Input allocation in multi-output settings: nonparametric robust efficiency measurements*

Barnabé Walheer†

August 6, 2018

Abstract

Comparing decision making units to detect their potential efficiency improvement, without relying on parametric unverifiable assumptions about the production process, is the goal of nonparametric efficiency analysis (such as FDH, DEA). While such methods have demonstrated their practical usefulness, practitioners sometimes have doubts about their fairness. In multi-output settings, two main limitations could give credit to their doubts: (1) the production process is modelled as a "black box", i.e. it is implicitly assumed that all the inputs produce simultaneously all the outputs; (2) only techniques investigating for outliers in all output directions simultaneously exist. In this paper, we tackle these two limitations by presenting two new nonparametric robust efficiency measurements for multi-output settings. Our new measurements present several attractive features. First, they increase the realism of the modelling by taking the links between inputs and outputs into account, and thus tackle (1). Second, they provide flexibility in the outlier detection exercise, and thus also tackle (2). Overall, our new measurements better use the data available, and can be seen as natural extensions of well-known nonparametric robust efficiency measurements for multi-output contexts. To demonstrate the usefulness of our method, we propose both a simulation and an empirical application.

Keywords: DEA; FDH; multi-output; order$-m$; order$-\alpha$; robust.

*We thank the Editors Thomas Archibald and Jonathan Crook and the referee for their comments that have improved the paper substantially.

†International Business School Suzhou, Xi'An Jiaotong-Liverpool University, China. email: Barnabe.Walheer@xjtlu.edu.cn.

1 Introduction

Nonparametric efficiency analysis of production activities is a technique used to evaluate a decision making unit (DMU; such as a firm, a plant, or a utility) by comparing its input-output performance to that of other DMUs operating in a similar technological environment. Nonparametric efficiency analysis is not based on some unverifiable parametric/functional specification of the production technology, but instead it 'lets the data speak for themselves' by reconstructing the production technology using the available data and by imposing technology axioms (such as monotonicity, convexity, or returns-to-scale). The most popular nonparametric efficiency models are Data Envelopment Analysis (DEA; Charnes, Cooper and Rhodes (1978)) and Free Disposal Hull (FDH; Deprins, Simar and Tulkens (1984)). Refer to Färe, Grosskopf and Lovell (1994), Cooper, Seiford and Zhu (2004), Cooper, Seiford and Tone (2007), Fried, Lovell and Schmidt (2008), and Cook and Seiford (2009) for reviews. Nonparametric efficiency analysis has demonstrated its practical usefulness for practitioners as the efficiency improvements detected could lead to potential cost reductions/profit improvements.

In general, nonparametric efficiency models (such as FDH and DEA) use all the data to reconstruct the production technology. As a consequence, these methods are very sensitive to outliers. The impacts on the results of the efficiency analysis could be huge since outliers disproportionately influence the performance evaluation of the DMUs. To solve that issue, Cazals, Florens and Simar (2002), Aragon, Daouia and Thomas-Agnan (2005), Daouia and Simar (2005, 2007a), Wheelock and Wilson (2008), Simar and Vanhems (2012), and Simar, Vanhems and Wilson (2012) amongst others, have proposed two nonparametric robust efficiency measurements: the order$-m$ efficiency measurement (where m can be viewed as a trimming parameter); and the order$-\alpha$ efficiency measurement (analogous to traditional quantile function). These measurements reconstruct the production technology using sub-samples of the data set. As such, they are less sensitive to outliers, i.e. more robust.[1]

[1]Note that other methods to deal with outliers in efficiency analysis have been proposed. We consider both the order$-m$ and the order$-\alpha$ efficiency measurements as they are established as the standard procedures for detecting and correcting for the presence of outliers. For alternative techniques, see, for example, Wilson (1993, 1995), Porembski, Breitenstein and Alpar (2005), Banker and Chang (2006), Johnson and McGinnis (2008), Chen and Johnson (2010), Tran, Shively and Preckel (2010). All these procedures should not be seen as competitors, but as complements since detecting outliers (especially in multi-output contexts) is not an easy task. In principle, these

Recently, much effort has been utilized to increase the realism of nonparametric efficiency models in multi-output contexts. In particular, available information about the links between inputs and outputs have been used to open the "black box"modelling of the efficiency analysis. Notable examples include Network DEA (see, for example, Färe and Grosskopf (2000), Färe, Grosskopf and Whittaker (2007), and Tone and Tsutsui (2009)), the disaggregating output-input vector DEA (see Salerian and Chan (2005)), and DEA-R (see Despic, Despic and Paradi (2007)). These methodologies consider different ways of allocating inputs to outputs. Network DEA assumes that all inputs can be fully allocated to outputs, while the disaggregating output-input vector DEA and DEA-R assume the opposite, i.e. inputs can only be partially allocated to outputs. Inspiriting by those initial attempts and building on the initial idea of Cherchye et al (2008); Cherchye et al (2013) and Cherchye, De Rock, and Walheer (2016) have suggested a new nonparametric multi-output efficiency technique taking the links between inputs and outputs into account. Their technique models each output separately by its own production technology, and thus gives the option to consider any type of input allocation. As a result, their new model provides a unifying framework that is consistent with the approaches cited above. Moreover, their nonparametric multi-output efficiency measurement has, in general, a greater ability to detect inefficient behaviour. Their model has been used and extended in, for example, Cherchye, De Rock, and Walheer (2015), Walheer (2016a, b, 2018b, f, g), Zhou et al (2016), Ding et al (2017), Silva (2018), and Walheer and Zhang (2018).

In the paper, we propose the robust counterparts of the nonparametric multi-output efficiency measurement of Cherchye et al (2013) and Cherchye, De Rock and Walheer (2016). Indeed, while their efficiency measurement has several attractive features, it is also sensitive to outliers. We introduce the concepts of order$-\mathbf{m}$ and order$-\boldsymbol{\alpha}$ multi-output efficiency (note that, this time, $\mathbf{m}$ and $\boldsymbol{\alpha}$ are vectors, see Section 2.3). The two new robust efficiency measurements give the option to investigate for outliers for each output production process individually. In other words, they provide flexibility in the outlier detection exercise. In practice, this is captured by output-specific robustness parameters (i.e. output-specific m and α). Consequently, the two new robust efficiency measurements could be seen as natural extensions in multi-output contexts of the concept of order$-m$ and order$-\alpha$ efficiency measurements. All in all, these two new nonparametric robust multi-output efficiency mea-

alternative methods could be adapted for the multi-output contexts described in the paper.

surements better use the data available, and provide extra valuable information about the (in)efficient behaviours and outlier profiles of the DMUs.

To demonstrate the usefulness of our two new nonparametric robust multi-output efficiency measurements, we rely on both a simulation and an empirical application for electricity plants in the United States. Our method is particularly useful to study these plants. Indeed, more and more plants produce different types of electricity generation. As such, outliers could be different for each type of electricity generation. That is, outliers could have different profiles for every electricity generation. Besides this advantage, our nonparametric robust multi-output efficiency measurements also provide a more realic model of the plant production process by linking the inputs to each type of electricity generation. Finally, our measurements are also better able to detect inefficient behaviour.

The rest of this paper unfolds as follows. In Section 2, we define our notion of robust multi-output efficiency measurements, and show how to estimate these measurements in practice. In Section 3, we present our simulation and application for US electricity plants. In Section 4, we present our conclusions.

2 Methodology

We develop nonparametric robust efficiency measurements for multi-output contexts. As an initial step, we briefly review the existing efficiency measurement for these contexts. By doing so, we also extend this initial definition in two directions. First, we rewrite the measurement allowing for efficiency investigation for every output. Second, we provide the probabilistic formulation of the efficiency measurement. Next, we define our notion of robust efficiency measurements for multi-output contexts. Finally, we provide the nonparametric estimators.

In the following, we consider the input orientation. The methodology is easily extended to the output orientation, or to other non-oriented efficiency measurements (as the directional distance function or slacks-based efficiency measurements). We do not report here for the sake of compactness.

2.1 Multi-output efficiency

We consider that DMUs use P inputs, captured by the vector $\mathbf{x} \in \mathbb{R}_+^P$, to produce Q outputs, captured by the vector $\mathbf{y} \in \mathbb{R}_+^Q$. For multi-output contexts, it is often the case that the inputs are used differently to produce every output. That is, the inputs are allocated to each individual output. In particular, we denote by $\mathbf{x}^q \in \mathbb{R}_+^P$ the inputs used to produce output q (denoted by y^q in the following). Building on this concept, we naturally model the production process by means of output-specific technology sets. Before formally defining these sets, we start by discussing three different ways to allocate the inputs. Next, we define the notion of multi-output efficiency measurement.

Input allocation. We consider three ways to allocate the inputs to each individual output:

- The *output-specific* inputs are allocated to the outputs. Let $\alpha_p^q \in [0,1]$, with $\sum_{q=1}^{Q} \alpha_p^q = 1$, represents the fraction of the p-th output-specific input quantity that is allocated to output q. These types of input have been considered in, for example, Färe and Grosskopf (2000), Färe, Grosskopf and Whittaker (2007), Tone and Tsutsui (2009), Cherchye et al (2013), Walheer (2016a, b, 2018h), Zhou et al (2016), Silva (2018), and Walheer and Zhang (2018).

- The *joint* inputs are simultaneously used to produce all the outputs. These types of input have been considered in, for example, Cherchye et al (2013), Cherchye, De Rock, and Walheer (2016), Ding et al (2017), Walheer (2018b, d, f, g), and Walheer and Zhang (2018).

- The *sub-joint* inputs are simultaneously used to produce a sub-set of outputs. These types of input have been considered in, for example, Salerian and Chan (2005), Despic, Despic, and Paradi (2007), Cherchye, De Rock, and Walheer (2015), and Walheer (2018b, c, h).

Attractively, we can relate the inputs used to produce output q to the initial input vector as follows:

$$\mathbf{x}_p^q = \begin{cases} \mathbf{x}_p & \text{if input } p \text{ is joint or sub-joint and used to produce output } q, \\ \alpha_p^q \cdot \mathbf{x}_p & \text{if input } p \text{ is output-specific and used to produce output } q, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In words, when input p is used jointly to produce all or a subset of outputs, the input quantity to produce output q corresponds to the initial input quantity; while when input p is allocated between outputs, only the fraction α_p^q of the input quantity is used to produce output q. See Section 4 for an illustration.

Output-specific technology sets. The previous definition of the inputs $\mathbf{x}^q$ used to produce output q allows us to characterize naturally each output q by its own production technology, captured, for each q, by a production possibility set defined as follows:

$$T^q = \{(\mathbf{x}^q, y^q) \in \mathbb{R}_+^{P+1} \mid \mathbf{x}^q \text{ can produce } y^q\}. \tag{2}$$

The set T^q contains all the combinations of output-specific, joint and sub-joint inputs (in $\mathbf{x}^q$) that can produce the output quantity y^q. In general, some regularity conditions, captured by technology axioms, are assumed for the production possibility sets. They are also assumed to avoid a trivial reconstruction of the technology (see Section 2.4 for more discussion). As we consider output-specific production possibility sets, we naturally define the technology axioms at the output level. In particular, we assume that the sets satisfy the following two axioms:

A1 (freely disposable outputs): $(\mathbf{x}^q, y^q) \in T^q$ and $y^q \geq y^{q'} \implies (\mathbf{x}^q, y^{q'}) \in T^q$.

A2 (freely disposable inputs): $(\mathbf{x}^q, y^q) \in T^q$ and $\mathbf{x}^{q'} \geq \mathbf{x}^q \implies (\mathbf{x}^{q'}, y^q) \in T^q$.

In words, **A1** says that it is always possible to produce less output for a given input quantity, and **A2** implies that more inputs never reduces the output.[2] These two axioms form a minimal attractive set of axioms for practical works. Additionally, it is often assumed that the production possibility sets are convex, defined at the output level as follows:

A3 (convexity): $(\mathbf{x}^q, y^q) \in T^q$ and $(\mathbf{x}^{q'}, y^{q'}) \in T^q \implies \forall \lambda \in [0,1] : \lambda(\mathbf{x}^q, y^q) + (1 -$

[2]We remark that $\mathbf{x}^{q'} \geq \mathbf{x}^q$ should be read as an element-by-element inequality between two vectors.

$\lambda)(\mathbf{x}^{q'}, y^{q'}) \in T^q$.

It is worth noting that imposing convexity implies more structure on theses sets (captured by **A3**). This is also why in the following we consider both options. See, for example, Deprins, Simar and Tulkens (1984), Kerstens and Vanden Eeckaut (1999), Briec, Kerstens and Vanden Eeckaut (2004), Podinovski (2004a, b), Leleu (2009), and Huang et al (2013) for related works without assuming convexity of the technology.

We end this part by providing two remarks. First, extra axioms can fairly easily be considered; for example weak disposability of the inputs/outputs (see Färe et al (1989), and Podinovski and Kuosmanen (2011)), relaxed convexity assumptions (see Petersen (1990) and Bogetoft (1996)), returns-to-scale assumptions (see Banker et al (2004) and Walheer (2018e)), and will not change the essence of the paper, but only impact the nonparametric estimators (see Section 2.4). Second, we point out that assuming these technology axioms for each output-specific technology T^q is less restrictive than assuming similar axioms for the aggregate production possibility set (i.e. the production possibility set for $(\mathbf{x}, \mathbf{y})$). Intuitively, this comes from the observation that, in general, there are no specific connections between the output-specific production possibility sets and the aggregate production possibility set. A similar discussion has been made by Walheer (2016a, b, 2018f) when considering sector- and country-level production possibility sets. Assuming specific technology axioms for the sector technologies do not imply that these axioms are fulfilled by the country technologies. In our context, this feature is particularly attractive since it means that the extra advantages of the method do not come with the drawback of imposing more structure about the production process. We refer also to Cherchye et al (2013) for more detail.

Multi-output efficiency measurement. To measure efficiency, we make use of an adapted Debreu (1951) − Farell (1957) input efficiency measurement. In particular, it is defined for a DMU operating at $(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ as:

$$\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \inf\{\theta \mid \forall q : (\theta \mathbf{x}^q, y^q) \in T^q\}. \tag{3}$$

In words, $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ defines the maximal potential radial input reduction that still allows for producing output $\mathbf{y}$. Generally, $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ is between 0 and 1, and a lower value of $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ indicates greater technical inefficiency.

$\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ depends only on the output-specific production possibility sets, and can thus be seen as a natural extension of the Debreu (1951) $-$ Farell (1957) input efficiency measurement for the multi-output contexts. Note that if there is only one output in the production process, both measurements coincide.

Also, we point out that the common θ in (3) gives the benefit of the doubt to the evaluated DMUs. Intuitively, there is no particular reason why the potential input reductions should be the same for every output q. (In general, they are different; see Section 2.2, and Section 3 for an illustration). By imposing a common input reduction, i.e. a common θ in (3), we evaluate the DMUs in the best possible light. This spirit is intrinsically present in nonparametric efficiency models. In practice, this spirit is captured by the multipliers in the programmings (see Section 2.4 for more discussion). It transpires that we can also interpret $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ as the upper bound of the efficiency measurements for multi-output settings. This will become even clearer when considering the reformulation of $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ in Section 2.2. At this point, it is important to remark that the upper bound feature of $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ does not mean that it reduces ability to detect inefficiency behaviour. In fact, in general, it detects more inefficiency behaviour than more standard efficiency measurements (i.e. those that are based on the aggregate production possibility set; see Section 3 for an illustration).

2.2 Reformulations of efficiency

The initial definition of the multi-output efficiency measurement in (3) cannot be used for defining the robust counterparts. As such, in this part, we extend the initial definition in two directions. Firstly, we reformulate the initial definition in terms of output-specific efficiency measurements. As a result, it gives us the option to investigate for efficiency behaviour for every output. Next, we show that the multi-output efficiency measurement can be defined exclusively in terms of output-specific conditional density functions. This new definition provides a natural extension of the definition of Cazals, Florens and Simar (2002) and Daraio and Simar (2005) for multi-output contexts.

Technical reformulation. $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ can alternatively be obtained using a two-steps procedure for a DMU operating at $(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$:

- Evaluate efficiency for every couples $(y^q, \mathbf{x}^q)$:

$$\theta^q(y^q, \mathbf{x}^q) = \inf\{\theta^q \mid (\theta^q \mathbf{x}^q, y^q) \in T^q\}. \tag{4}$$

- Take the maximum of the $\theta^q(y^q, \mathbf{x}^q)$'s:

$$\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \max_{q=1,\ldots,Q} \{\theta^q(y^q, \mathbf{x}^q)\}. \tag{5}$$

The $\theta^q(y^q, \mathbf{x}^q)$'s have a nice interpretation: they are the efficiency measurements for each individual output q. As $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$, $\theta^q(y^q, \mathbf{x}^q)$ is between 0 and 1. When $\theta^q(y^q, \mathbf{x}^q) = 1$, it indicates efficient behaviour for output q, while $\theta^q(y^q, \mathbf{x}^q) < 1$ indicates the opposite. These measurements give extra valuable information about the (in)efficiency behavior of the DMUs, and, in particular, gives the option to better understand why a DMU is declared as efficient or inefficient.

The benefit of the doubt spirit of $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ can also be seen by (5). Indeed, by taking the maximum, we guarantee that the DMUs are evaluated in the best possible light. Note that it is fairly easy to use other relationships, for example, the minimum, a weighted average, the median, or any quantiles, to define $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$. We choose to rely on the maximum since, as explained previously, it is the most natural procedure given the spirit of nonparametric efficiency models.[3] Also, as discussed previously, when relying on the maximum, $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ can be seen as natural extensions of the Debreu (1951) $-$ Farell (1957) efficiency measurement. Finally, note that, in general, other procedures require additional assumptions about the production process (e.g. a specific relationship between the aggregate production set and the output-specific production sets) and/or more data (e.g. the prices).[4]

Probabilistic reformulation. We can also rewrite the multi-output efficient measurement using output-specific conditional density functions. This mirrors the definition of Cazals, Florens and Simar (2002) and Daraio and Simar (2005) for the multi-output contexts. This rewriting is particularly useful for us since it facilitates

[3] We remark that when assuming a common θ in (3), we can also interpreted $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ as a shadow cost efficiency measurement. See Cherchye et al (2013) and Cherchye, De Rock and Walheer (2016) for more discussion.

[4] See, for example, Färe and Zelenyuk (2003), Kuosmanen, Cherchye and Sipilainen (2006), Färe and Karagiannis (2017), and Walheer (2018a) for related discussion.

the definition of the robust multi-output efficiency measurements. Formally, we first define the concept of output-specific conditional distribution function of $\mathbf{x}^q$ given $Y^q \geq y^q$ as follows:

$$F^q_{\mathbf{X}^q|Y^q}(\mathbf{x}^q \mid y^q) = \text{Prob}(\mathbf{X}^q \leq \mathbf{x}^q \mid Y^q \geq y^q). \tag{6}$$

The support of the output-specific conditional distribution F^q is the production possibility set T^q, and it can be interpreted as the probability to be dominated, i.e. producing more output with less inputs. It transpires that the output-specific and the multi-output efficiency measurements can equivalently be rewritten in terms of the output-specific conditional distributions as follows:

$$\theta^q(y^q, \mathbf{x}^q) = \inf\{\theta^q \mid F^q_{\mathbf{X}^q|Y^q}(\theta^q \mathbf{x}^q \mid y^q) > 0\}. \tag{7}$$

$$\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \inf\{\theta \mid \forall q : F^q_{\mathbf{X}^q|Y^q}(\theta \mathbf{x}^q \mid y^q) > 0\}. \tag{8}$$

Once more, we highlight the benefit of the doubt spirit of $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ captured by the common θ in (8). Also, when there is only one output (i.e. $q = 1$), our definition coincides with the one of Cazals, Florens and Simar (2002).

2.3 Robust multi-output efficiency

Robust efficiency measurements have been proposed to overcome the issue of sensitivity to outliers of efficiency measurements. Indeed, the presence of outliers could represent an important issue for the evaluation performance as they disproportionately influence the efficiency analysis. Two techniques, suggested by Cazals, Florens and Simar (2002), Aragon, Daouia and Thomas-Agnan (2005), Daouia and Simar (2005, 2007), Wheelock and Wilson (2008), Simar and Vanhems (2012), and Simar, Vanhems and Wilson (2012), have been established as the standards to solve this issue: order$-m$ (where m can be viewed as a trimming parameter) and order$-\alpha$ (analogous to traditional quantile function) efficiency measurements.

While the multi-output efficiency measurement has several attractive features (e.g. increasing the realism of the analysis; higher discriminatory power), it is also sensitive to outliers. As such, in the following, we propose the robust counterparts by defining the concepts of order$-\mathbf{m}$ and order$-\boldsymbol{\alpha}$ multi-output efficiency (note that, this time, $\mathbf{m}$ and $\boldsymbol{\alpha}$ are vectors). The two new robust efficiency measurements come with an extra desirable feature: they give the option to investigate for outliers for each output production process individually. In other words, they provide flexibility in the outlier detection procedure. Indeed, for multi-output settings, there is no specific reason why outliers should be the same for every output, and why an outlier could not be extreme only in the production of one or a subset of outputs As such, for those contexts, investigating for the presence of outliers in every output direction is of great interest as it represents extra valuable information. In practice, this is captured by output-specific robustness parameters (i.e. output-specific m and α). Consequently, the two new robust efficiency measurements could be seen as natural extensions in multi-output contexts of those well-established robust efficiency measurements.

Order$-\boldsymbol{\alpha}$ multi-output efficiency. Aragon, Daouia and Thomas-Agnan (2003) and Daouia and Simar (2005) proposed the concept of order$-\alpha$ efficiency measurement, where α is a quantile parameter that lies between 0 and 1. The basic idea is to define the robust efficiency measurement when restricting the conditional density function to be larger than a certain quantile. For the multi-output context, we rely on output-specific conditional density functions, giving naturally the option to define output-specific quantile parameters. In particular, we make use of α_q to denote the quantile parameter connected to output q, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)'$ to denote the vector of the output-specific quantile parameters. Building on this concept, we can define the order-$\boldsymbol{\alpha}$ multi-output efficiency measurement for a DMU operating at $(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ as:

$$\theta_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \inf\{\theta \mid \forall q : F^q_{\mathbf{X}^q|Y^q}(\theta \mathbf{x}^q \mid y^q) > 1 - \alpha_q\}. \tag{9}$$

An initial observation is that by letting $\boldsymbol{\alpha} = \mathbf{1}$ (i.e. $\alpha_q = 1$ for every q), $\theta_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ coincides with $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$. As such, $\theta_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ can be seen as a generalization of the multi-output efficiency measurement. Next, $\theta_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ benchmarks, for every q, the evaluated DMU against the input level not exceeded by $(1 - \alpha_q) * 100\%$ of DMUs among the population of units producing an output level of at least y^q. As such, $\theta_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ could be larger than 1. When $\theta_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = 1$, then the evaluated DMU is efficient at the level $\boldsymbol{\alpha} * 100\%$ (since it is dominated, for each q, by DMUs producing more than y^q with a probability $1 - \alpha_q$). When $\theta_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) < 1$, the evaluated DMU is inefficient. When $\theta_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) > 1$, the evaluated DMU is super-efficient, i.e. it can still increase its input and remains efficient.

Attractively, we can also rewrite $\theta_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ using a similar two-steps procedure as the one used for $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ in (4) − (5):

- Evaluate efficiency for every couples $(y^q, \mathbf{x}^q)$:

$$\theta^q_{\alpha_q}(y^q, \mathbf{x}^q) = \inf\{\theta^q \mid F^q_{\mathbf{X}^q|Y^q}(\theta^q \mathbf{x}^q \mid y^q) > 1 - \alpha_q\}. \tag{10}$$

- Take the maximum of the $\theta^q_{\alpha_q}(y^q, \mathbf{x}^q)$'s:

$$\theta_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \max_{q=1,\ldots,Q} \left\{\theta^q_{\alpha_q}(y^q, \mathbf{x}^q)\right\}. \tag{11}$$

$\theta^q_{\alpha_q}(y^q, \mathbf{x}^q)$ represents the order$-\alpha_q$ efficiency measurement, and thus gives a robust efficiency measurement for output q. Clearly, when $\alpha_q = 1$, we obtain that $\theta^q_{\alpha_q}(y^q, \mathbf{x}^q) = \theta^q(y^q, \mathbf{x}^q)$. This also reveals that the procedure in (10) − (11) coincides with the procedure in (4) − (5) when $\alpha_q = 1$ for every q. The order$-\alpha_q$ efficiency measurements have to be interpreted in an analogous manner as $\theta_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$, but here at the output level. Also, these measurements give the option to investigate for the presence of outliers for every output individually. That is, they give the option to better understand why a DMU is defined as an outlier.

Order−m multi-output efficiency. The concept of order$-m$ efficiency measurement has been introduced by Cazals, Florens and Simar (2002) in the univariate case, and by Daraio and Simar (2005) in the multivariate case. Two steps are needed to obtain the order$-m$ efficiency measurement. First, compute B efficiency measurements using random sub-samples of size m (usually, B should be large enough, see Section 3). Second, compute the order$-m$ efficiency measurement by taking the expectation of the B efficiency measurements. It turns out that the order$-m$ efficiency measurement is less sensible to extreme values in the sample, i.e. more robust to outliers.

For multi-output contexts, we define the multi-output efficiency measurement using output-specific technologies. As a result, we naturally obtain the option to set output-specific trimming parameters, denoted, for output q, by m_q. In practice, we define the concept of order-m_q efficiency measurement for a DMU operating at

$(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q)$ as follows:

$$\theta^q_{m_q}(y^q, \mathbf{x}^q) = E\left(\eta^q_{m_q}(y^q, \mathbf{x}^q) \mid Y^q \geq y^q\right), \tag{12}$$

$$= \int_0^\infty \left(1 - F_{\mathbf{X}^q|Y^q}(u\mathbf{x}^q \mid y^q)\right)^{m_q} du, \tag{13}$$

where $\eta^q_{m_q}(y^q, \mathbf{x}^q) = \inf\{\eta^q \mid (\eta^q \mathbf{x}^q, y^q) \in T^q_{m_q}\}$ is the output-specific efficiency measurement obtained for a random sub-sample of size m_q. Note that, in this definition $T^q_{m_q}$ is the production possibility set obtained for m_q *iid* random DMUs obtained using the output-specific conditional distribution function $F_{\mathbf{X}^q|Y^q}(. \mid y^q)$. As such, this set is random explaining why $\eta^q_{m_q}(y^q, \mathbf{x}^q)$ is random.

As done previously, we obtain the order$-\mathbf{m}$ multi-output efficiency measurement by taking the maximum of the order$-m_q$ efficiency measurements (where $\mathbf{m} = (m_1, \dots, m_Q)'$):

$$\theta_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q) = \max_{q=1,\dots,Q} \left\{\theta^q_{m_q}(y^q, \mathbf{x}^q)\right\}. \tag{14}$$

The interpretation of $\theta_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q)$ is analogous to the one of $\theta(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q)$, when restricting to sub-samples of $\mathbf{m}$ comparison partners. Also, $\theta_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q)$ can be bigger than 1, which indicates that the DMU operating at $(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q)$ is more efficient than the selected peers (in that case, it is important to investigate for (in)efficient behaviour by looking at the order$-m_q$ efficiency measurements). The interpretation of the output-specific measurement $\theta^q_{m_q}(y^q, \mathbf{x}^q)$ is similar, but applies to each individual q. Note that when m_q increases (i.e. $m_q \to \infty$), $\theta^q_{m_q}(y^q, \mathbf{x}^q) \to \theta^q(y^q, \mathbf{x}^q)$. If this holds true for every output q, we obtain that $\theta_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q)$ tends to $\theta(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q)$. We can also see this property by rewriting the integral in (13) as follows:

$$\theta^q_{m_q}(y^q, \mathbf{x}^q) = \theta^q(y^q, \mathbf{x}^q) + \int_{\theta^q(y^q, \mathbf{x}^q)}^\infty \left(1 - F_{\mathbf{X}^q|Y^q}(u\mathbf{x}^q \mid y^q)\right)^{m_q} du. \tag{15}$$

When $m_q \to \infty$, the integral converges though 0. Indeed, $1 - F_{\mathbf{X}^q|Y^q}(u\mathbf{x}^q \mid y^q)$ is by definition smaller than 1; and any number power ∞ converges to 0. Consequently, the order$-\mathbf{m}$ multi-output efficiency measurement can also be seen as a generalization of the multi-output efficiency measurement.

2.4 Nonparametric estimators

The previous definitions of the multi-output and robust multi-output efficiency measurements are not useful in practice as they are based on unknown (output-specific) production possibility sets and unknown (output-specific) conditional density functions. As explained in the Introduction, an advantage of the methodology is to reconstruct the technology using the data. As such, the starting point of the nonparametric estimation is the set of observed outputs and inputs. Suppose we observe a data set D for n DMUs. For each DMU $k \in \{1, \ldots, n\}$, we observe the output vector $\mathbf{y}_k$, and the allocation of the inputs as joint, sub-joint and output-specific inputs, i.e. $\mathbf{x}_k^1, \ldots, \mathbf{x}_k^Q$, giving the following data set D:

$$D = \{(\mathbf{y}_k, \mathbf{x}_k^1, \ldots, \mathbf{x}_k^Q) \mid k = 1, \ldots, n\}. \tag{16}$$

In practice, the unknown production possibility set is reconstructed using the "minimum extrapolation" principle. This principle states that the reconstruction must be the smallest set consistent with the chosen technology axioms. In addition, this principle also assumes that what is observed is certainly feasible.[5] This assumption implies that no measurement error is present in the production process, revealing the importance of robust efficiency measurements to correct for the deterministic feature of the nonparametric estimators.

As noted previously in Section 2.1, chosen convexity or not is probably the most debatable axiom for practical works. As such, in the following, we assume that free disposability of the inputs and outputs are fulfilled (i.e. axioms **A1** and **A2** are satisfied), but not convexity (i.e. axiom **A3** is not satisfied). That is, we consider the Free Disposal Hull (FDH) estimators. The Data Envelopment Analysis (DEA) estimators, i.e. the estimators assuming convexity of the production possibility set, are given in the Appendix. Both estimators are, by nature, nonparametric, as no functional or parametric assumptions are made about the production possibility sets.

Multi-output efficiency. The FDH estimator of the output-specific production possibility set is defined for output q as follows:

$$\widehat{T}^q_{FDH} = \left(\; (\mathbf{x}^q, y^q) \in \mathbb{R}_+^{P+1} \;\middle|\; \mathbf{x}_k^q \leq \mathbf{x}^q; y_k^q \geq y^q, k = 1, \ldots, n \; \right). \tag{17}$$

[5] Formally, $(\mathbf{y}_k, \mathbf{x}_k^1 \ldots, \mathbf{x}_k^Q) \in D \implies \forall q : (\mathbf{x}_k^q, y_k^q) \in T^q$.

By plugging-in the FDH estimator of the production possibility set in the definitions of the multi-output efficiency measurement in (3) and (5), we obtain the FDH estimator of the multi-output efficiency measurement for a DMU operating at $(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ as:

$$\widehat{\theta}_{FDH}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \inf\{\theta \mid \forall q : (\theta \mathbf{x}^q, y^q) \in \widehat{T}^q_{FDH}\}, \tag{18}$$

$$= \max_{q=1,\ldots,Q} \left\{\widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q)\right\}, \tag{19}$$

where, for every output q: $\widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q) = \inf\{\theta^q \mid (\theta^q \mathbf{x}^q, y^q) \in \widehat{T}^q_{FDH}\}$. In practice, the FDH estimator $\widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q)$ can be computed, for $q = 1, \ldots, Q$, by solving the mixed integer linear program **(MILP-1)**:

$$\widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q) = \min_{\lambda^q_k \ (k \in \{1,\ldots,n\})} \left(\theta^q \ \middle| \ \begin{array}{l} \sum_{k=1}^n \lambda^q_k \mathbf{x}^q_k \leq \theta^q \mathbf{x}^q ; \sum_{k=1}^n \lambda^q_k y^q_k \geq \lambda^q_k y^q \\ \sum_{k=1}^n \lambda^q_k = 1, \forall k : \lambda^q_k \in \{0;1\}, \theta^q \geq 0 \end{array} \right).$$

$\widehat{\theta}_{FDH}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ and $\widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q)$ have to be interpreted in an analogous manner than $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ and $\theta^q(y^q, \mathbf{x}^q)$. That is, both estimators are between 0 and 1, and lower values induce greater inefficient behaviour. Also, we point out that $\widehat{\theta}_{FDH}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) \geq \theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ and $\widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q) \geq \theta^q(y^q, \mathbf{x}^q)$ for every q. In words, the estimators evaluate the DMUs in the best light, captured by the multipliers λ^q_k. As such, the estimators give the benefit of the doubt to the DMUs. This can also be seen by noting that $\widehat{T}^q_{FDH} \subseteq T^q$, following immediately by the spirit of the "minimum extrapolation" principle.

Interestingly, the FDH estimators can also be obtained by replacing the (output-specific) conditional density distribution F^q by its empirical counterpart (where $\mathcal{I}(z)$ is the indicator function: $\mathcal{I}(z) = 1$ if z is true, $\mathcal{I}(z) = 0$ otherwise) defined, for output q, as follows:

$$\widehat{F}^q_{\mathbf{X}^q|Y^q}(\mathbf{x}^q \mid y^q) = \frac{\sum_{k=1}^n \mathcal{I}(\mathbf{X}^q_k \leq \mathbf{x}^q, Y^q_k \geq y^q)}{\sum_{k=1}^n \mathcal{I}(Y^q_k \geq y^q)}, \tag{20}$$

By plugging-in these empirical output-specific density functions in (7) and (8), we obtain:

$$\widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q) = \inf\{\theta^q \mid \widehat{F}^q_{\mathbf{X}^q|Y^q}(\theta^q \mathbf{x}^q \mid y^q) > 0\}. \tag{21}$$

$$\widehat{\theta}_{FDH}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \inf\{\theta \mid \forall q : \widehat{F}^q_{\mathbf{X}^q|Y^q}(\theta \mathbf{x}^q \mid y^q) > 0\}. \tag{22}$$

It turns out that the statistical formulation of the multi-output efficiency measurements when using the empirical output-specific conditional densities shares a one-to-one relationship with the FDH estimators of the multi-output efficiency measurements. See the Appendix for the DEA estimators, i.e. the nonparametric estimators assuming convexity of the output-specific production possibility sets.

Robust multi-output efficiency. We obtain the nonparametric estimators of the order$-\boldsymbol{\alpha}$ multi-output efficiency for a DMU operating at $(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ by plugging-in the empirical output-specific conditional density functions in (9) and (11):

$$\widehat{\theta}_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \inf\{\theta \mid \forall q : \widehat{F}^q_{\mathbf{X}^q|Y^q}(\theta \mathbf{x}^q \mid y^q) > 1 - \alpha_q\}, \tag{23}$$

$$= \max_{q=1,\ldots,Q} \left\{ \widehat{\theta}^q_{\alpha_q}(y^q, \mathbf{x}^q) \right\}, \tag{24}$$

where $\widehat{\theta}^q_{\alpha_q}(y^q, \mathbf{x}^q) = \inf\{\theta^q \mid \widehat{F}^q_{\mathbf{X}^q|Y^q}(\theta^q \mathbf{x}^q \mid y^q) > 1 - \alpha_q\}$. When $\alpha_q = 1$, the estimator $\widehat{\theta}^q_{\alpha_q}(y^q, \mathbf{x}^q)$ coincides with the estimator $\widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q)$. When it holds true for every output q, i.e. $\boldsymbol{\alpha} = \mathbf{1}$, we obtain that $\widehat{\theta}_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \widehat{\theta}_{FDH}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$. Therefore, $\widehat{\theta}_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ is defined without assuming convexity of the output-specific production possibility sets. For the nonparametric estimators when assuming convexity of the technology sets, we refer to the Appendix.

In practice computing the output-specific conditional density functions may be a difficult task. Attractively, $\widehat{\theta}_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ can be obtained by the following routine **(RTN-1)**:

1. For $q = 1, \ldots, Q$:
   (a) Compute $M^q_y = \sum_{k=1}^n \mathcal{I}(Y^q_k \geq y^q)$;
   (b) Let $\psi^q_k = \max\left(\frac{(x^q_k)_1}{(x^q)_1}, \ldots, \frac{(x^q_k)_p}{(x^q)_p}\right)$ for $k = 1 \ldots, n$;
   (c) Denote $\psi^q_{(j)}$ the output-specific j-th order statistic of the observations ψ^q_k (for $j = 1, \ldots, M^q_y$);
   (d) Compute $\widehat{\theta}^q_{\alpha_q}(y^q, \mathbf{x}^q)$:

$$\widehat{\theta}^q_{\alpha_q}(y^q, \mathbf{x}^q) = \begin{cases} \psi^q_{((1-\alpha_q)M^q_y)}, \text{ if } (1-\alpha_q)M^q_y \in \mathbb{N} \\ \psi^q_{(((1-\alpha_q)M^q_y)+1)}, \text{ otherwise} \end{cases} ;$$

2. Compute $\widehat{\theta}_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \max_{q=1,\ldots,Q} \left\{ \widehat{\theta}^q_{\alpha_q}(y^q, \mathbf{x}^q) \right\}$.

(RTN-1) is used to reduce the computer computations. Also, **(RTN-1)** can be seen as a natural extension of the routine suggested by Daraio and Simar (2007b) for the multi-output context. Indeed, the first step has to be done for every output separately, and in step (b), we find the inputs allocated to output q. When the production process consists of one output, both routines coincide. Finally, we point out that this routine can also be used to compute $\widehat{\theta}_{FDH}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ by letting $\boldsymbol{\alpha} = \mathbf{1}$.

Next, to obtain the nonparametric estimator of the order$-\mathbf{m}$ multi-output efficiency measurement, we first have to compute the output-specific efficiency measurements obtained for random sub-samples of size m_q. These random sub-samples are defined using the empirical output-specific conditional distribution function $\widehat{F}^q_{\mathbf{X}^q|Y^q}(. \mid y^q)$. We obtain that the random output-specific efficiency measurement for output q of a DMU operating at $(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ is given by:

$$\widehat{\eta}^q_{m_q}(y^q, \mathbf{x}^q) = \inf \left\{ \eta^q \mid (\eta^q \mathbf{x}^q, y^q) \in \widehat{T}^q_{m_q} \right\}, \tag{25}$$

where $\widehat{T}^q_{m_q}$ is the production possibility set obtained for m_q *iid* random DMUs obtained using the output-specific conditional distribution function $\widehat{F}^q_{\mathbf{X}^q|Y^q}(. \mid y^q)$.

Therefore, the nonparametric estimator of the order-α_q is obtained by taking the expectation of these random output-specific efficiency measurements:

$$\begin{aligned} \widehat{\theta}^q_{m_q}(y^q, \mathbf{x}^q) &= \widehat{E}\left(\widehat{\eta}^q_{m_q}(y^q, \mathbf{x}^q) \mid Y^q \geq y^q \right), & (26) \\ &= \int_0^\infty \left(1 - \widehat{F}^q_{\mathbf{X}^q|Y^q}(u\mathbf{x}^q \mid y^q) \right)^{m_q} du. & (27) \end{aligned}$$

Thus, the nonparametric estimator of the order$-\mathbf{m}$ multi-output efficiency measurement is obtained by taking the maximum of the nonparametric estimators of the order$-m$ efficiency measurements:

$$\theta_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \max_{q=1,\ldots,Q} \left\{ \widehat{\theta}^q_{m_q}(y^q, \mathbf{x}^q) \right\}, \tag{28}$$

As done in (15), we can connect the nonparametric estimators of $\widehat{\theta}^q_{m_q}(y^q, \mathbf{x}^q)$ and

$\widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q)$ as follows:

$$\widehat{\theta}^q_{m_q}(y^q, \mathbf{x}^q) = \widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q) + \int_{\widehat{\theta}^q_{FDH}(y^q,\mathbf{x}^q)}^{\infty} \left(1 - \widehat{F}^q_{\mathbf{X}^q|Y^q}(u\mathbf{x}^q \mid y^q)\right)^{m_q} du. \quad (29)$$

Thus, when $m_q \to \infty$, the integral vanishes making $\widehat{\theta}^q_{m_q}(y^q, \mathbf{x}^q) \to \widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q)$. When $m_q \to \infty$ for every output q, we have that $\theta_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) \to \widehat{\theta}^q_{FDH}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$. As such, $\theta_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ is not defined assuming convexity. The convex counterpart for the nonparametric estimator is defined in the Appendix.

In practice solving the integral may be difficult. Attractively, as for the order$-\boldsymbol{\alpha}$ multi-output efficiency measurement, a routine can be used to compute the order$-\mathbf{m}$ multi-output efficiency measurement **(RTN-2)**:

1. For each output $q = 1, \ldots, Q$:
   (a) For $b = 1, \ldots, B$:
      i. Draw a sample of size m_q with replacement among those input levels such that $Y^q \geq y^q$ and denote those inputs by $(\mathbf{x}^q_{1b}, \ldots, \mathbf{x}^q_{m_q b})$;
      ii. Compute $\widehat{\eta}^{q,b}_{m_q}(y^q, \mathbf{x}^q)$ by solving the mixed integer linear program **(MILP-2)**:

$$\widehat{\eta}^{q,b}_{m_q}(y^q, \mathbf{x}^q) = \min_{\lambda^q_k\ (k \in \{1,\ldots,m_q\})} \left(\eta^q \;\middle|\; \begin{array}{l} \sum_{k=1}^{m_q} \lambda^q_k \mathbf{x}^q_{kb} \leq \eta^q \mathbf{x}^q, \sum_{k=1}^{m_q} \lambda^q_k = 1, \\ \forall k : \lambda^q_k \in \{0; 1\}, \eta^q \geq 0 \end{array} \right);$$

   (b) Compute $\widehat{\theta}^q_{m_q}(y^q, \mathbf{x}^q) = \frac{1}{B} \sum_{b=1}^{B} \widehat{\eta}^{q,b}_{m_q}(y^q, \mathbf{x}^q)$;
2. Compute $\widehat{\theta}_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \max_{q=1,\ldots,Q} \left\{ \widehat{\theta}^q_{m_q}(y^q, \mathbf{x}^q) \right\}$.

We can see **(RTN-2)** as a natural extension of the routine suggested by Daraio and Simar (2007b) for the multi-output context. Indeed, the first step has to be done for every output separately, and in step (a), there are the inputs allocated to output q. Note that when there is only one output, our routines coincide.

3 Simulation and empirical application

To demonstrate the benefits of the new robust multi-output efficiency measurements for multi-output contexts, we provide both a simulation and an empirical application for US electricity plants.[6]

3.1 Simulation

We consider a setting with 3 outputs (y^1, y^2, and y^3) and 1 input (x) used jointly to produce the 3 outputs. We only present a simulation when considering more than one output since, as discussed previously, when the production process consists of only one output, our method coincides with the well-established methods for detecting and correcting for the presence of outliers. We define the three outputs as follows:

$$y^q = (x)^\beta * \exp(-u^q), \text{ for } q = 1, 2, 3, \tag{30}$$

where $\beta = 0.5$, x is uniform on $(0, 1)$, and u^q is an exponential with mean of $\frac{1}{3}$. This simulation is very similar to the scenario proposed in Simar (2003), but extends to the multi-output setting. We generate 100 DMUs, and add the three outliers described in Table 1.

Outlier	x	y^1	y^2	y^3
A	0.5	0.95	0.3	0.4
B	0.6	0.9	0.95	0.5
C	0.7	0.95	0.90	0.95

Table 1: Inputs and outputs for the outliers

The three outliers present different profiles. Outlier A is extreme only in the production of output 1; outlier B is extreme in the production of output 1 and 2; and outlier C is extreme in the production of the three outputs.

As explained in Section 2.1, an advantage of the multi-output approach is to allocate the inputs to the output production process. For our simulation, we consider the simplest case when only one input is used to produce all the outputs. Note that we also consider that simple case to obtain a fair comparison with the procedure when output-specific production sets are not considered (see Table 2). We obtain that $x^1 = x$, $x^2 = x$, and $x^3 = x$ using (1). In words, the input x is used to

[6] We make use of Matlab to compute the nonparametric estimators for both the simulation and the empirical application. Note that the computations are very fast. This also comes from the output-specific nature of the robust multi-output efficiency measurements. We only have to solve problems with one output and several inputs. See also Simar (2003) for more detail about the computation time of robust efficiency measurements.

produce the three outputs simultaneously, and thus appears in the three output-specific production processes.

Following Simar (2003), we compute the leave-one-out order-$\mathbf{m}$ multi-output efficiency measurement and order-m_q efficiency measurements, denoted by $\widehat{\theta}^{(k)}_{\mathbf{m}}(\mathbf{y}, x^1, x^2, x^3)$ and $\widehat{\theta}^{q,(k)}_{m_q}(y^q, x^q)$, for $q = 1, 2, 3$, respectively; and define an outlier for the overall (output q) production process a DMU that presents leave-one-out efficiency measurements larger than 1 when $\mathbf{m}$ (m_q) increases. We choose to compute the leave-one-out order-$\mathbf{m}$ multi-output efficiency measurements for $\mathbf{m} = \mathbf{10}$, $\mathbf{m} = \mathbf{25}$, $\mathbf{m} = \mathbf{50}$, $\mathbf{m} = \mathbf{75}$, $\mathbf{m} = \mathbf{100}$, and $\mathbf{m} = \mathbf{150}$; but also when considered each output individually, i.e. $m_q = 10$, $m_q = 25$, $m_q = 50$, $m_q = 75$, $m_q = 100$, and $m_q = 150$, for $q = 1, 2, 3$. Finally, we rely on $B = 200$ as in Simar (2003). Note that we obtain similar results with higher value for B. The efficiency results for the three outliers using **(RTN-2)** are displayed in Table 2. For comparison purposes, we also compute the leave-one-out order$-m$ efficiency measurements when not considering output-specific production possibility sets (i.e. the ones defined by Cazals, Florens and Simar (2002) and Daraio and Simar (2005)), denoted by $\widehat{\theta}^{(k)}_{m}(\mathbf{y}, x)$.

For the aggregate production level, i.e. using $\widehat{\theta}^{(k)}_{\mathbf{m}}(\mathbf{y}, x^1, x^2, x^3)$, we find that A, B, and C are outliers. A similar conclusion can be made using $\widehat{\theta}^{(k)}_{m}(\mathbf{y}, x)$. While this observation represents valuable information, it does not say much about the profiles of the outliers. Are they outliers for the overall production process? Or only for one or several outputs? Of course, using Table 1, we can answer those questions, but for more settings with several inputs (and different types of input allocation), it is more complex.

The distinguishing feature of our method is captured by $\widehat{\theta}^{1,(k)}_{m_1}(y^1, x^1)$, $\widehat{\theta}^{2,(k)}_{m_2}(y^2, x^2)$, and $\widehat{\theta}^{3,(k)}_{m_3}(y^3, x^3)$. These extra three efficiency measurements give us the option to investigate for outlier behaviour for output 1, 2, and 3, respectively. As such, they allow us to distinguish the profiles of the outliers. $\widehat{\theta}^{1,(k)}_{m_1}(y^1, x^1)$ reveals that A, B, and C are outliers. Next, $\widehat{\theta}^{2,(k)}_{m_2}(y^2, x^2)$ shows that B, and C are outliers for the production of output 2. As such, efficiency results can be provided for A for that output. In fact, A is inefficient for the production of output 2. Finally, $\widehat{\theta}^{3,(k)}_{m_3}(y^3, x^3)$ suggests that only C is an outlier for output 3; while B is efficient, and A is close to the efficient situation. These conclusions are consistent with our observations made from Table 1.

Overall, this simple simulation demonstrates that our new multi-output efficiency measurements provide more flexibility to study the profiles of the outliers. In partic-

Efficiency	Outlier	$m = 10$	$m = 25$	$m = 50$	$m = 75$	$m = 100$	$m = 150$
$\widehat{\theta}_{m}^{(k)}(\mathbf{y}, x)$	A	Inf	Inf	Inf	Inf	Inf	Inf
	B	5.35	5.32	5.31	5.30	5.29	5.28
	C	5.75	5.75	5.75	5.75	5.75	5.75
Efficiency	Outlier	$\mathbf{m = 10}$	$\mathbf{m = 25}$	$\mathbf{m = 50}$	$\mathbf{m = 75}$	$\mathbf{m = 100}$	$\mathbf{m = 150}$
$\widehat{\theta}_{\mathbf{m}}^{(k)}(\mathbf{y}, x^1, x^2, x^3)$	A	Inf	Inf	Inf	Inf	Inf	Inf
	B	4.45	4.25	4.16	4.14	4.12	4.11
	C	5.25	5.23	5.22	5.21	5.20	5.20
Efficiency	Outlier	$m_1 = 10$	$m_1 = 25$	$m_1 = 50$	$m_1 = 75$	$m_1 = 100$	$m_1 = 150$
$\widehat{\theta}_{m_1}^{1,(k)}(y^1, x^1)$	A	Inf	Inf	Inf	Inf	Inf	Inf
	B	4.21	4.18	4.16	4.14	4.12	4.11
	C	5.25	5.23	5.22	5.21	5.20	5.20
Efficiency	Outlier	$m_2 = 10$	$m_2 = 25$	$m_2 = 50$	$m_2 = 75$	$m_2 = 100$	$m_2 = 150$
$\widehat{\theta}_{m_2}^{2,(k)}(y^2, x^2)$	A	0.68	0.67	0.67	0.67	0.67	0.67
	B	4.45	4.25	4.15	4.05	3.98	3.98
	C	4.82	4.75	4.62	4.55	4.55	4.54
Efficiency	Outlier	$m_3 = 10$	$m_3 = 25$	$m_3 = 50$	$m_3 = 75$	$m_3 = 100$	$m_3 = 150$
$\widehat{\theta}_{m_3}^{3,(k)}(y^3, x^3)$	A	0.98	0.95	0.92	0.92	0.92	0.92
	B	1	1	1	1	1	1
	C	4.67	4.25	4.17	4.05	4.00	4.00

Table 2: Multi-output efficiency results for the outliers

ular, using our method, we are able to provide more valuable information about the efficiency behaviours, and thus better use the data. This is highlighted, for example, when looking at A. With our method, we can conclude that, indeed, A is an outlier but only for the production of output 1, while A is not efficient for output 2, and close to being efficient for output 3.

We end this simulation with three remarks. First, we note that besides A, B, and C, other DMUs have been pointed out as potential outliers (in fact, 10 observations in total). We do not report those cases here, as the aim of this simulation is to demonstrate the usefulness of our technique, and its greater flexibility. Second, we note that our conclusions about A, B, and C remain valid when using the order$-\boldsymbol{\alpha}$ multi-output efficiency measurement, and when imposing convexity. Finally, in general, it is better to rely on both input- and output-oriented efficiency measurements when investigating for outliers.

3.2 Empirical application

Benchmarking electric utilities is a popular topic in the nonparametric efficiency literature. See, for example, Yaisawang and Klein (1994), Färe, Grosskopf, Noh and Weber (2005), Sarkis and Cordeiro (2012), Cherchye, De Rock, and Walheer (2015), and Walheer (2018f, g) for analyses of US electric utilities, Goto and Tsutsui (1998), Hattori (2002) and Tone and Tsutsui (2007) for analyses of both Japanese and US electric utilities, and Korhonen and Luptacik (2004) for an analysis of European electric utilities. Generally, the input orientation is mostly chosen in this context. Indeed, it may reasonably be assumed that the electricity generated is exogenously defined, which means that the size of the electricity market (or number of consumers) falls beyond control of the electric utilities. Nevertheless, the plants can still minimize their inputs (or costs) for a given electricity production.

The major difference of our efficiency analysis is that we recognize the multi-output nature of the electricity plants. That is, we do not consider the electricity generation as a whole, but rather split the electricity generation into renewable (e.g. wind, solar, geothermal) and non-renewable (e.g. coal, oil, gas) electricity generation. As a result, we can allocate the inputs to each type of electricity generation. In general, two main inputs are selected when electricity plants are of interest: nameplate capacity (used as a proxy for total assets), and the quantity of fuel used. Clearly, these two inputs are used not in the same manner to generate electricity. Indeed, while nameplate capacity is used in the production of all types, the fuel quantity is only used for non-renewable electricity. Nameplate capacity is thus a joint input, and the fuel quantity is an output-specific input completely allocated to the non-renewable electricity production.[7]

Our two new robust multi-output efficiency measurements are particularly useful for analyzing the electricity plants. Indeed, plants may differ in the proportion of renewable/non-renewable electricity generation they produce. As such, there is no reason why outliers should be the same for each type of electricity, and why an outlier could not be extreme only for the production of one type of electricity. Using our two new measurements, we are able to identify extreme behaviours for every electricity

[7] At this point, it is useful to remark that additional inputs (as the total number of employees, the generator capacity and the boiler capacity) and undesirable by-products of the use of fuel as input (as SO_2, NO_x and CO_2 emissions) are also considered by some studies. For simplicity and to match with all previous studies, we do not consider these additional inputs and undesirable by-products for our empirical study.

generation, and thus better understand why a plant is categorized as an outlier.

To present our application, we first discuss the specificities of our set-up, and present the data. Subsequently, we present the results of the empirical application.

Data and input and output selection. We have taken the data from the *eGRID* system that was developed by the Environmental Protection Agency of the United States. *eGRID* stands for a comprehensive source of data of all electricity plants in the US. In particular, we use the *eGRID* 2012 version 1.0, and concentrate our analysis on plants that are multi-output producers. That is, in this context, plants that produce both renewable (y^1) and non-renewable electricity (y^2). Also, we have two inputs: nameplate capacity (x_1) and quantity of fuel used (x_2). Nameplate capacity is a joint input and the fuel quantity is an output-specific input completely allocated to the production of non-renewable electricity. Using the notations of Section 2.1, we have for each plant k:

$$\mathbf{y}_k = \begin{bmatrix} y_k^1 \\ y_k^2 \end{bmatrix}, \mathbf{x}_k = \begin{bmatrix} x_{1k} \\ x_{2k} \end{bmatrix}, \alpha_2^1 = 0, \alpha_2^2 = 1,$$

$$\mathbf{x}_k^1 = \begin{bmatrix} x_{1k} \\ \alpha_2^1 * x_{2k} \end{bmatrix} = \begin{bmatrix} x_{1k} \\ 0 \end{bmatrix}, \text{ and } \mathbf{x}_k^2 = \begin{bmatrix} x_{1k} \\ \alpha_2^2 * x_{2k} \end{bmatrix} = \begin{bmatrix} x_{1k} \\ x_{2k} \end{bmatrix}. \tag{31}$$

We end with a sample of 343 plants. Table 3 reports the corresponding descriptive statistics for the different inputs and outputs.

	Outputs		Inputs	
	Renewable Energy (MWh)	Non-Renewable Energy (MWh)	Nameplate Capacity (MW)	Fuel (MMBtu)
Min	1.43	1.29	1.2	49
Mean	122,350	175,020	77.51	2,900,600
Max	607,280	8,474,234	1,755	82,691,000
Std	135,400	657,080	154.97	6,662,200

Table 3: Descriptive statistics for the 343 plants

As expected, multi-output plants produce, on average, more non-renewable than renewable electricity. Also, the minimum, maximum and standard deviation for the outputs and the inputs reveal the presence of heterogeneity in the sample, indicating that outliers might be an issue for the efficiency analysis. Also, the amplitude of the heterogeneity is not the same for every input/output. As such, the number and the profile of outliers could be different for each type of electricity.

Multi-output efficiency results. We compute the nonparametric estimators for the multi-output efficiency measurement without and with assuming convexity of the technology, denoted by $\widehat{\theta}_{FDH}(\mathbf{y}, \mathbf{x}^1, \mathbf{x}^2)$ and $\widehat{\theta}_{DEA}(\mathbf{y}, \mathbf{x}^1, \mathbf{x}^2)$, using **(MILP-1)** and **(LP-1)**, respectively. For comparison purposes, we also compute two extra efficiency measurements: $\widehat{\theta}_{FDH}(\mathbf{y}, \mathbf{x})$ and $\widehat{\theta}_{DEA}(\mathbf{y}, \mathbf{x})$. These two efficiency measurements are the standard FDH and DEA efficiency measurements of Deprins, Simar and Tulkens (1984) and Charnes, Cooper and Rhodes (1978), respectively. Standard in the sense that they do not consider output-specific production possibility sets, nor the allocation of the inputs. In practice, they are also computed by mixed integer linear programming and linear programming. We refer to their paper for more detail.[8] We present, in Table 4, the descriptive statistics of the computed efficiency measurements.

Estimator	Min	Mean	Median	Max	St. dev.	# efficient	% efficient
$\widehat{\theta}_{FDH}(\mathbf{y}, \mathbf{x})$	0.1154	0.8977	1	1	0.1944	219	63.85
$\widehat{\theta}_{DEA}(\mathbf{y}, \mathbf{x})$	0.1154	0.8864	1	1	0.2020	207	60.35
$\widehat{\theta}_{FDH}(\mathbf{y}, \mathbf{x}^1, \mathbf{x}^2)$	0.0736	0.7417	0.7568	1	0.2415	101	29.45
$\widehat{\theta}_{DEA}(\mathbf{y}, \mathbf{x}^1, \mathbf{x}^2)$	0.0710	0.7326	0.7381	1	0.2423	93	27.11

Table 4: Multi-output efficiency results

The main conclusion from those results is that the multi-output efficiency measurements have a greater ability to detect inefficiency behaviour. Clearly, this is supported by the descriptive statistics: the minimums, means and medians are smaller when using the multi-output efficiency measurements, meaning that more inefficient behaviours have been detected. As a consequence, fewer plants are declared as efficient. Importantly, this fact holds true irrespective of whether convexity is assumed or not for the technology. Note that the difference between the FDH and DEA estimators are only due to the extra assumption of convexity of DEA.

[8] Interestingly, the more standard FDH and DEA estimators could also be obtained as a particular case of **(MILP-1)** and **(LP-1)** by imposing that $\lambda_k^q = \lambda_k, \forall q$, i.e. a common λ for every output q.

Attractively, as explained in Section 2.2, the multi-output efficiency measurements can also be used to investigate for (in)efficient behaviour at the output (electricity in this context) level. Therefore, it offers the option to better understand the efficiency results at the aggregate production level, i.e. for all electricity generations together. Table 5 gives the results of the output-specific efficiency measurements when convexity is not assumed ($\widehat{\theta}^1_{FDH}(y^1, \mathbf{x}^1)$ and $\widehat{\theta}^2_{FDH}(y^2, \mathbf{x}^2)$), and when convexity is assumed ($\widehat{\theta}^1_{DEA}(y^1, \mathbf{x}^1)$ and $\widehat{\theta}^2_{DEA}(y^2, \mathbf{x}^2)$). The results highlight that only a few plants produce both outputs efficiently (seven in the non-convex setting and three in the convex setting). Also, plants are on average more efficient for the production of non-renewable electricity (y^2) than for the production of renewable electricity (y^1). A possible explanation for these results is that plants have been producing non-renewable electricity for decades, while the production of renewable electricity is more recent. See also Table 6 for analysis of specific plants.

Estimator	Min	Mean	Median	Max	St. dev.	# efficient	% efficient
$\widehat{\theta}^1_{FDH}(y^1, \mathbf{x}^1)$	0.0736	0.7102	0.7225	1	0.2073	35	10.20
$\widehat{\theta}^2_{FDH}(y^2, \mathbf{x}^2)$	0.0651	0.8122	0.8661	1	0.3329	73	21.28
$\widehat{\theta}^1_{DEA}(y^1, \mathbf{x}^1)$	0.0710	0.6973	0.7108	1	0.2870	34	9.91
$\widehat{\theta}^2_{DEA}(y^2, \mathbf{x}^2)$	0.0631	0.7820	0.7761	1	0.2911	62	18.08

Table 5: Output-specific efficiency results

Outlier detection. To detect the outliers in the sample, we make use of the same procedure as the one used for the simulation in Section 3.1. Thus, we compute the leave-one-out order-$\mathbf{m}$ multi-output efficiency measurements for several values of $\mathbf{m}$, but also the leave-one-out order-m_1 and order-m_2 efficiency measurements for several values of m_1 and m_2 (we have two outputs in the production process). A plant is declared as an outlier for the overall production, output 1, and output 2, when the leave-one-out efficiency measurements remain larger than 1 when $\mathbf{m}$, m_1, or m_2 increase, respectively. For comparison purposes, we also compute the number of outliers with the more standard procedure (i.e. based on m). We use **(RTN-2)** for the computational aspects.

The results are presented in Table 6. Both methods find a similar number of outliers for the overall production processes: 21 and 25. The difference between both methods may be attributed to the allocation of the inputs to the electricity generation process when considering the multi-output efficiency measurement. That is, by increasing the realism of the modelling of the plant production process by allocating the inputs to the outputs, more outliers are discovered. The advantage of our method is captured by $\widehat{\theta}^1_{m_1}(y^1, \mathbf{x}^1)$ and $\widehat{\theta}^2_{m_2}(y^2, \mathbf{x}^2)$. Based on these two output-specific efficiency measurements, we find that there are 14 outliers for the production of renewable electricity, and 13 for the production of non-renewable electricity. It implies that only two plants are outliers for the productions of both outputs. This is clearly less than 21 when relying on $\widehat{\theta}_m(\mathbf{y}, \mathbf{x})$.

Estimator	# outliers	% outliers
$\widehat{\theta}_m(\mathbf{y}, \mathbf{x})$	21	6.12
$\widehat{\theta}_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \mathbf{x}^2)$	25	7.29
$\widehat{\theta}^1_{m_1}(y^1, \mathbf{x}^1)$	14	4.08
$\widehat{\theta}^2_{m_2}(y^2, \mathbf{x}^2)$	13	3.79

Table 6: Number of outliers

Robust multi-output efficiency results. We present in Table 7, the nonparametric estimators of the robust multi-output efficiency measurements. As expected, it is confirmed that the multi-output efficiency measurements better identify inefficiency behaviour of the plants. Also, it is confirmed that plants are on average more efficient in the production of non-renewable electricity.

Estimator	Min	Mean	Median	Max	St. dev.	# efficient	% efficient
$\widehat{\theta}_m(\mathbf{y}, \mathbf{x})$	0.1157	0.9024	1	1.02	0.2054	248	72.30
$\widehat{\theta}_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \mathbf{x}^2)$	0.1295	0.7744	0.7722	1.09	0.2568	122	35.57
$\widehat{\theta}^1_{m_1}(y^1, \mathbf{x}^1)$	0.1295	0.7516	0.7672	1.09	0.2058	56	16.33
$\widehat{\theta}^2_{m_2}(y^2, \mathbf{x}^2)$	0.0851	0.8594	0.8811	1.07	0.2790	90	26.24

Table 7: Robust multi-output efficiency results

To illustrate, once more, the benefit of the output-specific modelling, we report in Table 8 the decomposition for selected plants. Plant # 19 has a nonparametric estimator of 0.5569 when considering the multi-output model, while it increases to 0.7059 when the multi-output modelling is not considered. It reveals the higher ability to detect inefficient behaviour of our method. The output-specific nonparametric estimators highlight that Plant # 19 is more efficient for the production of non-renewable electricity. Plant # 87 is an outlier for the production of non-renewable electricity, but not for the production of renewable electricity (0.6829). Clearly, both overall efficiency measurements confirm the outlier profile of this plant. Using the robust output-specific efficiency measurements, we learn that it is because of output 2. In a similar vein, we learn from Table 8 that Plant # 201 is an outlier for both types of electricity generation, while Plant # 341 is an outlier only for output 1.

Plant	$\widehat{\theta}^1_{m_1}(y^1, \mathbf{x}^1)$	$\widehat{\theta}^2_{m_2}(y^2, \mathbf{x}^2)$	$\widehat{\theta}_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \mathbf{x}^2)$	$\widehat{\theta}_m(\mathbf{y}, \mathbf{x})$
# 19	0.4072	0.5569	0.5569	0.7059
# 87	0.6829	(Outlier)	(Outlier)	(Outlier)
# 201	(Outlier)	(Outlier)	(Outlier)	(Outlier)
# 321	(Outlier)	0.5325	(Outlier)	(Outlier)

Table 8: Robust multi-efficiency results for selected plants

This illustration shows how the method could be used to provide more valuable information about both the (in)efficiency behaviour of the plants, and their outlier profile.

4 Conclusion

Nonparametric efficiency measurements have demonstrated their usefulness to detect inefficient behaviours of decisions making units leading to potential cost reductions/profit improvements. These methods are particularly attractive for practitioners since no unverifiable parametric assumptions are required about the technology. Nevertheless, in multi-output settings, practitioners sometimes have doubts about the fairness of such methods. Two limitations could give credit to their doubts. Firstly, the production process is modelled as a "black box", i.e. it is implicitly assumed that all the inputs produce simultaneously all the outputs. Next, only techniques investigating for outliers in all output directions simultaneously exist.

In this paper, we suggested two new nonparametric robust efficiency measurements that tackle these two limitations. The new measurements present several attractive features. First, they increase the realism of the modelling by taking the links between inputs and outputs into account. Second, they give more flexibility for the outlier detection exercise, as they have the ability to detect outliers in each output direction.

Three, they have, in general, a greater ability to detect inefficient behaviour. Four, they can be seen as natural extensions in multi-output contexts of well-established procedures dealing with outliers in efficiency analysis. All in all, the two new nonparametric robust efficiency measurements better use the available information contained in the data. We proposed both a simulation and an empirical application to the electrity plants in the United States to demonstrate the usefulness of our method.

References

- Aragon Y., Daouia A., Thomas-Agnan C. (2005). "Nonparametric frontier estimation: A conditional quantile-based approach", *Econometric Theory* 21, 358-389.
- Banker R. D., Chang H. (2006). "The super-efficiency procedure for outlier identification, not for ranking efficient units", *European Journal of Operational Research* 175, 1311-1320.
- Banker R. D., Cooper W. W., Seiford L. M., Thrall R. M., Zhu J. (2004). "Returns to scale in different DEA models", *European Journal of Operational Research* 154, 345-362.
- Bogetoft P. (1996). "DEA on relaxed convexity assumptions", *Management Science* 42, 457-465.
- Briec W., Kerstens K., Vanden Eeckaut P. (2004). "Non-convex Technologies and Cost Functions: Definitions, Duality and Nonparametric Tests of Convexity", *Journal of Economics* 81, 155-192.
- Cazals, C., Florens J. P., Simar L. (2002). "Nonparametric frontier estimation: A robust approach", *Journal of Econometrics* 106, 1-25.
- Charnes A., Cooper W. W., Rhodes E. (1978). "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research* 2, 429-444.
- Chen W.-C., Johnson A. L. (2010). "A unified model for detecting efficient and inefficient outliers in data envelopment analysis", *European Journal of Operational Research* 37, 417-425.
- Cherchye L., De Rock B., Dierynck B., Roodhooft F., Sabbe J. (2013). "Opening the Black Box of Efficiency Measurement: Input Allocation in Multi-Output Settings", *Operations Research* 61, 1148-1165.
- Cherchye L., De Rock B., Vermeulen F. (2008). "Analyzing Cost-Efficient Production Behavior Under Economies of Scope: A Nonparametric Methodology", *Operations Research* 56 (204-221.
- Cherchye L., De Rock B., Walheer B. (2015). "Multi-output efficiency with good and bad outputs", *European Journal of Operational Research* 240, 872-881.

- Cherchye L., De Rock B., Walheer B. (2016). "Multi-Output Profit Efficiency And Directional Distance Functions", *Omega* 61, 100-109.
- Cook W. D., Seiford L. M. (2009). "Data Envelopment Analysis (DEA) - Thirty years on", *European Journal of Operational Research* 192, 1-17.
- Cooper W. W., Seiford L. M., Tone K. (2007). "Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software, Second Edition", *Springer Edition.*
- Cooper W. W., Seiford L. M., Zhu J. (2004). "Handbook on Data Envelopment Analysis, Second Edition", *Springer Edition.*
- Daouia A., Gijbels I. (2011). "Robustness and inference in nonparametric partial frontier modeling", *Journal of Econometrics* 161, 147-165.
- Daouia A., Simar L. (2005). "Robust nonparametric estimators of monotone boundaries", *Journal of Multivariate Analysis* 96, 311-331.
- Daouia A., Simar L. (2007). "Nonparametric efficiency analysis: A multivariate conditional quantile approach", *Journal of Econometrics* 140, 375-400.
- Daraio C., Simar L. (2005). "Introducing environmental variables in nonparametric frontier models: A probabilistic approach", *Journal of Productivity Analysis* 24, 93-121.
- Daraio C., Simar L. (2007a). "Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications", *Springer.*
- Daraio C., Simar L. (2007b). "Conditional Nonparametric Frontier Models for Convex and Non Convex Technologies: A unifying Approach", *Journal of Productivity Analysis* 28, 13-32.
- Debreu G. (1951). "The Coefficient of Resource Utilization", *Econometrica* 19(3), 273-292.
- Deprins D., Simar L., Tulkens H. (1984). "Measuring labor inefficiency in post offices", In *The Performance of Public Enterprises: Concepts and Measurements. Amsterdam: North-Holland, Marchand M., Pestieau P. and H. Tulkens (eds.)*, 243-267.

- Despic O., Despic M., Paradi J. (2007). "DEA-R: ratio-based comparative efficiency model, its mathematical relation to DEA and its use in applications", *Journal of Productivity Analysis* 28, 33-44.

- Ding J., Dong W., Liang L., Zhu J. (2017). "Goal congruence analysis in multi-Division Organizations with shared resources based on data envelopment analysis", *European Journal of Operational Research* 263, 961-973.

- Färe R., Grosskopf S. (2000). "Network DEA", *Socio-Economic Planning Sciences* 34, 35-49.

- Färe R., Grosskopf S., Lovell C. A. K. (1994). "Production Frontier", *Cambridge University Press.*

- Färe R., Grosskopf S., Lovell C. A. K., Pasurka C. (1989). "Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach", *The Review of Economics and Statistics* 71, 90-98.

- Färe R., Grosskopf S., Noh D. W., Weber W. (2005). "Characteristics of a polluting technology: theory and practice", *Journal of Econometrics* 126, 469-492.

- Färe R., Grosskopf S., Whittaker G. (2007). "Network DEA", *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis, J. Zhu and W. Cook, Eds. Springer.*

- Färe R., Karagiannis G. (2017). "The denominator rule for share-weighting aggregation", *European Journal of Operational Research* 260, 1175-1180.

- Färe R., Zelenyuk (2003). "Aggregation of Cost Efficiency: Indicators and Indexes Across Firms", *European Journal of Operational Research* 146, 615-620.

- Farrell M. (1957). "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society* Series 1, General, 120, Part 3, 253-281.

- Fried H., Lovell C. A. K., Schmidt S. (2008). "The Measurement of Productive Efficiency and Productivity Change", *Oxford University Press.*

- Goto M., Tsutsui M. (1998). "Comparison of Productive and Cost Efficiencies Among Japanese and US Electric Utilities", *Omega* 2, 177-194.

- Hattori T. (2002). “Relative performance of US and Japanese electricity distribution: an application of stochastic frontier analysis”, *Journal of Productivity Analysis* 18, 269-284.

- Huang C.-W., Ting C.-T., Lin C.-H., Lin C.-T. (2013). “Measuring non-convex metafrontier efficiency in international tourist hotels”, *Journal of the Operational Research Society* 64, 250-259.

- Johnson A. L., McGinnis L. F. (2008). “Outlier detection in two-stage semiparametric DEA models”,*European Journal of Operational Research* 187, 629-635.

- Kerstens K., Vanden Eeckaut P. (1999). “Estimating returns to scale using non-parametric deterministic technologies: A new method based on goodness-of-fit”, *European Journal of Operational Research* 113(1): 206-214.

- Korhonen P., Luptacik M. (2004). “Eco-efficiency of power plants: An extension of data envelopment analysis”, *European Journal of Operational Research* 154, 437-446.

- Kuosmanen T., Cherchye L., Sipilainen T. (2006). “The law of one price in data envelopment analysis: restricting weight flexibility across firms”, *European Journal of Operational Research* 170 , 735-757.

- Leleu H. (2009). “Mixing DEA and FDH models together”, *Journal of Operational Research Society* 60(12), 1730-1739.

- Petersen N. C. (1990). “Data envelopment analysis on a relaxed set of assumptions”, *Management Science* 36, 305-314.

- Podinovski V. V. (2004a). “On the linearization of reference technologies for testing returns to scale in FDH models”, *European Journal of Operational Research* 152(3), 800-802.

- Podinovski V. V. (2004b). “Local and global returns to scale in performance measurement”, *Journal of Operational Research Society* 55(2): 170-178.

- Podinovski V. V., Kuosmanen T. (2011). “Modelling weak disposability in data envelopment analysis under relaxed convexity assumptions”, *European Journal of Operational Research* 211 (3), 577-585.

- Porembski M., Breitenstein K., Alpar P. (2005). "Visualizing efficiency and reference relations in data envelopment analysis with an application to the branches of a German bank", *Journal of Productivity Analysis* 23, 203-221.

- Salerian, J. and Chan C. (2005). "Restricting multiple-output multiple-input DEA models by disaggregating the output-input vector", *Journal of Productivity Analysis* 24, 5-29.

- Sarkis J., Cordeiro J. J. (2012). "Ecological modernization in the electrical utility industry: An application of a bads-goods DEA model of ecological and technical efficiency", *European Journal of Operational Research* 219, 386-395.

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

- Simar L. (2003). "Detecting outliers in frontier models: A simple approach", *Journal of Productivity Analysis* 20, 391-424.

- Simar L., Vanhems A. (2012). "Probabilistic characterization of directional distances and their robust versions", *Journal of Econometrics* 166, 342-354.

- Simar L., Vanhems A., Wilson W. (2012). "Statistical inference for DEA estimators of directional distances", *European Journal of Operational Research* 220, 853-864.

- Tone K., Tsutsui M. (2007). "Decomposition of cost efficiency and its application to Japanese-US electric utility comparisons", *Socio-Economic Planning Sciences* 41, 91-106.

- Tone K., Tsutsui M. (2009). "Network DEA: A slacks-based measure approach", *European Journal of Operational Research* 197, 243-252.

- Tran N. A., Shively G., Preckel P. (2010). "A new method for detecting outliers in Data Envelopment Analysis", *Applied Economics Letters* 17, 313-316.

- Walheer B. (2016a). "A multi-sector nonparametric production-frontier analysis of the economic growth and the convergence of the European countries", *Pacific Economic Review*, 21(4), 498-524.

- Walheer B. (2016b). "Growth and Convergence of the OECD countries: A Multi-Sector Production-Frontier Approach", *European Journal of Operational Research* 252, 665-675.
- Walheer B. (2018a). "Aggregation of metafrontier technology gap ratios: the case of European sectors in 1995-2015", *European Journal of Operational Research* 269, 1013-1026.
- Walheer B. (2018b). "Cost Malmquist productivity index: an output-specific approach for group comparison", *Journal of Productivity Analysis* 49(1), 79-94.
- Walheer B. (2018c). "Disaggregation of the Cost Malmquist Productivity Index with joint and output-specific inputs", *Omega* 75, 1-12.
- Walheer B. (2018d). "Economic growth and greenhouse gases in Europe: A non-radial multi-sector nonparametric production-frontier analysis", *Energy Economics* 74, 51-62.
- Walheer B. (2018e). "Is constant returns-to-scale a restrictive assumption for sector-level empirical macroeconomics? The case of Europe", *Applied Economics Letters* doi.org/10.1080/13504851.2018.1458186
- Walheer B. (2018f). "Malmquist productivity index for multi-output producers: An application to electricity generation plants", *Socio-Economic Planning Sciences* doi.org/10.1016/j.seps.2018.02.003
- Walheer B. (2018g). "Scale efficiency for multi-output cost minimizing producers: the case of the US electricity plants", *Energy Economics* 70, 26-36.
- Walheer B. (2018h), Scale, congestion, and technical efficiency of European countries: a sector-based nonparametric approach", *Empirical Economics* doi.org/10.1007/s00181-018-1426-7
- Walheer B., Zhang L.-J. (2018). "Profit Luenberger and Malmquist-Luenberger indexes for multi-activity decision making units: the case of the star-rated hotel industry in China", *Tourism Management* 69, 1-11.
- Wheelock C. D., Wilson P. W. (2008). "Non-parametric, unconditional quantile estimation for efficiency analysis with an application to Federal Reserve check processing operations", *Journal of Econometrics* 145 (209-225.

- Wilson P. W. (1993). “Detecting outliers in deterministic nonparametric frontier models with multiple outputs”, *Journal of Business and Economic Statistics* 11, 319-323.
- Wilson P. W. (1995). “Detecting influential observations in data envelopment analysis”, *Journal of Productivity Analysis* 6, 27-45.
- Yaisawarng S., Klein J. D. (1994). “The effects of sulfur-dioxide controls on productivity change in the United States electric-power industry”, *Review of Economics and Statistics* 76, 447-460.
- Zhou D. Q., Wu F., Zhou X., Zhou P. (2016). “Output-specific energy efficiency assessment: A data envelopment analysis approach”, *Applied Energy* 177, 117-126.

Appendix

In this Appendix, we provide the nonparametric estimators of the multi-output efficiency measurement, the order$-\boldsymbol{\alpha}$ multi-output efficiency measurement, and the order$-\mathbf{m}$ multi-output efficiency measurement for a DMU operating at $(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ when convexity (i.e. axiom **A3**) is assumed. We denote these estimators by $\widehat{\theta}_{DEA}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$, $\widehat{\theta}^C_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ and $\widehat{\theta}^C_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$, respectively.

When convexity of the output-specific production possibility sets is assumed, we obtain the DEA estimator of these sets, defined for every output q, as follows:

$$\widehat{T}^q_{DEA} = \left((\mathbf{x}^q, y^q) \in \mathbb{R}^{P+1}_+ \;\middle|\; \begin{array}{l} \sum_{k=1}^n \lambda^q_k \mathbf{x}^q_k \leq \mathbf{x}^q, \sum_{k=1}^n \lambda^q_k y^q_k \geq \lambda^q_k y^q \\ \sum_{k=1}^n \lambda^q_k = 1, \forall k : \lambda^q_k \geq 0 \end{array} \right). \tag{32}$$

As such, the DEA estimator of the multi-output efficiency measurement for a DMU operating at $(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ is given by:

$$\widehat{\theta}_{DEA}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) = \inf\{\theta \mid \forall q : (\theta \mathbf{x}^q, y^q) \in \widehat{T}^q_{DEA}\}, \tag{33}$$

$$= \max_{q=1,\ldots,Q} \left\{ \widehat{\theta}^q_{DEA}(y^q, \mathbf{x}^q) \right\}, \tag{34}$$

where, for every output q: $\widehat{\theta}^q_{DEA}(y^q, \mathbf{x}^q) = \inf\{\theta^q \mid (\theta^q \mathbf{x}^q, y^q) \in \widehat{T}^q_{DEA}\}$. In practice, $\widehat{\theta}^q_{DEA}(y^q, \mathbf{x}^q)$ can be computed, for $q = 1, \ldots, Q$, by solving the linear program **(LP-1)**:

$$\widehat{\theta}^q_{DEA}(y^q, \mathbf{x}^q) = \min_{\lambda^q_k \; (k \in \{1,\ldots,n\})} \left(\theta^q \;\middle|\; \begin{array}{l} \sum_{k=1}^n \lambda^q_k \mathbf{x}^q_k \leq \theta^q \mathbf{x}^q, \sum_{k=1}^n \lambda^q_k y^q_k \geq \lambda^q_k y^q \\ \sum_{k=1}^n \lambda^q_k = 1, \forall k : \lambda^q_k \geq 0, \theta^q \geq 0 \end{array} \right).$$

The interpretation of $\widehat{\theta}_{DEA}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ and $\widehat{\theta}^q_{DEA}(y^q, \mathbf{x}^q)$ are analogous to the interpretation of $\theta(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ and $\theta^q(y^q, \mathbf{x}^q)$. We also point out that, in general, $\widehat{\theta}_{FDH}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q) \geq \widehat{\theta}_{DEA}(\mathbf{y}, \mathbf{x}^1, \ldots, \mathbf{x}^Q)$ and $\widehat{\theta}^q_{FDH}(y^q, \mathbf{x}^q) \geq \widehat{\theta}^q_{DEA}(y^q, \mathbf{x}^q)$ for every q. It reveals that imposing convexity can only decrease the efficiency measurements. This also reveals that imposing this axiom should be well justified. Note that it is similar to say that $\widehat{T}^q_{FDH} \subseteq \widehat{T}^q_{DEA}$.

Also, we remark that the FDH estimators are sufficient to compute the DEA estimators. This property, demonstrated by Daraio and Simar (2007b), is particular useful for our discussion. Indeed, it implies that it is enough to define the nonparametric estimators without assuming convexity. Firstly, the DEA estimators of the output-specific production possibility set can be obtained as the convex hull (denoted by $\mathcal{CH}$) of the FDH estimators:

$$\widehat{T}^q_{DEA} = \mathcal{CH}\left(\widehat{T}^q_{FDH}\right). \tag{35}$$

Next, building on this relationship, the DEA estimator of the multi-output efficiency measurement can be obtained by convexifying the output-specific FDH input efficient input level **(LP-2)**:

$$\widehat{\theta}^q_{DEA}(y^q, \mathbf{x}^q) = \min_{\lambda^q_k \ (k\in\{1,\dots,n\})} \left(\theta^q \;\middle|\; \begin{array}{l} \sum_{k=1}^n \lambda^q_k \widehat{\mathbf{x}}^q_{FDH,k} \leq \theta^q \mathbf{x}^q, \sum_{k=1}^n \lambda^q_k y^q_k \geq \lambda^q_k y^q \\ \sum_{k=1}^n \lambda^q_k = 1, \forall k : \lambda^q_k \geq 0, \theta^q \geq 0 \end{array} \right),$$

where $\widehat{\mathbf{x}}^q_{FDH,k} = \widehat{\theta}^q_{FDH}(y^q_k, \mathbf{x}^q_k) * \mathbf{x}^q_k$ is the FDH-input efficient level of output q. As such, $\widehat{\theta}_{DEA}(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q)$ can be obtained after solving **(LP-1)** for every q, or after solving **(MILP-1)** and **(LP-2)** for every q.

Next, the nonparametric estimator of the order-$\boldsymbol{\alpha}$ efficiency measurement when imposing convexity is defined for a DMU operating at $(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q)$ as follows:

$$\widehat{\theta}^C_{\boldsymbol{\alpha}}(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q) = \max_{q=1,\dots,Q} \left\{ \widehat{\theta}^{q,C}_{\alpha_q}(y^q, \mathbf{x}^q) \right\}, \tag{36}$$

where $\widehat{\theta}^{q,C}_{\alpha_q}(y^q, \mathbf{x}^q)$ is obtained for every output q by solving the following linear program **(LP-3)**:

$$\widehat{\theta}^{q,C}_{\alpha_q}(y^q, \mathbf{x}^q) = \min_{\lambda^q_k \ (k\in\{1,\dots,n\})} \left(\theta^q \;\middle|\; \begin{array}{l} \sum_{k=1}^n \lambda^q_k \widehat{\mathbf{x}}^q_{\alpha_q,k} \leq \theta^q \mathbf{x}^q; \sum_{k=1}^n \lambda^q_k y^q_k \geq \lambda^q_k y^q \\ \sum_{k=1}^n \lambda^q_k = 1, \forall k : \lambda^q_k \geq 0, \theta^q \geq 0 \end{array} \right),$$

where $\widehat{\mathbf{x}}^q_{\alpha_q,k} = \widehat{\theta}^q_{\alpha_q}(y^q_k, \mathbf{x}^q_k) * \mathbf{x}^q_k$ is the α_q-input efficient level of output q.

Finally, the nonparametric estimator of the order-$\mathbf{m}$ efficiency measurement when imposing convexity is defined for a DMU operating at $(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q)$ as follows:

$$\widehat{\theta}^C_{\mathbf{m}}(\mathbf{y}, \mathbf{x}^1, \dots, \mathbf{x}^Q) = \max_{q=1,\dots,Q} \left\{ \widehat{\theta}^{q,C}_{m_q}(y^q, \mathbf{x}^q) \right\}, \tag{37}$$

where $\widehat{\theta}^{q,C}_{m_q}(y^q, \mathbf{x}^q)$ is obtained for every output q by solving the following linear program **(LP-4)**:

$$\widehat{\theta}^{q,C}_{m_q}(y^q,\mathbf{x}^q) = \min_{\lambda^q_k\ (k\in\{1,\dots,n\})} \left(\theta^q \;\middle|\; \begin{array}{l} \sum_{k=1}^n \lambda^q_k \widehat{\mathbf{x}}^q_{m_q,k} \le \theta^q \mathbf{x}^q, \sum_{k=1}^n \lambda^q_k y^q_k \ge \lambda^q_k y^q \\ \sum_{k=1}^n \lambda^q_k = 1, \forall k : \lambda^q_k \ge 0, \theta^q \ge 0 \end{array} \right),$$

where $\widehat{\mathbf{x}}^q_{m_q,k} = \widehat{\theta}^q_{m_q}(y^q_k,\mathbf{x}^q_k) * \mathbf{x}^q_k$ is the m_q-input efficient level of output q.

As a final remark, we point out that when $\alpha_q = 1$, we have that $\widehat{\theta}^{q,C}_{\alpha_q}(y^q,\mathbf{x}^q) = \widehat{\theta}^q_{DEA}(y^q,\mathbf{x}^q)$, and thus when $\boldsymbol{\alpha} = \mathbf{1}$, $\widehat{\theta}^C_{\boldsymbol{\alpha}}(\mathbf{y},\mathbf{x}^1,\dots,\mathbf{x}^Q) = \widehat{\theta}_{DEA}(\mathbf{y},\mathbf{x}^1,\dots,\mathbf{x}^Q)$. In a similar vein, when $m_q \to \infty$, we have that $\widehat{\theta}^{q,C}_{m_q}(y^q,\mathbf{x}^q) \to \widehat{\theta}^q_{DEA}(y^q,\mathbf{x}^q)$, implying that when $\mathbf{m} \to \boldsymbol{\infty}$, $\widehat{\theta}^C_{\mathbf{m}}(\mathbf{y},\mathbf{x}^1,\dots,\mathbf{x}^Q) \to \widehat{\theta}_{DEA}(\mathbf{y},\mathbf{x}^1,\dots,\mathbf{x}^Q)$.