9749	10174	18	19
Average				**20**	**26**

Section 2 has defined technologies as well as the traditional radial efficiency measure along with the alternative nonradial Färe and Lovell (1978) efficiency measure. Section 3 has developed three working hypotheses on how the analysis of peers in frontier technologies facilitates learning. To empirically test these three working hypotheses we define three metrics that can help in quantifying the empirical effects. Section 4 describes five existing secondary data sets with a variety of specifications and sample sizes. These existing data sets are supplemented with one large sample of Belgian traffic control centres.

Section 5 reports the detailed empirical results. Evaluating the efficiency distributions we find that the divergences in efficiency are more pronounced for the C versus NC comparison relative to the radial versus nonradial efficiency measures. Furthermore, for the radial versus nonradial efficiency measures the divergence is highly dependent on sample size.

In terms of the number of peers working hypothesis WH_1, radial efficiency measures have overall more peers than nonradial efficiency measures on C but especially on NC technologies. Thus, one may consider opting for a nonradial instead of a radial efficiency measure to simplify the learning process for managers and policy makers (e.g., regulators). Furthermore, the peer similarity working hypothesis WH_2 as measured by the Jaccard Similarity Index between radial versus nonradial approaches is substantially higher than the similarity between C and NC approaches. The latter similarity can become extremely low under large sample sizes, though this effect is slightly mitigated under nonradial measures. For managers and policy makers peer similarity is highest for radial versus nonradial efficiency under NC, and for C versus NC under a nonradial efficiency measure: thus, the choice of efficiency measure makes least impact under NC and the choice of C versus NC technology has the least effect under a nonradial efficiency measure, respectively.

While under a NC technology inefficiency is always related to vector-dominance, our empirical analysis reveals that the large majority of best practices under a C technology do not outperform the inefficient DMUs in terms of vector-dominance or peer dominance (working hypothesis WH_3). On average, at most about only 1/5 of the C&DF peers and about 1/4 of the C&FL peers are vector-dominating, respectively, with a noticeable impact of increasing sample size. This result implies that convexity may greatly hinder the learning processes: this is a managerial argument questioning convexity. Apart from the reasons exposed in the Introduction, when learning from peers is crucial for managers and policy makers, then one should be critical w.r.t. C: we recommend using the Simar and Zelenyuk (2006) and the Kneip et al. (2016) tests.

Overall, if one wants to minimize the number of peers, then it is clear that the nonradial FL efficiency measure is a better choice than the traditional radial one. Peer similarity over model variations is higher between radial versus nonradial approaches than between C and NC approaches. Finally, traditional C methods have a rather low peer dominance compared to the NC approach. Thus, the combination of the logical, the statistical, and the managerial arguments against convexity is rather overwhelming in our point of view. Perfect time divisibility is very implausible on logical grounds in almost all application contexts. Most but not all statistical tests reject convexity in our particular empirical samples. It is important to apply such tests of convexity in all empirical applications. Finally, the managerial concerns developed in this paper regarding the learning from peers show that convexity may greatly hinder the learning processes. Overall, we think convexity is an axiom that should be scrutinized in all these three respects in all future methodological innovations as well as in empirical applications.

One potential avenue for future research is to investigate how the learning of peers works under C and NC cost function estimates. Moreover, it could be good to replicate the current results with big to huge data sets solely (instead of including also small data sets as we have done). Furthermore, for robustness sake it could be good to duplicate results with another resemblance measure than the JSI. One limitation of the analysis –pointed out by a referee– is that we treat all efficient observations on an equal footing: this may be remedied in future research, e.g., by using economic value (e.g., cost) functions, as already suggested by the first point above.

CRediT authorship contribution statement

Kristiaan Kerstens: Writing – original draft, Supervision, Investigation, Data curation, Writing – review & editing, Validation, Methodology, Formal analysis, Conceptualization. **Bart Roets:** Writing – review & editing, Visualization, Software, Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization, Investigation, Data curation. **Ignace Van de Woestyne:** Writing – review & editing, Visualization, Supervision, Methodology, Formal analysis, Writing – original draft, Validation, Software, Investigation, Conceptualization. **Shirong Zhao:** Visualization, Software, Methodology, Formal analysis, Writing – review & editing, Validation, Resources, Investigation, Conceptualization.

Acknowledgments

Shirong Zhao acknowledges the support from the National Natural Science Foundation of China (grant number 72401056).

Appendices. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ejor.2025.07.062.

References

Apon, A. W., Ngo, M. E., Payne, L. B., & Wilson, P. W. (2015). Assessing the effect of high performance computing capabilities on academic research output. *Empirical Economics*, *48*(1), 283–312.

Batagelj, V., & Bren, M. (1995). Comparing resemblance measures. *Journal of Classification*, *12*(1), 73–90.

Bogetoft, P., Tama, J. M., & Tind, J. (2000). Convex input and output projections of nonconvex production possibility sets. *Management Science*, *46*(6), 858–869.

Boles, J. N. (1966). Efficiency squared–efficient computation of efficiency indexes. In *Proceedings of the annual meeting (Western farm economics association), 39 (August 15-17, 1966* (pp. 137–142). Western Agricultural Economics Association.

Bouhnik, S., Golany, U., Passy, S. T., Hackman, B., & Vlatsa, D. A. (2001). Lower bound restrictions on intensities in Data Envelopment Analysis. *Journal of Productivity Analysis*, *16*(3), 241–261.

Briec, W., Dumas, K., Kerstens, A., & Stenger, A. (2022). Generalised commensurability properties of efficiency measures: Implications for productivity indicators. *European Journal of Operational Research*, *303*(3), 1481–1492.

Briec, W., Kerstens, K., & Van de Woestyne, I. (2022). Nonconvexity in production and cost functions: An exploratory and selective review. In S. C. Ray, R. Chambers, & S. Kumbhakar (Eds.), *Handbook of production economics*: *Vol. 2*, (pp. 721–754). Singapore: Springer.

Casula, M., Rangarajan, N., & Shields, P. (2021). The potential of working hypotheses for deductive exploratory research. *Quality & Quantity*, *55*(5), 1703–1725.

Cesaroni, G. (2011). A complete FDH efficiency analysis of a diffused production network: The case of the Italian driver and vehicle agency. *International Transactions in Operational Research*, *18*(2), 205–229.

Cesaroni, G., Kerstens, K., & Van de Woestyne, I. (2017). Global and local scale characteristics in convex and nonconvex nonparametric technologies: A first empirical exploration. *European Journal of Operational Research*, *259*(2), 576–586.

Chavas, J. P., & Kim, K. (2015). Nonparametric analysis of technology and productivity under non-convexity: A neighborhood-based approach. *Journal of Productivity Analysis*, *43*(1), 59–74.

Daraio, C., & Simar, L. (2016). Efficiency and benchmarking with directional distances: A data-driven approach. *Journal of the Operational Research Society*, *67*(7), 928–944.

De Borger, B., Ferrier, G., & Kerstens, K. (1998). The choice of a technical efficiency measure on the Free Disposal Hull reference technology: A comparison using US banking data. *European Journal of Operational Research*, *105*(3), 427–446.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, *19*(3), 273–292.

Emrouznejad, A., Parker, B. R., & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, *42*(3), 151–157.

Emrouznejad, A., & Yang, G. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*, *61*, 4–8.

Epstein, M., & Henderson, J. (1989). Data Envelopment Analysis for managerial control and diagnosis. *Decision Sciences*, *20*(1), 90–119.

Fan, Y., Li, Q., & Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models. *Journal of Business & Economic Statistics*, *14*(4), 460–468.

Fan, Y., & Ullah, A. (1999). On goodness-of-fit tests for weakly dependent processes using kernel method. *Journal of Nonparametric Statistics*, *11*(1), 337–360.

Färe, R., Grosskopf, S., & Logan, J. (1983). The relative efficiency of Illinois electric utilities. *Resources and Energy*, *5*(4), 349–367.

Färe, R., Grosskopf, J., Logan, S., & Lovell, C. A. K. (1985). Measuring efficiency in production: With an application to electric utilities. In A. Dogramaci, & N. Adam (Eds.), *Managerial issues in productivity analysis* (pp. 185–214). Boston: Kluwer.

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1994). *Production frontiers*. Cambridge: Cambridge University Press.

Färe, R., & Lovell, C. A. K. (1978). Measuring the technical efficiency of production. *Journal of Economic Theory*, *19*(1), 150–162.

Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A: General*, *120*(3), 253–281.

Ferrier, G., Kerstens, K., & Vanden Eeckaut, P. (1994). Radial and nonradial technical efficiency measures on a DEA reference technology: A comparison using US banking data. *Recherches Économiques de Louvain*, *60*(4), 449–479.

Ghahraman, A., & Prior, D. (2016). A learning ladder toward efficiency: Proposing network-based stepwise benchmark selection. *Omega*, *63*, 83–93.

Haag, S., Jaska, P., & Semple, J. (1992). Assessing the relative efficiency of agricultural production units in the Blackland Prairie, Texas. *Applied Economics*, *24*(5), 559–565.

Hackman, S. T. (2008). *Production economics: integrating the microeconomic and engineering perspectives*. Berlin: Springer.

Halme, M., Korhonen, P., & Eskelinen, J. (2014). Non-convex value efficiency analysis and its application to bank branch sales evaluation. *Omega*, *48*, 10–18.

Hung, N. M., Le Van, C., & Michel, P. (2009). Non-convex aggregate technology and optimal economic growth. *Economic Theory*, *40*(3), 457–471.

Kneip, Alois, Simar, Léopold, & Wilson, Paul W. (2015). When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores. *Econometric Theory*, *31*(2), 394–422.

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

Krüger, J. J. (2018). Direct targeting of efficient DMUs for benchmarking. *International Journal of Production Economics*, *199*, 1–6.

Li, Q. (1996). Nonparametric testing of closeness between two unknown distribution functions. *Econometric Reviews*, *15*(1), 261–274.

Li, Q., Maasoumi, E., & Racine, J. S. (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. *Journal of Econometrics*, *148*(2), 186–200.

López-Torres, L., Johnes, C., Elliott, J., & Polo, C. (2021). The effects of competition and collaboration on efficiency in the UK independent school sector. *Economic Modelling*, *96*, 40–53.

Mas-Colell, A. (1987). Non-convexity. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The new Palgrave: a dictionary of economics* (pp. 653–661). London: Palgrave Macmillan.

Månsson, J. (2003). How can we use the result from a DEA analysis? Identification of firm-relevant reference units. *Journal of Applied Economics*, *6*(1), 157–175.

O'Loughlin, C. T., & Wilson, P. W. (2021). Benchmarking the performance of US municipalities. *Empirical Economics*, *60*(6), 2665–2700.

Parkan, C. (1987). Measuring the efficiency of service operations: An application to bank branches. *Engineering Cost and Production Economics*, *12*(1–4), 237–242.

Roets, B., Verschelde, M., & Christiaens, J. (2018). Multi-output efficiency and operational safety: An analysis of railway traffic control centre performance. *European Journal of Operational Research*, *271*(1), 224–237.

Romer, P. M. (1990). Are nonconvexities important for understanding growth? *American Economic Review*, *80*(2), 97–103.

Ruggiero, J., & Bretschneider, S. (1998). The weighted Russell measure of technical efficiency. *European Journal of Operational Research*, *108*(2), 438–451.

Ruiz, J. L., & Sirvent, I. (2022a). Benchmarking within a DEA framework: Setting the closest targets and identifying peer groups with the most similar performances. *International Transactions in Operational Research*, *29*(1), 554–573.

Ruiz, J. L., & Sirvent, I. (2022b). Identifying suitable benchmarks in the way toward achieving targets using Data Envelopment Analysis. *International Transactions in Operational Research*, *29*(3), 1749–1768.

Russell, R. R., & Schworm, W. (2009). Axiomatic foundations of efficiency measurement on data-generated technologies. *Journal of Productivity Analysis*, *31*(2), 77–86.

Russell, R. R., & Schworm, W. (2011). Properties of inefficiency indexes on <input, output> space. *Journal of Productivity Analysis*, *36*(2), 143–156.

Shephard, R. W. (1967). The notion of a production function. *Mathematical Methods of Operations Research*, *11*(1), 209–232.

Shephard, R. W. (1970). *Theory of cost and production functions*. Princeton: Princeton University Press.

Silva Portela, M. C. A., Borges, P. C., & Thanassoulis, E. (2003). Finding closest targets in non-oriented DEA models: The case of convex and non-convex technologies. *Journal of Productivity Analysis*, *19*(2–3), 251–269.

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

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

Topcu, T. G., Triantis, K., & Roets, B. (2019). Estimation of the workload boundary in socio-technical infrastructure management systems: The case of Belgian railroads. *European Journal of Operational Research*, *278*(1), 314–329.

Wilson, P. W. (2021). U.S. banking in the post-crisis era: New results from new methods. In C. F. Parmeter, & R. C. Sickles (Eds.), *Advances in efficiency and productivity analysis* (pp. 233–264). Berlin: Springer.

Zhu, J. (1996). Data Envelopment Analysis with preference structure. *Journal of the Operational Research Society*, *47*(1), 136–150.