



Centre for Efficiency and Productivity Analysis

**Working Paper Series
No. WP04/2017**

Convexity, Disposability and Returns to Scale in Production Analysis

Manh D. Pham and Valentine Zelenyuk

Date: October 2017

**School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia**

ISSN No. 1932 - 4398

Convexity, Disposability and Returns to Scale in Production Analysis

Manh D. Pham*¹ and Valentin Zelenyuk^{1,2}

¹*School of Economics, The University of Queensland, Australia*

²*Centre for Efficiency and Productivity Analysis, The University of Queensland, Australia*

October 20, 2017

Abstract

Adequate modelling of undesirable outputs is a key aspect for any performance analysis of economic systems. A nonparametric approach assuming jointly weak disposability of desirable and undesirable outputs inspired by Shephard (1974) has gained substantial popularity in addressing this issue. Recently, researchers were offered an alternative that is to use multiple scaling factors (rather than a single one as in the Shephard's (1974) approach) when imposing weak disposability in practice. In this paper we discover new properties and relationships between the two approaches, which in turn sheds some new light on the problem and offers reconciling solutions.

Key words: Data envelopment analysis, Nonparametric productivity analysis, Weak disposability, Undesirable outputs, Environmental performance.

JEL classification: C14, C44, C51, D24, M11

*Corresponding to: ducmanh.pham@uq.net.au

1 Introduction

Modelling undesirable (bad) outputs is an important and challenging task in economics analysis. On the one hand, this might attribute to the modern view on sustainable development which raises the importance of harmonizing economic development with the healthiness of natural systems. On the other hand, from the management viewpoint, it is also necessary to fully account for bad outputs as they might have considerable impacts on production processes.¹ Several examples below well illustrate that modelling undesirable outputs has been strongly motivated in a wide variety of industries.

In 1989, Färe et al., by investigating paper mills operating in the U.S., demonstrated that conventional production analysis may be seriously misleading if they ignore undesirable outputs which are subject to different degrees of regulatory constraint across the dataset. In the banking industry, Park and Weber (2006) claimed that non-performing loans might appear and be written off eventually, raising a need for an efficiency measure to account for these undesirable by-products. Also underlining the importance of accounting for bad outputs in the lending process, Fukuyama and Weber (2010) included nonperforming loans in their network structure model to analyze the performance of Japanese banks. In energy production, a recent survey of Sueyoshi et al. (2017) indicated that the matter of bad outputs receives serious attention from a large number of researchers since pollution, particularly CO_2 emission, can have severe consequences on development of a sustainable society. Zhou et al. (2008) studied the carbon emission performance of eight world regions and found that different types of environmental technologies are likely to have distinguishable effects on efficiency measurement. In analyzing performance of hospitals, Arocena and García-Prado (2007) also highlighted the relevance of bad outputs (quality issues) in ensuring that efficiency gains are not with reduction in service quality.

In the literature, mathematical models of production date back to at least some classical works of Samuelson (1947); Dorfman et al. (1958); Shephard (1953, 1970).

¹There is also an analogous phenomenon regarding inputs termed congestion (see Färe and Svensson, 1980, for a theoretical discussion).

Early treatments of bad outputs in production analysis appeared in works of Baumol and Oates (1975, 1988) who treated them as freely (strongly) disposable inputs. Here it is worth noting that weak and strong disposability of inputs was first modelled in activity analysis models (AAM) by Shephard (1974). Following up on Shephard (1970, 1974), Färe and Grosskopf (1983b) introduced a weak output measure of technical efficiency to evaluate loss due to lack of free disposability of bad outputs.² Its empirical illustration can be found in, e.g., an application to steam electric plants in the U.S. by Färe et al. (1986). Using a different approach, Pittman (1983) extended the multilateral productivity index introduced by Caves, Christensen and Diewert (1982) to model undesirable outputs. In contrast, following Shephard (1974), Färe et al. (1989) developed and implemented a performance measure accounting for bad outputs which differed from Pittman (1983) in two major points: (i) an enhanced hyperbolic efficiency measure was used instead of the superlative productivity index, and (ii) nonparametric AAM was employed rather than a parametric model (translog transformation function). Since Färe et al. (1989), the framework where inputs and good outputs are strongly disposable whereas bad outputs are jointly weak disposable with good outputs has been applied widely in the growing stream of nonparametric research on bad outputs.³

In a comment to the paper of Hailu and Veeman (2001), Färe and Grosskopf (2003) clarified the framework under variable returns to scale (VRS), emphasizing the importance of a parameter representing jointly weak disposability of good and bad outputs in AAM as inspired by Shephard (1974). Subsequently, Kuosmanen (2005) proposed to use multiple scaling parameters in lieu of a single one (as since Shephard (1974)), which started an ongoing debate on whether a single or multiple scaling factors should be used.

Under constant returns to scale (CRS) and nonincreasing returns to scale (NIRS),

²Note that most early works on weak and strong disposability focused on inputs, including Borts and Mishan (1962); Maxwell (1965); Sitorus (1966); McFadden (1978); Färe and Svensson (1980); Färe and Grosskopf (1983a); Färe et al. (1983). Subsequently, weak disposability of outputs appeared in Färe and Grosskopf (1983b); Färe et al. (1985). See Färe et al. (1994) for a more detailed literature review.

³The Google Scholar web search found about 17,600 results for “weak disposability” and “bad outputs” and about 15,500 results for “weak disposability” and “undesirable outputs” (as of July 23, 2017).

the two approaches are identical since both of the single and multiple scaling factors collapse to unity in the optimization problem (Färe and Grosskopf, 2009). However, it might be too restrictive to assume CRS or NIRS in production analysis (Tyteca, 1996). Indeed, under CRS and to some extent under NIRS, efficiency of an individual may be measured towards peer firms of very different sizes whereas under VRS, efficiency is measured towards peer firms of more similar sizes. In this article we shed some new light on the relationship between the two approaches under VRS from various theoretical perspectives.

In a nutshell, the contribution of this paper is twofold. Firstly, we present new theoretical findings which to some extent help reconcile the two approaches. Secondly, we point out and prove several important properties of reference technology sets which have not been unveiled before and propose a new additional reference technology set. This yields a more complete taxonomy that can be considered as an extension of the spectrum of reference technology sets in the traditional context (without bad outputs) which was constructed mainly from a series of works in *Management Science* (Banker et al., 1984; Banker and Maindiratta, 1986; Petersen, 1990; Bogetoft, 1996; Bogetoft et al., 2000). Using this taxonomy, researchers can opt for a reference technology set which matches with their particular purposes, production technologies, and datasets.

The paper is organized as follows. Section 2 formalizes the key concepts and notions in nonparametric production analysis and summarizes the ongoing debate. Section 3 presents our theoretical findings, including new theorems, numerical evidence and a taxonomy of reference technology sets. Section 4 concludes. Proofs of new theorems are presented in Appendix A.

2 Preliminaries

2.1 Foundation

We denote inputs, good outputs and bad outputs by column vectors $x \in \mathbb{R}_+^N$, $v \in \mathbb{R}_+^M$, and $w \in \mathbb{R}_+^J$, respectively.⁴ The true (or hypothetical) production technology is represented by

$$Y = \{(v, w, x) \in \mathbb{R}_+^M \times \mathbb{R}_+^J \times \mathbb{R}_+^N : x \text{ can produce } (v, w)\}. \quad (1)$$

Equivalently, Y can be characterized via the output correspondence $P : \mathbb{R}_+^N \rightarrow 2^{\mathbb{R}_+^M \times \mathbb{R}_+^J}$ which defines the output sets

$$P(x) = \{(v, w) \in \mathbb{R}_+^M \times \mathbb{R}_+^J : (v, w, x) \in Y\}, \quad x \in \mathbb{R}_+^N, \quad (2)$$

or via the input correspondence $L : \mathbb{R}_+^M \times \mathbb{R}_+^J \rightarrow 2^{\mathbb{R}_+^N}$ which defines the input requirement sets

$$L(v, w) = \{x \in \mathbb{R}_+^N : (v, w, x) \in Y\}, \quad (v, w) \in \mathbb{R}_+^M \times \mathbb{R}_+^J. \quad (3)$$

Next, we adopt the axiomatic approach to production theory of Shephard (1953, 1970) and Färe and Primont (1995) by assuming that Y satisfies several standard regularity axioms as follows.

A1 No free lunch: $(v, w, 0) \notin Y \quad \forall (v, w) \geq 0$.⁵

A2 Producing nothing is possible: $\exists x_o \in \mathbb{R}_+^N : (0, 0, x) \in Y \quad \forall x \geq x_o$.⁶

A3 $P(x)$ is bounded for all $x \in \mathbb{R}_+^N$.

⁴The concepts and notations we use are inspired by Shephard (1953, 1970); Färe (1988); Kuosmanen (2005); Kuosmanen and Podinovski (2009); Podinovski and Kuosmanen (2011).

⁵For $a, b \in \mathbb{R}^m$, “ $a \geq b$ ” or “ $b \leq a$ ” means all elements of a is greater than or equal to the corresponding elements of b (i.e., $a - b \in \mathbb{R}_+^m$), “ $a \geq b$ ” or “ $b \leq a$ ” means “ $a \geq b$ ” and at least one element of a is greater than the corresponding element of b (i.e., $a - b \in \mathbb{R}_+^m \setminus \{0_m\}$); “ $a > b$ ” or “ $b < a$ ” means all elements of a are greater than the corresponding elements of b (i.e., $a - b \in \mathbb{R}_{++}^m$).

⁶ x_o represents a minimum cost which might be needed to establish the production, e.g., fixed cost.

A4 Y is a closed set.

A5 Strong disposability of inputs: $(v, w, x) \in Y \Rightarrow (v, w, x^*) \in Y \forall x^* \geq x$.

A6 Strong disposability of good outputs: $(v, w, x) \in Y \Rightarrow (v^*, w, x) \in Y \forall v^* \leq v$.

A7 Jointly weak disposability of good outputs and bad outputs: $(v, w, x) \in Y \Rightarrow (\theta v, \theta w, x) \in Y \forall \theta \in [0, 1]$.⁷

Axiom **A7** reflects that bad outputs cannot be disposed freely. Instead, they must be reduced in proportion to the reduction of good outputs (see, e.g., Shephard (1970, 1974); Färe et al. (1989); Chung et al. (1997); Färe et al. (2005), for more discussions).⁸

Last but not least, in the lemma below we recall duality between quasiconcavity of the input (output) correspondence and the output (input requirement) sets, which was mentioned in Shephard (1953, 1970); Färe (1988).⁹

Lemma 1. *Given a technology set Y , we have*

- (a) *The output correspondence P characterizing Y is quasiconcave if and only if the input requirement set $L(v, w)$ is convex for all $(v, w) \in \mathbb{R}_+^M \times \mathbb{R}_+^J$,*
- (b) *The input correspondence L characterizing Y is quasiconcave if and only if the output set $P(x)$ is convex for all $x \in \mathbb{R}_+^N$.*

Proof. See Appendix B. □

2.2 Implementing jointly weak disposability of good and bad outputs in AAM

In practice, Y is unknown and must be estimated from a sample of observations:

$\mathcal{S}_K = \{(v^k, w^k, x^k) : k = 1, \dots, K\}$ assuming Axioms **A1-A7**.¹⁰ Although Axiom **A7**

⁷There are several terms used to refer to the parameter θ , e.g., “abatment factor” (Kuosmanen, 2005), “disposability parameter” or “scaling factor” (Färe and Grosskopf, 2009). In this paper we call θ “scaling factor”.

⁸Several studies also impose null-jointness of good and bad outputs: $(v, 0, x) \in Y \Rightarrow v = 0$.

⁹A corrsponce $R : \mathbb{R}^n \rightarrow 2^{\mathbb{R}^m}$ is quasiconcave on \mathbb{R}^n if $R((1 - \lambda)a + \lambda b) \supseteq R(a) \cap R(b) \forall a, b \in \mathbb{R}^n, \lambda \in [0, 1]$ (Shephard, 1953, 1970; Färe, 1988).

¹⁰It is worth noting that convexity is not invoked as a prior axiom here.

is conceded widely in studies concerning undesirable outputs, how to implement it in nonparametric AAM is still debatable as mentioned in Section 1. Färe and Grosskopf (2003), in the spirit of Shephard (1974), clarified an estimation of Y under VRS as follows:

$$\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K) = \left\{ (v, w, x) : v \leq \theta \sum_{k=1}^K z^k v^k; w = \theta \sum_{k=1}^K z^k w^k; x \geq \sum_{k=1}^K z^k x^k; \sum_{k=1}^K z^k = 1; z^k \geq 0, k = 1, \dots, K; 0 \leq \theta \leq 1 \right\}. \quad (4)$$

Subsequently, Kuosmanen (2005) proposed another estimator of Y under VRS:

$$\hat{Y}_{\mathcal{X}}(\mathcal{S}_K) = \left\{ (v, w, x) : v \leq \sum_{k=1}^K \theta^k z^k v^k; w = \sum_{k=1}^K \theta^k z^k w^k; x \geq \sum_{k=1}^K z^k x^k; \sum_{k=1}^K z^k = 1; z^k \geq 0, 0 \leq \theta^k \leq 1, k = 1, \dots, K \right\}. \quad (5)$$

It is worth noting that the only difference between the two approaches lies at the parameter representing Axiom **A7**: a single scalar θ is used in $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K)$ whereas in $\hat{Y}_{\mathcal{X}}(\mathcal{S}_K)$ each observation is given a different parameter $(\theta^1, \dots, \theta^K)$. For this reason, henceforward we refer to the former and latter methods as the single and multiple scaling factor approaches, respectively. Kuosmanen and Podinovski (2009) were the first to point out that $\hat{Y}_{\mathcal{X}}(\mathcal{S}_K)$ is the smallest convex reference technology set derived from the dataset \mathcal{S}_K . Meanwhile, Färe and Grosskopf (2009) claimed that $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K)$ satisfies convexity of output sets.

Recently, Podinovski and Kuosmanen (2011) pointed out that $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K)$ is not the smallest hull of the data satisfying convexity of output sets and proposed two new estimators of Y on the basis of relaxation of convexity. Firstly, Podinovski and Kuosmanen (2011) imposed convexity on only the output sets and proposed a minimal reference technology set satisfying this condition. In light of Lemma 1, its respective input correspondence is quasiconcave. Therefore, we will refer to this estimator as quasiconcave-input-correspondence estimator of the technology set with free dispos-

ability of inputs and good outputs and jointly weak disposability of good and bad outputs, denoted by $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}(\mathcal{S}_K)$. The mathematical expression of $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}(\mathcal{S}_K)$ is:

$$\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}(\mathcal{S}_K) = \left\{ (v, w, x) : v \leq \sum_{k=1}^K \theta^k z^k v^k; w = \sum_{k=1}^K \theta^k z^k w^k; \sum_{k=1}^K z^k = 1; \right. \\ \left. z^k \geq 0, 0 \leq \theta^k \leq 1, \text{ if } z^k > 0 \text{ then } x \geq x^k, k = 1, \dots, K \right\}. \quad (6)$$

The second estimator proposed by Podinovski and Kuosmanen (2011) assumes no convexity at all and is defined as:

$$\hat{Y}_{\mathcal{P}\mathcal{H}}^{QFDH}(\mathcal{S}_K) = \left\{ (v, w, x) : v \leq \sum_{k=1}^K \theta^k z^k v^k; w = \sum_{k=1}^K \theta^k z^k w^k; x \geq \sum_{k=1}^K z^k x^k; \right. \\ \left. \sum_{k=1}^K z^k = 1; z^k \in \{0, 1\}, 0 \leq \theta^k \leq 1, k = 1, \dots, K \right\}. \quad (7)$$

This estimate resembles the Free Disposal Hull (FDH) introduced by Deprins et al. (1984), except that an additional equation related to bad outputs is added to the set of constraints and the scaling parameters θ^k ($k = 1, \dots, K$) appear in constraints on good and bad outputs. Hence, in this paper we will refer to this estimator as quasi-free-disposal-hull estimator of the technology set with free disposability of inputs and good outputs and jointly weak disposability of good and bad outputs, denoted by $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QFDH}(\mathcal{S}_K)$. We provide a graphical depiction of the four estimators $\hat{Y}_{\mathcal{H}}$, $\hat{Y}_{\mathcal{F}\mathcal{G}}$, $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}$, and $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QFDH}$ in Appendix C.

3 Theoretical discussion

3.1 Interaction between disposability and returns to scale

Basically, in order to construct a reference technology set for AAM, one needs to use two important characteristics: (i) disposability, which is represented by the scaling parameters, inequality and equality signs in constraints regarding inputs, good and bad outputs, and (ii) returns to scale, which is represented by constraints on intensity

variables z and their sum. In principle, the reference technology set estimated by the single scaling factor approach is always a subset of that estimated by the multiple scaling factor approach. However, returns to scale can partially compensate for disposability in the sense that it adds to reference technology sets activities which are not included when only disposability is assumed. Particularly, under CRS and NIRS, the scaling parameters are unnecessary in AAM (i.e., they can be set to 1) (Färe and Grosskopf, 2009), hence the single and multiple scaling factor approaches are identical.

Next we investigate the VRS case by a numerical example. Consider a dataset consisting of two firms B_1 and B_2 using the same amount of inputs to produce one good output and one bad output (v, w) : $B_1 = (2, 1)$ and $B_2 = (1, 2)$ (Figure 1).¹¹ The reference technology set extrapolated by assuming VRS, strong disposability of v and jointly weak disposability of v and w are the shaded area in Figure 1c. On the one hand, the VRS assumption helps include points lying in the triangle $B_1B_2B_3$, which are omitted when only strong disposability of v and jointly weak disposability of v and w are assumed (Figure 1a). On the other hand, compared to Figure 1d, points lying in the triangle OB_1B_4 are excluded from the reference technology set due to jointly weak disposability of v and w . In addition, when B_1 moves along the segment B_1B_4 toward B_4 , the reference technology set inflates and approaches the set associated with the case where v is strongly disposable (Figure 1d).

The above example well illustrates the ability of returns to scale in adding to reference technology sets (without regard to whether the single or multiple scaling factor approach is used) activities which are not included when only the disposability characteristic is employed.

3.2 When do the two paradigms coincide?

As mentioned in Section 3.1, the single and multiple scaling factor approaches are known to be identical in three cases: (i) CRS, (ii) NIRS, and (iii) VRS if $x^1 = \dots = x^K$. Under VRS, we discover more cases which imply several important implications. Prior

¹¹Note that the reference technology sets estimated by two approaches are alike if all decision-making-units (DMUs) use the same amount of inputs (Kuosmanen and Podinovski, 2009).

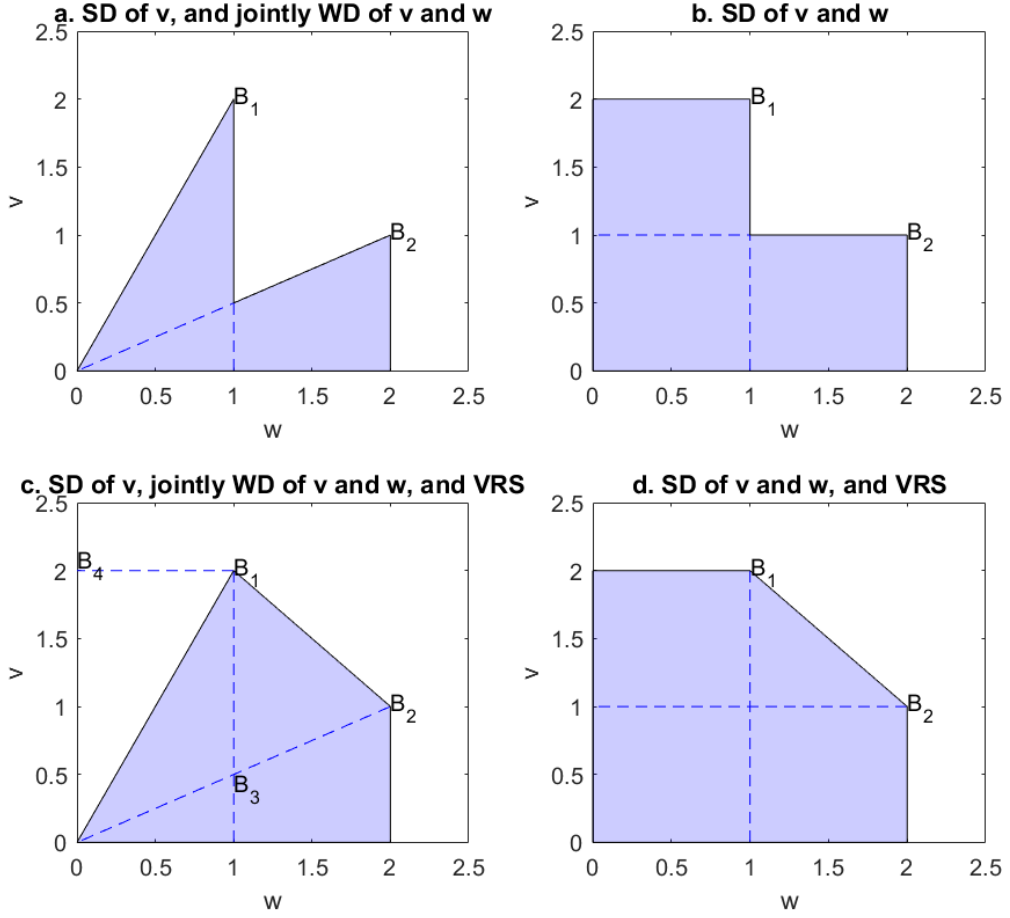


Figure 1: Illustration of interaction between disposability and returns to scale in construction of output sets. SD stands for Strong Disposability, WD stands for Weak Disposability.

to presenting these new results, we highlight two interesting and obvious properties below.

P1 $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K) \subseteq \hat{Y}_{\mathcal{H}}(\mathcal{S}_K)$ for any \mathcal{S}_K .

P2 $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_{K_1}) \subseteq \hat{Y}_{\mathcal{F}g}(\mathcal{S}_{K_2}), \hat{Y}_{\mathcal{H}}(\mathcal{S}_{K_1}) \subseteq \hat{Y}_{\mathcal{H}}(\mathcal{S}_{K_2})$ for any $\mathcal{S}_{K_1} \subseteq \mathcal{S}_{K_2}$.

The following theorem shows a special case where two estimators $\hat{Y}_{\mathcal{F}g}$ and $\hat{Y}_{\mathcal{H}}$ yield identical reference technology sets.

Theorem 1. For any \mathcal{S}_K , we have

(a) $\hat{Y}_{\mathcal{H}}(\mathcal{S}_K \cup \{(\tilde{v}, \tilde{w}, \tilde{x})\}) = \hat{Y}_{\mathcal{H}}(\mathcal{S}_K) \forall (\tilde{v}, \tilde{w}, \tilde{x}) \in \hat{Y}_{\mathcal{H}}(\mathcal{S}_K)$,

(b) $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K^*) = \hat{Y}_{\mathcal{H}}(\mathcal{S}_K^*) = \hat{Y}_{\mathcal{H}}(\mathcal{S}_K)$ where $\mathcal{S}_K^* = \mathcal{S}_K \cup \{(0_M, 0_J, x^k)\}_{k=1}^K$.

Proof. See Appendix A.1. □

Theorem 1a hints that the fact that an extrapolated activity becomes an observed one does not change the reference technology set estimated by $\hat{Y}_{\mathcal{X}}$. Meanwhile, this does not always hold true for $\hat{Y}_{\mathcal{F}g}$. Interestingly, when inactivity DMUs $(0_M, 0_J, x^k)$ ($k = 1, \dots, K$) become observed ones, $\hat{Y}_{\mathcal{F}g}$ and $\hat{Y}_{\mathcal{X}}$ produce identical reference technology sets. Note that those inactivity DMUs, which appear in $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K)$ as extrapolated DMUs, are hypothetical (i.e., not observed in reality) but theoretically feasible by Axiom **A2**. Equally important, it can be deduced from Theorem 1 that multiple scaling factors are unnecessary if the dataset is of the form similar to \mathcal{S}_K^* , i.e., for any input vector observed, there exists in the sample an inactivity DMU corresponding to that input vector.

Furthermore, we discover that both of the single and multiple scaling factor approaches are definitely identical when convexity of Y is relaxed partly or entirely. First, note that estimators $\hat{Y}_{\mathcal{P}\mathcal{X}}^{QC}$ and $\hat{Y}_{\mathcal{P}\mathcal{X}}^{QFDH}$, proposed by Podinovski and Kuosmanen (2011) as mentioned in Section 2.2, follow the multiple scaling factor approach. Naturally, we consider their analogues following the single scaling factor approach that is to replace different scaling factors $\theta^1, \dots, \theta^K$ by only a single θ . We denote these analogues by $\hat{Y}_{\mathcal{F}g}^{QC}$ and $\hat{Y}_{\mathcal{F}g}^{QFDH}$, respectively. The mathematical expression of $\hat{Y}_{\mathcal{F}g}^{QC}(\mathcal{S}_K)$ is:

$$\hat{Y}_{\mathcal{F}g}^{QC}(\mathcal{S}_K) = \left\{ (v, w, x) : v \leq \theta \sum_{k=1}^K z^k v^k; w = \theta \sum_{k=1}^K z^k w^k; \sum_{k=1}^K z^k = 1; \right. \\ \left. 0 \leq \theta \leq 1; z^k \geq 0, \text{ if } z^k > 0 \text{ then } x \geq x^k, k = 1, \dots, K \right\}. \quad (8)$$

Clearly, $\hat{Y}_{\mathcal{F}g}^{QC}(\mathcal{S}_K)$ is a subset of $\hat{Y}_{\mathcal{P}\mathcal{X}}^{QC}(\mathcal{S}_K)$ since setting $\theta^1 = \dots = \theta^K = \theta$ in $\hat{Y}_{\mathcal{P}\mathcal{X}}^{QC}(\mathcal{S}_K)$ gives $\hat{Y}_{\mathcal{F}g}^{QC}(\mathcal{S}_K)$. It turns out that $\hat{Y}_{\mathcal{P}\mathcal{X}}^{QC}(\mathcal{S}_K)$ and $\hat{Y}_{\mathcal{F}g}^{QC}(\mathcal{S}_K)$ are, in fact, always identical as summarized in the following theorem.

Theorem 2. *For any \mathcal{S}_K , we have*

$$\hat{Y}_{\mathcal{F}g}^{QC}(\mathcal{S}_K) = \hat{Y}_{\mathcal{P}\mathcal{X}}^{QC}(\mathcal{S}_K). \quad (9)$$

Proof. See Appendix A.2. □

The implication of Theorem 2 is twofold. Firstly, it contributes to reconciliation of the single and multiple scaling factor approaches as the two turn out to be identical when only convexity of output sets is assumed. Secondly, it helps reduce the number of optimization variables in related AAM optimization problems since only one variable θ is used in $\hat{Y}_{\mathcal{F}\mathcal{G}}^{QC}(\mathcal{S}_K)$ instead of K variables $\theta^1, \dots, \theta^K$ in $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}(\mathcal{S}_K)$. Note that this reduction can be very substantial when the sample size is large (e.g., 999 variables can be eliminated for a sample of 1000 observations). Hence, the computational speed can be improved significantly, especially in situations where optimization problems are nonlinear due to several efficiency measures (e.g., the hyperbolic efficiency measure by Färe et al. (1989)). To illustrate, for each $K \in \{100, 200, 500, 1000\}$, we generated 1000 datasets \mathcal{S}_K where $N, M, J = 2$, $v_i^k, w_i^k, x_i^k \sim \text{Uniform}(1, 10)$ for all $k = 1, \dots, K$, $i = 1, 2$. For each dataset \mathcal{S}_K , we used $\hat{Y}_{\mathcal{F}\mathcal{G}}^{QC}$ and $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}$ to estimate the efficiency score of the first observation in \mathcal{S}_K using the directional distance function measure (DDF) (Diewert, 1983; Chambers et al., 1996, 1998):

$$DDF(v, w, x) = \max_{\alpha} \{ \alpha : (v + \alpha g_v, w - \alpha g_w, x - \alpha g_x) \in Y \}, \quad (10)$$

where g_v, g_w, g_x are directional vectors with all elements equal to 1.

The results show that the estimation following the single scaling factor approach ($\hat{Y}_{\mathcal{F}\mathcal{G}}^{QC}$) is, on average, faster than that following the multiple scaling factor approach ($\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}$). Moreover, the timing gap inflates as the sample size increases, e.g., when $K = 1000$, the average runtime of the program based on $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}$ is about 1.66 times that based on $\hat{Y}_{\mathcal{F}\mathcal{G}}^{QC}$ (Figure 2).¹² Such a difference is particularly substantial if one needs to use bootstrap or other simulation-based methods requiring many replications.

Next, we consider the mathematical expression of $\hat{Y}_{\mathcal{F}\mathcal{G}}^{QFDH}$ (i.e., the analogue of

¹²Configuration of the machine used for simulation: Intel Core i5-5200 CPU 2.20Ghz (4CPUs), 8192MB RAM. Optimization is done using the mix-integer linear programming solver “intlinprog” of MATLAB® (version R2017a 64-bit) with the same options (e.g., tolerance, maximum iterations) for both approaches.

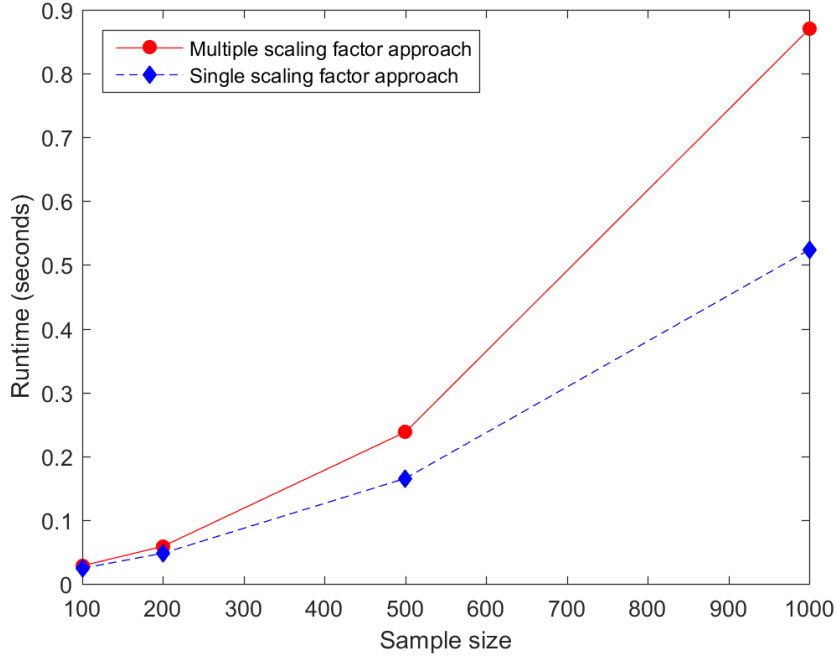


Figure 2: Average time needed to compute DDF efficiency score of one observation

$\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}$) below.

$$\hat{Y}_{\mathcal{F}g}^{QFDH}(\mathcal{S}_K) = \left\{ (v, w, x) : v \leq \theta \sum_{k=1}^K z^k v^k; w = \theta \sum_{k=1}^K z^k w^k; x \geq \sum_{k=1}^K z^k x^k; \sum_{k=1}^K z^k = 1; 0 \leq \theta \leq 1; z^k \in \{0, 1\}, k = 1, \dots, K \right\}. \quad (11)$$

Again, $\hat{Y}_{\mathcal{F}g}^{QFDH}(\mathcal{S}_K)$ is a subset of $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}(\mathcal{S}_K)$ since $\hat{Y}_{\mathcal{F}g}^{QC}(\mathcal{S}_K)$ can be obtained from $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}(\mathcal{S}_K)$ by setting $\theta^1 = \dots = \theta^K = \theta$. Interestingly, these two estimators are also identical as described below.

Theorem 3. For any \mathcal{S}_K , we have

$$\hat{Y}_{\mathcal{F}g}^{QFDH}(\mathcal{S}_K) = \hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}(\mathcal{S}_K). \quad (12)$$

Proof. See Appendix A.3. □

Theorem 3 also contributes to reconciling the single and multiple scaling factor approaches by confirming that they are identical when no convexity is assumed. Unlike the case of Theorem 2, it is unnecessary to compare computational speed between opti-

mization problems based on $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}$ and $\hat{Y}_{\mathcal{F}\mathcal{G}}^{QFDH}$ since their computational procedures are identical and similar to that of the traditional FDH.¹³

3.3 A taxonomy of reference technology sets

Until Podinovski and Kuosmanen (2011), the works contributed to the debate regarding the single and multiple scaling factor approaches have led to four estimators of the true technology: $\hat{Y}_{\mathcal{K}}$, $\hat{Y}_{\mathcal{F}\mathcal{G}}$, $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QC}$, and $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}$. These estimators can be classified according to parts of the true technology imposed with convexity: the whole technology ($\hat{Y}_{\mathcal{K}}$), only the output sets ($\hat{Y}_{\mathcal{F}\mathcal{G}}$ and $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QC}$), and no convexity at all ($\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}$).

Interestingly, about two decades ago the literature had somewhat a similar debate with respect to the conventional AAM without bad outputs, which forms up a spectrum of reference technology sets by relaxing the assumption of convexity: (i) the true technology is convex (Banker et al., 1984), (ii) the true technology is not necessarily convex, but both of the output sets and input sets are convex (Bogetoft, 1996; Bogetoft et al., 2000), (iii) only the output sets are convex (Petersen, 1990), (iv) only the input sets are convex (Petersen, 1990), and (v) no convexity at all (Deprins et al., 1984). We summarize these cases in Table 1.

Table 1: A spectrum of reference technology sets where all outputs are desirable

	Parts of technology imposed with convexity	Reference technology sets
(i)	The whole technology	Banker et al. (1984)
(ii)	Both output and input requirement sets	Bogetoft (1996), Bogetoft et al. (2000)
(iii)	Only output sets	Petersen (1990)
(iv)	Only input requirement sets	Petersen (1990)
(v)	No convexity at all	Deprins et al. (1984)

The debate between the single and multiple scaling factor approaches naturally complements the above debate with undesirable outputs taken into consideration. On the basis of this recognition, we summarize a taxonomy of reference technology sets for the AAM with undesirable outputs under VRS, which comprises five categories

¹³Podinovski and Kuosmanen (2011) also mentioned the procedure for computing $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}$ in practice. By definition, only one of z^k 's is equal to 1 while the others are zero. Thus, optimization using $\hat{Y}_{\mathcal{F}\mathcal{G}}^{QFDH}$ and $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}$ can be done by optimizing K sub-problems corresponding to $z^k = 1$ and $z^l = 0$ ($l \neq k$) where k varies from 1 to K , which are the same for both $\hat{Y}_{\mathcal{F}\mathcal{G}}^{QFDH}$ and $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}$, and then take the optimal of the K sub-problems.

similar to Table 1. In order to achieve this goal, we point out some properties of the current reference technology sets which have not been unveiled before and also propose a new reference technology set to fill in the taxonomy. The first property to reveal is described in the following theorem.

Theorem 4.

- (a) For any \mathcal{S}_K , the input requirement sets of $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K)$ are convex.
- (b) The input requirement sets of $\hat{Y}_{\mathcal{P}\mathcal{X}}^{QC}(\mathcal{S}_K)$ and therefore of $\hat{Y}_{\mathcal{F}g}^{QC}(\mathcal{S}_K)$ are not necessarily convex for any \mathcal{S}_K .

Proof. See Appendix A.4. □

On the one hand, Theorem 4a unveils that $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K)$, which satisfies convexity of output sets (Färe and Grosskopf, 2009), also satisfies convexity of input requirement sets. As it does not always satisfies convexity of the entire technology (Podinovski and Kuosmanen, 2011), we can place it in the category (ii) – both input sets and output sets are convex, but not necessarily the entire technology. On the other hand, Theorem 4b suggests that $\hat{Y}_{\mathcal{P}\mathcal{X}}^{QC}(\mathcal{S}_K)$ (or equivalently, $\hat{Y}_{\mathcal{F}g}^{QC}(\mathcal{S}_K)$ by Theorem 2), which was proved to be the smallest set satisfying convexity of output sets (Podinovski and Kuosmanen, 2011), does not always satisfy convexity of input requirement sets and hence, it should be placed in category (iii) – only the output sets are convex. Obviously categories (i) and (v) are filled by $\hat{Y}_{\mathcal{X}}(\mathcal{S}_K)$ and $\hat{Y}_{\mathcal{P}\mathcal{X}}^{QFDH}(\mathcal{S}_K)$ respectively, according to Kuosmanen and Podinovski (2009); Podinovski and Kuosmanen (2011).

The taxonomy is still incomplete since the category (iv) is currently empty. We fill this blank by proposing a new reference technology set defined as:

$$\hat{Y}_{\mathcal{J}}(\mathcal{S}_K) = \left\{ (v, w, x) : x \geq \sum_{k=1}^K z^k x^k; \sum_{k=1}^K z^k = 1; z^k \geq 0, 0 \leq \theta^k \leq 1, k = 1, \dots, K; \right. \\ \left. \text{if } z^k > 0 \text{ then } v \leq \theta^k v^k \text{ and } w = \theta^k w^k, k = 1, \dots, K \right\}. \quad (13)$$

An interesting property of $\hat{Y}_{\mathcal{J}}$ is presented below.

Theorem 5. For any \mathcal{S}_K , $\hat{Y}_{\mathcal{J}}(\mathcal{S}_K)$ is the minimal reference technology set satisfying convexity of the input requirement sets.

Proof. See Appendix A.5. □

By Theorem 5, we can position $\hat{Y}_{\mathcal{J}}(\mathcal{S}_K)$ in category (iv), completing the taxonomy. Additionally, its output correspondence is quasiconcave by Lemma 1. All in all, the taxonomy of reference technology sets can be summarized as in Table 2.

Table 2: A taxonomy of reference technology sets under VRS

	Parts imposed with convexity	AAM without bad outputs	AAM with bad outputs
(i)	The whole technology	Banker et al. (1984)	$\hat{Y}_{\mathcal{K}}$ (Kuosmanen, 2005)
(ii)	Both output and input requirement sets	Bogetoft (1996), Bogetoft et al. (2000)	$\hat{Y}_{\mathcal{F}\mathcal{G}}$ (Färe and Grosskopf, 2003)
(iii)	Only output sets	Petersen (1990)	$\hat{Y}_{\mathcal{P}\mathcal{K}}^{QC}$ (Podinovski and Kuosmanen, 2011)
(iv)	Only input requirement sets	Petersen (1990)	$\hat{Y}_{\mathcal{J}}$ (proposed in this paper)
(v)	No convexity at all	Deprins et al. (1984)	$\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}$ (Podinovski and Kuosmanen, 2011)

As proved in this paper, $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QC} \equiv \hat{Y}_{\mathcal{F}\mathcal{G}}^{QC}$, $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH} \equiv \hat{Y}_{\mathcal{F}\mathcal{G}}^{QFDH}$.

4 Conclusion

In this paper we have made two main contributions. Firstly, we shed new lights to the debate between the single and multiple scaling factor approaches by showing that: (i) returns to scale can partially compensate for jointly weak disposability of good and bad outputs, (ii) the single and multiple scaling factor approaches are identical in several scenarios, especially when convexity is relaxed (partly or entirely), and (iii) the single scaling factor approach appears to be more convenient than the multiple one in some aspects (e.g., the former involves fewer optimization variables and faster computational speed). Secondly, by linking two interesting debates in the literature, we help construct a comprehensive taxonomy of reference technology sets for AAM under VRS.

Furthermore, this paper suggests several directions for future research. Firstly, although the reference technology set presented in Färe and Grosskopf (2003) has been proved to satisfy convexity of both output and input requirement sets, whether it is

the minimal set satisfying this property has not been verified yet, raising a question for future works. Another direction would be to develop statistical properties for the reference technology sets in the presence of bad outputs, similar to those developed for the traditional AAM without bad outputs. For example, Korostelev et al. (1995) proved an asymptotic property of the FDH estimator, Kneip et al. (1998) proved the consistency of the Data Envelopment Analysis (DEA) estimators, Park et al. (2000) established the limiting distribution of the FDH estimator for the fully multivariate case. Hence, similar results should be established for the new estimators in the presence of bad outputs.

Appendix A Proofs of Theorems

A.1 Proof of Theorems 1

(a) Since $(\tilde{v}, \tilde{w}, \tilde{x}) \in \hat{Y}_{\mathcal{X}}(\mathcal{S}_K)$, there exist numbers $\tilde{z}^k \geq 0$ and $\tilde{\theta}^k \in [0, 1]$ ($k = 1, \dots, K$) such that:

$$\tilde{v} \leq \sum_{k=1}^K \tilde{\theta}^k \tilde{z}^k v^k, \tilde{w} = \sum_{k=1}^K \tilde{\theta}^k \tilde{z}^k w^k, \tilde{x} \geq \sum_{k=1}^K \tilde{z}^k x^k, \sum_{k=1}^K \tilde{z}^k = 1. \quad (14)$$

Therefore,

$$\begin{aligned} & \hat{Y}_{\mathcal{X}}(\mathcal{S}_K \cup \{(\tilde{v}, \tilde{w}, \tilde{x})\}) \\ &= \left\{ (v, w, x) : v \leq \sum_{k=1}^K \theta^k z^k v^k + \tilde{\theta} \tilde{z} \tilde{v}; w = \sum_{k=1}^K \theta^k z^k w^k + \tilde{\theta} \tilde{z} \tilde{w}; x \geq \sum_{k=1}^K z^k x^k + \tilde{z} \tilde{x}; \right. \\ & \quad \left. \sum_{k=1}^K z^k + \tilde{z} = 1; \tilde{z} \geq 0; 0 \leq \tilde{\theta} \leq 1; z^k \geq 0, 0 \leq \theta^k \leq 1, k = 1, \dots, K \right\} \quad (15) \end{aligned}$$

$$\begin{aligned} & \subseteq \left\{ (v, w, x) : v \leq \sum_{k=1}^K \theta^k z^k v^k + \tilde{\theta} \tilde{z} \sum_{k=1}^K \tilde{\theta}^k \tilde{z}^k v^k; w = \sum_{k=1}^K \theta^k z^k w^k + \tilde{\theta} \tilde{z} \sum_{k=1}^K \tilde{\theta}^k \tilde{z}^k w^k; \right. \\ & \quad \left. x \geq \sum_{k=1}^K z^k x^k + \tilde{z} \sum_{k=1}^K \tilde{z}^k x^k; \sum_{k=1}^K z^k + \tilde{z} = 1; \tilde{z} \geq 0; 0 \leq \tilde{\theta} \leq 1; z^k \geq 0, 0 \leq \theta^k \leq 1 \forall k \right\} \quad (16) \end{aligned}$$

$$\begin{aligned} &= \left\{ (v, w, x) : v \leq \sum_{k=1}^K (\theta^k z^k + \tilde{\theta} \tilde{z} \tilde{\theta}^k \tilde{z}^k) v^k; w = \sum_{k=1}^K (\theta^k z^k + \tilde{\theta} \tilde{z} \tilde{\theta}^k \tilde{z}^k) w^k; \sum_{k=1}^K z^k + \tilde{z} = 1; \right. \\ & \quad \left. x \geq \sum_{k=1}^K (z^k + \tilde{z} \tilde{z}^k) x^k; \tilde{z} \geq 0; 0 \leq \tilde{\theta} \leq 1; z^k \geq 0, 0 \leq \theta^k \leq 1, k = 1, \dots, K \right\} \quad (17) \end{aligned}$$

$$\begin{aligned} & \subseteq \left\{ (v, w, x) : v \leq \sum_{k=1}^K \hat{\theta}^k \hat{z}^k v^k; w = \sum_{k=1}^K \hat{\theta}^k \hat{z}^k w^k; x \geq \sum_{k=1}^K \hat{z}^k x^k; \right. \\ & \quad \left. \sum_{k=1}^K \hat{z}^k = 1; \hat{z}^k \geq 0, 0 \leq \hat{\theta}^k \leq 1, k = 1, \dots, K \right\} \quad (18) \end{aligned}$$

$$= \hat{Y}_{\mathcal{X}}(\mathcal{S}_K) \quad (19)$$

Here (16) follows by using (14), while (18) follows by substituting $\theta^k z^k + \tilde{\theta} \tilde{z} \tilde{\theta}^k \tilde{z}^k$ and $z^k + \tilde{z} \tilde{z}^k$ in (17) by $\hat{\theta}^k \hat{z}^k$ and \hat{z}^k , respectively. These substitutions are feasible because:

(i) $\sum_{k=1}^K \hat{z}^k = \sum_{k=1}^K z^k + \tilde{z} \sum_{k=1}^K \tilde{z}^k = \sum_{k=1}^K z^k + \tilde{z} = 1$, and (ii) $\theta^k z^k + \tilde{\theta} \tilde{\theta}^k \tilde{z}^k \leq z^k + \tilde{z} \tilde{z}^k = \hat{z}^k$, implying that $\exists \hat{\theta}^k \in [0, 1] : \theta^k z^k + \tilde{\theta} \tilde{\theta}^k \tilde{z}^k = \hat{\theta}^k \hat{z}^k$ ($k = 1, \dots, K$).

Thus, $\hat{Y}_{\mathcal{X}}(\mathcal{S}_K \cup \{(\tilde{v}, \tilde{w}, \tilde{x})\}) \subseteq \hat{Y}_{\mathcal{X}}(\mathcal{S}_K)$. Meanwhile, $\hat{Y}_{\mathcal{X}}(\mathcal{S}_K) \subseteq \hat{Y}_{\mathcal{X}}(\mathcal{S}_K \cup \{(\tilde{v}, \tilde{w}, \tilde{x})\})$ by Property **P2**. Therefore, we have $\hat{Y}_{\mathcal{X}}(\mathcal{S}_K) = \hat{Y}_{\mathcal{X}}(\mathcal{S}_K \cup \{(\tilde{v}, \tilde{w}, \tilde{x})\})$. \square

(b) Firstly, $\hat{Y}_{\mathcal{X}}(\mathcal{S}_K^*) = \hat{Y}_{\mathcal{X}}(\mathcal{S}_K)$ follows directly from part (a) since $(0_M, 0_J, x^k) \in \hat{Y}_{\mathcal{X}}(\mathcal{S}_K) \forall k = 1, \dots, K$. Hence it remains to prove that $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K^*) = \hat{Y}_{\mathcal{X}}(\mathcal{S}_K^*)$. We start by transforming $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K^*)$ as follows:

$$\begin{aligned} & \hat{Y}_{\mathcal{F}g}(\mathcal{S}_K^*) \\ &= \left\{ (v, w, x) : v \leq \theta \sum_{k=1}^K z^k v^k + \theta \sum_{k=1}^K \tilde{z}^k 0; w = \theta \sum_{k=1}^K z^k w^k + \theta \sum_{k=1}^K \tilde{z}^k 0; \right. \\ & \quad \left. x \geq \sum_{k=1}^K z^k x^k + \sum_{k=1}^K \tilde{z}^k x^k; \sum_{k=1}^K (z^k + \tilde{z}^k) = 1; z^k, \tilde{z}^k \geq 0, k = 1, \dots, K; 0 \leq \theta \leq 1 \right\} \\ &= \left\{ (v, w, x) : v \leq \theta \sum_{k=1}^K z^k v^k; w = \theta \sum_{k=1}^K z^k w^k; x \geq \sum_{k=1}^K (z^k + \tilde{z}^k) x^k; \right. \\ & \quad \left. \sum_{k=1}^K (z^k + \tilde{z}^k) = 1; z^k, \tilde{z}^k \geq 0, k = 1, \dots, K; 0 \leq \theta \leq 1 \right\} \end{aligned} \quad (20)$$

$$\begin{aligned} &= \left\{ (v, w, x) : v \leq \theta \sum_{k=1}^K z^k v^k; w = \theta \sum_{k=1}^K z^k w^k; x \geq \sum_{k=1}^K \hat{z}^k x^k; \right. \\ & \quad \left. \sum_{k=1}^K \hat{z}^k = 1; \hat{z}^k \geq z^k \geq 0, k = 1, \dots, K; 0 \leq \theta \leq 1 \right\}. \end{aligned} \quad (21)$$

Here (21) follows by substituting $z^k + \tilde{z}^k$ by \hat{z}^k . Now let (v^o, w^o, x^o) be any element of $\hat{Y}_{\mathcal{X}}(\mathcal{S}_K)$. By definition of $\hat{Y}_{\mathcal{X}}(\mathcal{S}_K)$, there exist numbers $z^k \geq 0$ and $\theta^k \in [0, 1]$ ($k = 1, \dots, K$) such that $v^o \leq \sum_{k=1}^K \theta^k z^k v^k$, $w^o = \sum_{k=1}^K \theta^k z^k w^k$, $x^o \geq \sum_{k=1}^K z^k x^k$, and $\sum_{k=1}^K z^k = 1$. For $k = 1, \dots, K$, let $\theta^k z^k = l^k$, then we have $z^k \geq l^k \geq 0 \forall k$, $v^o \leq \sum_{k=1}^K l^k v^k$, $w^o = \sum_{k=1}^K l^k w^k$, $x^o \geq \sum_{k=1}^K z^k x^k$, and $\sum_{k=1}^K z^k = 1$. Combined with (21), this leads to $(v^o, w^o, x^o) \in \hat{Y}_{\mathcal{F}g}(\mathcal{S}_K^*)$, implying $\hat{Y}_{\mathcal{X}}(\mathcal{S}_K) \subseteq \hat{Y}_{\mathcal{F}g}(\mathcal{S}_K^*)$ or equivalently, $\hat{Y}_{\mathcal{X}}(\mathcal{S}_K^*) \subseteq \hat{Y}_{\mathcal{F}g}(\mathcal{S}_K^*)$.

On the other hand, $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K^*) \subseteq \hat{Y}_{\mathcal{X}}(\mathcal{S}_K^*)$ by Property **P1**. Therefore, $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_K^*) = \hat{Y}_{\mathcal{X}}(\mathcal{S}_K^*)$, completing the proof. \square

A.2 Proof of Theorem 2

Firstly we show that $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}(\mathcal{S}_K) \subseteq \hat{Y}_{\mathcal{F}\mathcal{G}}^{QC}(\mathcal{S}_K)$. Indeed,

$$\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}(\mathcal{S}_K) = \left\{ (v, w, x) : v \leq \sum_{k=1}^K \theta^k z^k v^k; w = \sum_{k=1}^K \theta^k z^k w^k; \sum_{k=1}^K z^k = 1; \right. \\ \left. 0 \leq \theta^k \leq 1, z^k \geq 0, \text{ if } z^k > 0 \text{ then } x \geq x^k, k = 1, \dots, K \right\} \quad (22)$$

$$\subseteq \left\{ (v, w, x) : v \leq \theta \sum_{k=1}^K \tilde{z}^k v^k; w = \theta \sum_{k=1}^K \tilde{z}^k w^k; \sum_{k=1}^K \tilde{z}^k = 1; \right. \\ \left. 0 \leq \theta \leq 1, \tilde{z}^k \geq 0, \text{ if } \tilde{z}^k > 0 \text{ then } x \geq x^k, k = 1, \dots, K \right\}. \quad (23)$$

Expression (23) results from substituting $\theta^k z^k$ in (22) by $\theta \tilde{z}^k$ where $\theta = \sum_{k=1}^K \theta^k z^k$ and

$$\tilde{z}^k = \begin{cases} \frac{\theta^k z^k}{\theta} & \text{if } \theta \neq 0, \\ z^k & \text{if } \theta = 0. \end{cases}$$

Note that $0 \leq \theta = \sum_{k=1}^K \theta^k z^k \leq \sum_{k=1}^K z^k = 1$, and $\theta = 0$ implies $\theta^k z^k = 0 \forall k = 1, \dots, K$. Thus it is easy to verify that the above choice of θ and \tilde{z}^k satisfies conditions: (i) $\theta \tilde{z}^k = \theta^k z^k \forall k = 1, \dots, K$, (ii) $\sum_{k=1}^K \tilde{z}^k = 1$, and (iii) $z^k > 0$ if $\tilde{z}^k > 0 \forall k = 1, \dots, K$. Hence, the set expressed in (22) is a subset of that expressed in (23), which is equivalent to the mathematical expression of $\hat{Y}_{\mathcal{F}\mathcal{G}}^{QC}(\mathcal{S}_K)$. Consequently, $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}(\mathcal{S}_K) \subseteq \hat{Y}_{\mathcal{F}\mathcal{G}}^{QC}(\mathcal{S}_K)$.

On the other hand, $\hat{Y}_{\mathcal{F}\mathcal{G}}^{QC}(\mathcal{S}_K) \subseteq \hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}(\mathcal{S}_K)$ as $\hat{Y}_{\mathcal{F}\mathcal{G}}^{QC}(\mathcal{S}_K)$ is a special case of $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}(\mathcal{S}_K)$ when $\theta^1 = \dots = \theta^K = \theta$. Thus, we have $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QC}(\mathcal{S}_K) = \hat{Y}_{\mathcal{F}\mathcal{G}}^{QC}(\mathcal{S}_K)$. \square

A.3 Proof of Theorem 3

To prove the Theorem, we transform $\hat{Y}_{\mathcal{P}\mathcal{H}}^{QFDH}(\mathcal{S}_K)$ as follows:

$$\hat{Y}_{\mathcal{P}\mathcal{H}}^{QFDH}(\mathcal{S}_K) = \left\{ (v, w, x) : v \leq \sum_{k=1}^K \theta^k z^k v^k; w = \sum_{k=1}^K \theta^k z^k w^k; x \geq \sum_{k=1}^K z^k x^k; \right. \\ \left. \sum_{k=1}^K z^k = 1; z^k \in \{0, 1\}, 0 \leq \theta^k \leq 1, k = 1, \dots, K \right\} \quad (24)$$

$$= \left\{ (v, w, x) : \exists l \in \{1, \dots, K\} \text{ such that } v \leq \theta^l v^l, w = \theta^l w^l, \right. \\ \left. x \geq x^l, 0 \leq \theta^l \leq 1 \right\} \quad (25)$$

$$= \left\{ (v, w, x) : \exists l \in \{1, \dots, K\} \text{ such that } v \leq \theta v^l, w = \theta w^l, \right. \\ \left. x \geq x^l; 0 \leq \theta \leq 1 \right\} \quad (26)$$

$$= \left\{ (v, w, x) : v \leq \theta \sum_{k=1}^K z^k v^k; w = \theta \sum_{k=1}^K z^k w^k; x \geq \sum_{k=1}^K z^k x^k; \right. \\ \left. \sum_{k=1}^K z^k = 1; 0 \leq \theta \leq 1; z^k \in \{0, 1\}, k = 1, \dots, K \right\} \quad (27)$$

$$= \hat{Y}_{\mathcal{F}\mathcal{G}}^{QFDH}(\mathcal{S}_K). \quad (28)$$

Here (25) and (27) follow by the fact that $\sum_{k=1}^K z^k = 1$ and $z^k \in \{0, 1\} \forall k$ are equivalent to $\exists l \in \{1, \dots, K\}$ such that $z^l = 1$ and $z^k = 0 \forall k \neq l$. From the above transformations, we have the Theorem proved as desired. \square

A.4 Proof of Theorem 4

(a) Let $\hat{L}_{\mathcal{F}\mathcal{G}}(v, w | \mathcal{S}_K)$ be the input requirement set of $\hat{Y}_{\mathcal{F}\mathcal{G}}(\mathcal{S}_K)$ corresponding to the outputs (v, w) . For any (v^*, w^*) and for any $\tilde{x}, \tilde{\tilde{x}} \in \hat{L}_{\mathcal{F}\mathcal{G}}(v^*, w^* | \mathcal{S}_K)$, we will prove that

$$\hat{x} = \alpha \tilde{x} + (1 - \alpha) \tilde{\tilde{x}} \in \hat{L}_{\mathcal{F}\mathcal{G}}(v^*, w^* | \mathcal{S}_K), \quad \forall \alpha \in [0, 1]. \quad (29)$$

Case 1: $w^* \neq 0_J$. Since $w^* \neq 0_J$ and $\tilde{x}, \tilde{\tilde{x}} \in \hat{L}_{\mathcal{F}\mathcal{G}}(v^*, w^* | \mathcal{S}_K)$, there exist numbers $\tilde{\theta}, \tilde{\tilde{\theta}} \in (0, 1]$ and $\tilde{z}^k, \tilde{\tilde{z}}^k \geq 0$ ($k = 1, \dots, K$) such that

$$v^* \leq \tilde{\theta} \sum_{k=1}^K \tilde{z}^k v^k, w^* = \tilde{\theta} \sum_{k=1}^K \tilde{z}^k w^k, \tilde{x} \geq \sum_{k=1}^K \tilde{z}^k x^k, \sum_{k=1}^K \tilde{z}^k = 1; \\ v^* \leq \tilde{\tilde{\theta}} \sum_{k=1}^K \tilde{\tilde{z}}^k v^k, w^* = \tilde{\tilde{\theta}} \sum_{k=1}^K \tilde{\tilde{z}}^k w^k, \tilde{\tilde{x}} \geq \sum_{k=1}^K \tilde{\tilde{z}}^k x^k, \sum_{k=1}^K \tilde{\tilde{z}}^k = 1.$$

Now let $\hat{z}^k = \alpha \tilde{z}^k + (1 - \alpha) \tilde{\tilde{z}}^k$ ($k = 1, \dots, K$) and $\hat{\theta} = (\alpha / \tilde{\theta} + (1 - \alpha) / \tilde{\tilde{\theta}})^{-1}$, then we

have $\hat{z}^k \geq 0 \forall k$, $0 < \hat{\theta} \leq 1$, and

$$\begin{aligned} v^*/\hat{\theta} &= (\alpha/\tilde{\theta} + (1-\alpha)/\tilde{\theta})v^* \leq \alpha \sum_{k=1}^K \tilde{z}^k v^k + (1-\alpha) \sum_{k=1}^K \tilde{\tilde{z}}^k v^k = \sum_{k=1}^K \hat{z}^k v^k, \\ w^*/\hat{\theta} &= (\alpha/\tilde{\theta} + (1-\alpha)/\tilde{\theta})w^* = \alpha \sum_{k=1}^K \tilde{z}^k w^k + (1-\alpha) \sum_{k=1}^K \tilde{\tilde{z}}^k w^k = \sum_{k=1}^K \hat{z}^k w^k, \\ \hat{x} &= \alpha \hat{x} + (1-\alpha)\hat{x} \geq \alpha \sum_{k=1}^K \tilde{z}^k x^k + (1-\alpha) \sum_{k=1}^K \tilde{\tilde{z}}^k x^k = \sum_{k=1}^K \hat{z}^k x^k, \\ \sum_{k=1}^K \hat{z}^k &= \alpha \sum_{k=1}^K \tilde{z}^k + (1-\alpha) \sum_{k=1}^K \tilde{\tilde{z}}^k = \alpha + (1-\alpha) = 1. \end{aligned}$$

Thus, $\hat{x} \in \hat{L}_{\mathcal{F}g}(v^*, w^* | \mathcal{S}_K)$, implying that $\hat{L}_{\mathcal{F}g}(v^*, w^* | \mathcal{S}_K)$ is convex for any (v^*, w^*) where $w^* \neq 0_J$.

Case 2: $w^* = 0_J$. We transform $\hat{L}_{\mathcal{F}g}(v^*, 0_J | \mathcal{S}_K)$ as follows.

$$\hat{L}_{\mathcal{F}g}(v^*, 0_J | \mathcal{S}_K) = \left\{ x : (v^*, 0_J, x) \in \hat{Y}_{\mathcal{F}g}(\mathcal{S}_K) \right\} \quad (30)$$

$$\begin{aligned} &= \left\{ x : v^* \leq \theta \sum_{k=1}^K z^k v^k; 0_J = \theta \sum_{k=1}^K z^k w^k; x \geq \sum_{k=1}^K z^k x^k; \right. \\ &\quad \left. \sum_{k=1}^K z^k = 1; 0 \leq \theta \leq 1; z^k \geq 0, k = 1, \dots, K \right\} \quad (31) \end{aligned}$$

If $w^1 = \dots = w^K = 0_J$, from (31) we have

$$\begin{aligned} \hat{L}_{\mathcal{F}g}(v^*, 0_J | \mathcal{S}_K) &= \left\{ x : v^* \leq \theta \sum_{k=1}^K z^k v^k; x \geq \sum_{k=1}^K z^k x^k; \sum_{k=1}^K z^k = 1; \right. \\ &\quad \left. 0 \leq \theta \leq 1; z^k \geq 0, k = 1, \dots, K \right\} \\ &= \left\{ x : v^* \leq \sum_{k=1}^K z^k v^k; x \geq \sum_{k=1}^K z^k x^k; \sum_{k=1}^K z^k = 1; z^k \geq 0, k = 1, \dots, K \right\}, \end{aligned}$$

which is equivalent to the input requirement sets of the traditional DEA estimation without bad outputs and hence, $\hat{L}_{\mathcal{F}g}(v^*, 0_J | \mathcal{S}_K)$ is convex.

If there exist at least one $w^k \neq 0_J$, from (31) we have $\theta = 0$. As a consequence, if $v^* \neq 0_M$, we get $\hat{L}_{\mathcal{F}g}(v^*, 0_J | \mathcal{S}_K) = \emptyset$, which is convex by convention. On the other

hand, if $v^* = 0_M$, we then have

$$\hat{L}_{\mathcal{F}\mathcal{G}}(0_M, 0_J | \mathcal{S}_K) = \left\{ x : x \geq \sum_{k=1}^K z^k x^k; \sum_{k=1}^K z^k = 1; z^k \geq 0, k = 1, \dots, K \right\}, \quad (32)$$

which is a convex free disposal polyhedron (or hull) in the input-space (\mathbb{R}_+^N).

All in all, we have $\hat{L}_{\mathcal{F}\mathcal{G}}(v^*, w^* | \mathcal{S}_K)$ is convex for any \mathcal{S}_K , completing the proof. \square

(b) We prove the Theorem by showing a counterexample. Consider a dataset $\mathcal{S} = \{(v^*, w^*, x^1), (v^*, w^*, x^2)\}$ where $v^* \in \mathbb{R}_+^M$, $w^* \in \mathbb{R}_+^J$, $x^1 = (2, 1)'$, and $x^2 = (1, 2)'$. Let $\hat{L}_{\mathcal{P}\mathcal{X}}(v^*, w^* | \mathcal{S})$ be the input requirement set of $\hat{Y}_{\mathcal{P}\mathcal{X}}^{QC}(\mathcal{S})$ corresponding to the outputs (v^*, w^*) . Obviously, $x^1, x^2 \in \hat{L}_{\mathcal{P}\mathcal{X}}(v^*, w^* | \mathcal{S})$. We will show that $\hat{L}_{\mathcal{P}\mathcal{X}}(v^*, w^* | \mathcal{S})$ is not convex by indicating that $\hat{x} \notin \hat{L}_{\mathcal{P}\mathcal{X}}(v^*, w^* | \mathcal{S})$ where $\hat{x} = 0.5x^1 + 0.5x^2 = (1.5, 1.5)'$. For this particular scenario, by Theorem 2, we have

$$\begin{aligned} \hat{Y}_{\mathcal{P}\mathcal{X}}^{QC}(\mathcal{S}) = & \left\{ (v, w, x) : v \leq \theta(z^1 v^* + z^2 v^*); w = \theta(z^1 w^* + z^2 w^*); z^1 + z^2 = 1; \right. \\ & \left. z^1, z^2 \geq 0; 0 \leq \theta \leq 1; \text{if } z^k > 0 \text{ then } x \geq x^k, k = 1, 2 \right\} \end{aligned} \quad (33)$$

$$\begin{aligned} = & \left\{ (v, w, x) : v \leq \theta v^*; w = \theta w^*; z^1 + z^2 = 1; \right. \\ & \left. z^1, z^2 \geq 0; 0 \leq \theta \leq 1; \text{if } z^k > 0 \text{ then } x \geq x^k, k = 1, 2 \right\}. \end{aligned} \quad (34)$$

Suppose that $\hat{x} \in \hat{L}_{\mathcal{P}\mathcal{X}}(v^*, w^* | \mathcal{S})$. Then due to (34), we should have $\hat{x} \geq x^1$ or $\hat{x} \geq x^2$ because at least one of z^1, z^2 is greater than 0 since $z^1 + z^2 = 1$. However, by construction, neither $\hat{x} \geq x^1$ nor $\hat{x} \geq x^2$ is true, which is a contradiction. Consequently, $\hat{x} \notin \hat{L}_{\mathcal{P}\mathcal{X}}(v^*, w^* | \mathcal{S})$, implying that $\hat{L}_{\mathcal{P}\mathcal{X}}(v^*, w^* | \mathcal{S})$ is not convex, completing the proof. \square

A.5 Proof of Theorem 5

Here we need to prove that $\hat{Y}_{\mathcal{S}}(\mathcal{S}_K)$ satisfies disposability properties (strong disposability of inputs and good outputs, and jointly weak disposability of good and bad outputs) and convexity of input requirement sets. The disposability properties are ob-

vious. For the convexity of input requirement sets, let $\hat{L}_{\mathcal{J}}(v^*, w^* | \mathcal{S}_K)$ denote the input requirement set of $\hat{Y}_{\mathcal{J}}(\mathcal{S}_K)$ corresponding to an arbitrary pair of outputs (v^*, w^*) . We will prove that $\hat{x} = \alpha \tilde{x} + (1 - \alpha) \tilde{\tilde{x}} \in \hat{L}_{\mathcal{J}}(v^*, w^* | \mathcal{S}_K)$ for any $\tilde{x}, \tilde{\tilde{x}} \in \hat{L}_{\mathcal{J}}(v^*, w^* | \mathcal{S}_K)$ and any $\alpha \in [0, 1]$. To do so, note that there exist numbers $\tilde{\theta}^k, \tilde{\tilde{\theta}}^k \in [0, 1]$ and $\tilde{z}^k, \tilde{\tilde{z}}^k \geq 0$ ($k = 1, \dots, K$) such that

$$\begin{aligned} \tilde{x} &\geq \sum_{k=1}^K \tilde{z}^k x^k, \sum_{k=1}^K \tilde{z}^k = 1, \text{ if } \tilde{z}^k > 0 \text{ then } v^* \leq \tilde{\theta}^k v^k \text{ and } w^* = \tilde{\theta}^k w^k; \\ \tilde{\tilde{x}} &\geq \sum_{k=1}^K \tilde{\tilde{z}}^k x^k, \sum_{k=1}^K \tilde{\tilde{z}}^k = 1, \text{ if } \tilde{\tilde{z}}^k > 0 \text{ then } v^* \leq \tilde{\tilde{\theta}}^k v^k \text{ and } w^* = \tilde{\tilde{\theta}}^k w^k. \end{aligned}$$

Let $\hat{z}^k = \alpha \tilde{z}^k + (1 - \alpha) \tilde{\tilde{z}}^k$ and $\hat{\theta}^k = \min\{\tilde{\theta}^k, \tilde{\tilde{\theta}}^k\}$ ($k = 1, \dots, K$), we have: $0 \leq \hat{\theta}^k \leq 1$, $\hat{z}^k \geq 0$ ($k = 1, \dots, K$); $\hat{x} \geq \sum_{k=1}^K \hat{z}^k x^k$; $\sum_{k=1}^K \hat{z}^k = 1$; if $\hat{z}^k > 0$ then $v^* \leq \hat{\theta}^k v^k$ and $w^* = \hat{\theta}^k w^k \quad \forall k = 1, \dots, K$. Thus, $(v^*, w^*, \hat{x}) \in \hat{Y}_{\mathcal{J}}(\mathcal{S}_K)$, i.e., $\hat{x} \in \hat{L}_{\mathcal{J}}(v^*, w^* | \mathcal{S}_K)$.

Now we prove that $\hat{Y}_{\mathcal{J}}(\mathcal{S}_K)$ is the subset of any technology sets \hat{Y} that contains the dataset \mathcal{S}_K and satisfies disposability properties (strong disposability of inputs and good outputs, jointly weak disposability of good and bad outputs) and convexity of input requirement sets. To do so, consider an arbitrary point (v, w, x) in $\hat{Y}_{\mathcal{J}}(\mathcal{S}_K)$, we prove that $(v, w, x) \in \hat{Y}$ by showing that (v, w, x) can be reached after a finite number of transformations on the dataset based on disposability and convexity properties of \hat{Y} .

Indeed, since $(v, w, x) \in \hat{Y}_{\mathcal{J}}(\mathcal{S}_K)$, there exist $z^k \geq 0$ and $\theta^k \in [0, 1]$ ($k = 1, \dots, K$) such that $x \geq \sum_{k=1}^K z^k x^k$, $\sum_{k=1}^K z^k = 1$, if $z^k > 0$ then $v \leq \theta^k v^k$ and $w = \theta^k w^k$. Firstly for each $k = 1, \dots, K$, we have $(v^k, w^k, x^k) \in \hat{Y}$, then by rescaling outputs of (v^k, w^k, x^k) by θ^k , we obtain point $(\theta^k v^k, \theta^k w^k, x^k)$ which also belongs to \hat{Y} , or equivalently, $(\theta^k v^k, w, x^k) \in \hat{Y}$. Next, by properties of strong disposability of good outputs, we have $(v, w, x^k) \in \hat{Y}$ for $k = 1, \dots, K$. Now by convexity of input requirement sets, we get $(v, w, \sum_{k=1}^K z^k x^k) \in \hat{Y}$. Finally, by strong disposability of inputs, we also get $(v, w, x) \in \hat{Y}$, which completes the proof. \square

Appendix B Proof of Lemma 1

This proof is in the spirit of Shephard (1953) (p. 182). Here we present it in the context of bad outputs for convenience. The proof of part (a) is provided below while that of part (b) is not presented since it is similar to part (a).

$$\begin{aligned}
P \text{ is quasiconcave} &\Leftrightarrow P((1 - \lambda)\tilde{x} + \lambda\tilde{\tilde{x}}) \supseteq P(\tilde{x}) \cap P(\tilde{\tilde{x}}) \quad \forall \tilde{x}, \tilde{\tilde{x}} \in \mathbb{R}_+^N, \forall \lambda \in [0, 1] \\
&\Leftrightarrow (v, w) \in P((1 - \lambda)\tilde{x} + \lambda\tilde{\tilde{x}}) \quad \forall (v, w) \in P(\tilde{x}) \cap P(\tilde{\tilde{x}}), \tilde{x}, \tilde{\tilde{x}} \in \mathbb{R}_+^N, \forall \lambda \in [0, 1] \\
&\Leftrightarrow (v, w) \in P((1 - \lambda)\tilde{x} + \lambda\tilde{\tilde{x}}) \quad \forall (v, w) \in P(\tilde{x}) \wedge (v, w) \in P(\tilde{\tilde{x}}), \tilde{x}, \tilde{\tilde{x}} \in \mathbb{R}_+^N, \forall \lambda \in [0, 1] \\
&\Leftrightarrow (1 - \lambda)\tilde{x} + \lambda\tilde{\tilde{x}} \in L(v, w) \quad \forall \tilde{x} \in L(v, w) \wedge \tilde{\tilde{x}} \in L(v, w), \forall \lambda \in [0, 1], \forall (v, w) \in \mathbb{R}_+^M \times \mathbb{R}_+^J \\
&\Leftrightarrow (1 - \lambda)\tilde{x} + \lambda\tilde{\tilde{x}} \in L(v, w) \quad \forall \tilde{x}, \tilde{\tilde{x}} \in L(v, w), \forall \lambda \in [0, 1], \forall (v, w) \in \mathbb{R}_+^M \times \mathbb{R}_+^J \\
&\Leftrightarrow L(v, w) \text{ is convex for all } (v, w) \in \mathbb{R}_+^M \times \mathbb{R}_+^J
\end{aligned}$$

Therefore, that the output correspondence P is quasiconcave is equivalent to that all input requirement sets of Y is convex. □

Appendix C Graphical illustration of reference technology sets in three-dimensional space

This illustration is drawn using a sample in Kuosmanen and Podinovski (2009): $\mathcal{S}_D = \{D_1, D_2\}$ where $D_1 = (3, 4, 1)$ and $D_2 = (5, 1, 4)$. Because of the strong disposability of input, $\hat{Y}_{\mathcal{K}}(\mathcal{S}_D)$ and $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_D)$ are alike in the region where $x > 4$. Meanwhile, for $1 \leq x \leq 4$, the surfaces of the reference technology sets are different: the surface corresponding to $\hat{Y}_{\mathcal{K}}(\mathcal{S}_D)$ comprises flat facets and the set of points under these facets is convex (Figure 3a) whereas the surface of $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_D)$ is nonconvex and lying under that of $\hat{Y}_{\mathcal{K}}(\mathcal{S}_D)$ (Figure 3b). The surface of $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QC}(\mathcal{S}_D)$ also lies under that of $\hat{Y}_{\mathcal{F}g}(\mathcal{S}_D)$ (Figure 3c) and finally, $\hat{Y}_{\mathcal{P}\mathcal{K}}^{QFDH}(\mathcal{S}_D)$ appears to be the smallest set of the four (Figure 3d).

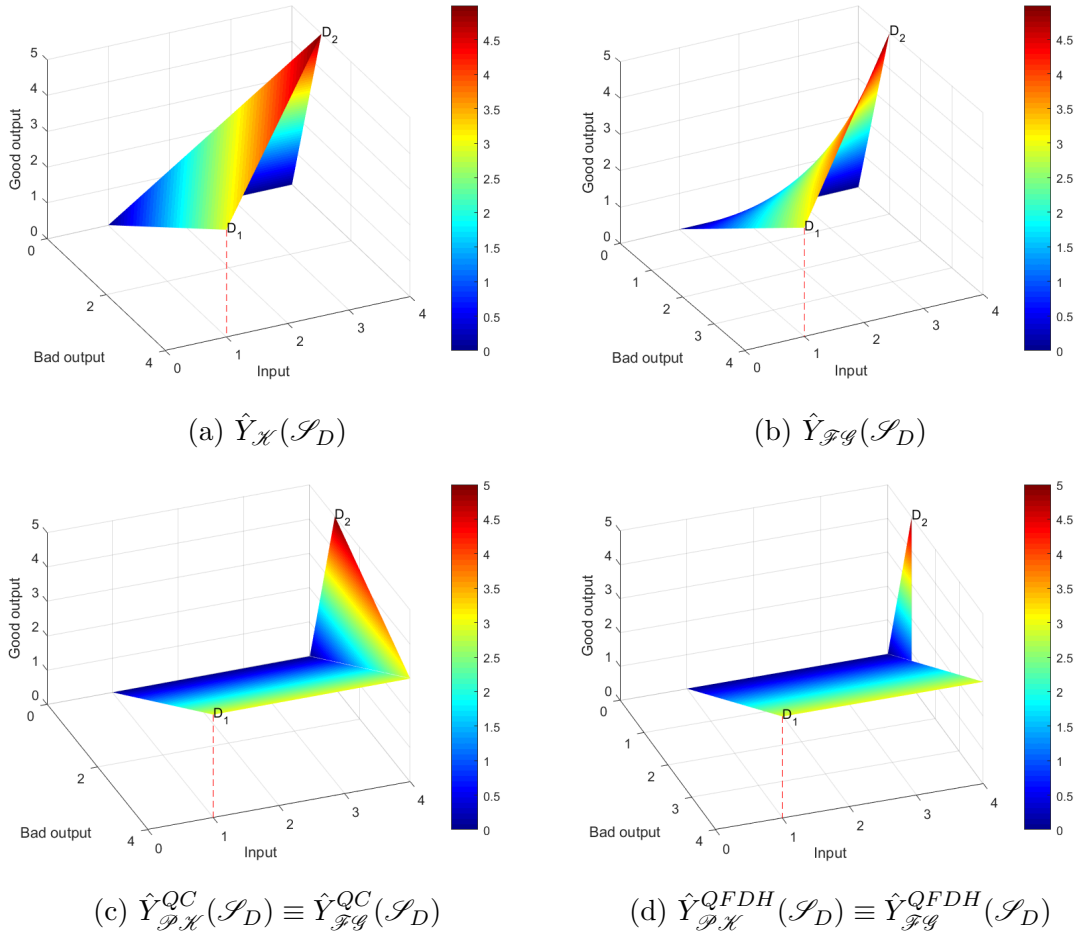


Figure 3: Graphical illustration of surfaces of reference technology sets

References

- Arocena, P. and García-Prado, A. (2007). Accounting for quality in the measurement of hospital performance: evidence from Costa Rica. *Health Economics*, 16(7):667–685.
- Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9):1078–1092.
- Banker, R. D. and Maindiratta, A. (1986). Piecewise loglinear estimation of efficient production surfaces. *Management Science*, 32(1):126–135.
- Baumol, W. J. and Oates, W. E. (1975). *The Theory of Environmental Policy*. Cambridge University Press, 1st edition.
- Baumol, W. J. and Oates, W. E. (1988). *The Theory of Environmental Policy*. Cambridge University Press, 2nd edition.
- Bogetoft, P. (1996). DEA on Relaxed Convexity Assumptions. *Management Science*, 42(3):457–465.
- Bogetoft, P., Tama, J. M., and Tind, J. (2000). Convex input and output projections of nonconvex production possibility sets. *Management Science*, 46(6):858–869.
- Borts, G. H. and Mishan, E. J. (1962). Exploring the “uneconomic region” of the production function. *The Review of Economic Studies*, 29(4):300–312.
- Caves, D. W., Christensen, L. R., and Diewert, W. E. (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. *The Economic Journal*, 92(365):73–86.
- Chambers, R., Chung, Y., and Färe, R. (1998). Profit, Directional Distance Functions, and Nerlovian Efficiency. *Journal of Optimization Theory and Applications*, 98(2):351–364.
- Chambers, R. G., Chung, Y., and Färe, R. (1996). Benefit and distance functions. *Journal of Economic Theory*, 70(2):407–419.
- Chung, Y. H., Färe, R., and Grosskopf, S. (1997). Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management*, 51(3):229–240.
- Deprins, D., Simar, L., and Tulkens, H. (1984). *The performance of public enterprises*, volume 33 of *Studies in Mathematical and Managerial Economics*, chapter Measuring Labour efficiency in Post Offices. North-Holland, Amsterdam.
- Diewert, W. E. (1983). The measurement of waste within the production sector of an open economy. *The Scandinavian Journal of Economics*, 85(2):159–179.
- Dorfman, R., Samuelson, P. A., and Solow, R. M. (1958). *Linear programming and economic analysis*. Courier Corporation.
- Färe, R. (1988). *Fundamentals of production theory*. Lecture Notes in Economics and Mathematical Systems. Springer.

- Färe, R. and Grosskopf, S. (1983a). Measuring congestion in production. *Zeitschrift für Nationalökonomie / Journal of Economics*, 43(3):257–271.
- Färe, R. and Grosskopf, S. (1983b). Measuring output efficiency. *European Journal of Operational Research*, 13(2):173 – 179.
- Färe, R. and Grosskopf, S. (2003). Nonparametric productivity analysis with undesirable outputs: Comment. *American Journal of Agricultural Economics*, 85(4):1070–1074.
- Färe, R. and Grosskopf, S. (2009). A comment on weak disposability in nonparametric production analysis. *American Journal of Agricultural Economics*, 91(2):535–538.
- Färe, R., Grosskopf, S., and Lovell, C. A. K. (1983). The structure of technical efficiency. *The Scandinavian Journal of Economics*, 85(2):181–190.
- Färe, R., Grosskopf, S., Lovell, C. A. K., and Pasurka, C. (1989). Multilateral Productivity Comparisons When Some Outputs are Undesirable: A Nonparametric Approach. *The Review of Economics and Statistics*, 71(1):90.
- Färe, R., Grosskopf, S., and Lovell, C. K. (1985). *The Measurement of Efficiency of Production*, volume 6. Springer.
- Färe, R., Grosskopf, S., and Lovell, C. K. (1994). *Production frontiers*. Cambridge University Press.
- Färe, R., Grosskopf, S., Noh, D.-W., and Weber, W. (2005). Characteristics of a polluting technology: theory and practice. *Journal of Econometrics*, 126(2):469–492.
- Färe, R. and Primont, D. (1995). *Multi-Output Production and Duality: Theory and Applications*. Springer.
- Färe, R., Grosskopf, S., and Pasurka, C. (1986). Effects on relative efficiency in electric power generation due to environmental controls. *Resources and Energy*, 8(2):167 – 184.
- Färe, R. and Svensson, L. (1980). Congestion of production factors. *Econometrica*, 48(7):1745–1753.
- Fukuyama, H. and Weber, W. L. (2010). A slacks-based inefficiency measure for a two-stage system with bad outputs. *Omega*, 38(5):398 – 409. Empirical Research in the EU Banking Sector and the Financial Crisis.
- Hailu, A. and Veeman, T. S. (2001). Non-parametric productivity analysis with undesirable outputs: An application to the canadian pulp and paper industry. *American Journal of Agricultural Economics*, 83(3):605–616.
- Kneip, A., Park, B. U., and Simar, L. (1998). A Note on the Convergence of Non-parametric DEA Estimators for Production Efficiency Scores. *Econometric Theory*, 14(6):783–793.
- Korostelev, A., Simar, L., and Tsybakov, A. (1995). Efficient estimation of monotone boundaries. *Annals of Statistics*, 23(2):476–489.

- Kuosmanen, T. (2005). Weak disposability in nonparametric production analysis with undesirable outputs. *American Journal of Agricultural Economics*, 87(4):1077–1082.
- Kuosmanen, T. and Podinovski, V. (2009). Weak Disposability in Nonparametric Production Analysis: Reply to Färe and Grosskopf. *American Journal of Agricultural Economics*, 91(2):539–545.
- Maxwell, W. D. (1965). Short-run returns to scale and the production of services. *Southern Economic Journal*, 32(1):1–14.
- McFadden, D. (1978). Cost, revenue, and profit functions. In Fuss, M. and McFadden, D., editors, *Production Economics: A Dual Approach to Theory and Applications*, volume 1, chapter 1. McMaster University Archive for the History of Economic Thought.
- Park, B. U., Simar, L., and Weiner, C. (2000). The FDH Estimator for Productivity Efficiency Scores: Asymptotic Properties. *Econometric Theory*, 16(6):855–877.
- Park, K. H. and Weber, W. L. (2006). A note on efficiency and productivity growth in the Korean Banking Industry, 1992–2002. *Journal of Banking & Finance*, 30(8):2371 – 2386.
- Petersen, N. C. (1990). Data envelopment analysis on a relaxed set of assumptions. *Management Science*, 36(3):305–314.
- Pittman, R. W. (1983). Multilateral productivity comparisons with undesirable outputs. *The Economic Journal*, 93(372):883–891.
- Podinovski, V. V. and Kuosmanen, T. (2011). Modelling weak disposability in data envelopment analysis under relaxed convexity assumptions. *European Journal of Operational Research*, 211(3):577–585.
- Samuelson, P. A. (1947). *Foundations of economic analysis*. Harvard University Press.
- Shephard, R. W. (1953). *Cost and Production functions*. Princeton University Press.
- Shephard, R. W. (1970). *Theory of Cost and Production Functions*. Princeton University Press.
- Shephard, R. W. (1974). *Indirect Production Functions*. Verlag Anton Hain, Meisenheim am Glan.
- Sitorus, B. L. (1966). Productive efficiency and redundant factors of production in traditional agriculture of underdeveloped countries: A note on measurement. *Proceedings of the Annual Meeting (Western Farm Economics Association)*, 39:153–158.
- Sueyoshi, T., Yuan, Y., and Goto, M. (2017). A literature study for DEA applied to energy and environment. *Energy Economics*, 62:104 – 124.
- Tyteca, D. (1996). On the measurement of the environmental performance of firms - a literature review and a productive efficiency perspective. *Journal of Environmental Management*, 46(3):281–308.
- Zhou, P., Ang, B., and Poh, K. (2008). Measuring environmental performance under different environmental DEA technologies. *Energy Economics*, 30(1):1–14.