Louvain School of Management

Efficiency Benchmarking of the Swedish Distribution System Operators Using the Convex Pairs Approach

Auteur : Simon Van Kerckhove
Promoteur: Per J. Agrell
Année académique 2019-2020

Abstract

The objective of this thesis is to implement the Convex Pairs approach, using an alternative convexity hypothesis to benchmark Swedish electricity distributors (DSO). Three models are developed, the first two being the classical approaches under the hypothesis of the Free Disposability Hull and of Varying Returns to Scale. The third is developed as an algorithm that creates artificial Convex Pairs based on the real observations, using blocking and polyhedron theories. It has the advantage to neglect the assumption of convexity in the technology set, which reflects the economic reality in some situations. An efficiency index is then implemented by solving as many linear programming problems as there are firms analysed. We implemented this algorithm for 40 DSOs characterised by 1 input and 5 output variables. In total, 174 non-dominated pairs were kept among all those created by the algorithm. The median efficiency using the new approach is found between those for FDH and VRS efficiency. However, the difference between the convex pairs approach efficiency and the VRS efficiency is too large to use the two as equivalent, depending on the convexity assumption required.

Acknowledgements

I would like to thank Pr Per Agrell for introducing me to the interesting subject of this thesis.

My thanks particularly go to my Operations and Research professors in Norway and Louvain-la-Neuve for their invaluable help.

A final thought is for my loved ones, I thank their continued support and unconditional encouragement, which allowed me to carry out this work. Thank you.

Declaration

UCLouvain
Louvain School of Management

Thesis submitted in partial fulfilment

of the requirements for the Degree of

Master in Business Engineering

I hereby certify that this is entirely my own work unless otherwise stated.

Brussels, January 2020.

Contents

List of Figures

Chapter 1

Introduction to benchmarking and regulation

1.1 Data Envelopment Analysis, the origins.

Charnes, Cooper, and Rhodes (1978) introduced the data envelopment analysis (DEA) method to extend the business cost efficiency measure developed by Farrell (1957). Those ideas are now widely accepted and still in use for applications in evaluating public programs. This therm refers to a set of so-called Decision Making Units which share a common set of inputs and outputs. Application of the models demands a fair transparency in data collection. Therefore, it is particularly suitable for public sector applications as the variables chosen may not be monetisable *stricto sensus*.

In the decades following the original publication, economists and management scientists extended this work into various mathematical programming models to evaluate *a posteriori* the relative efficiency of managerial decisions. The most general model, the *Free Disposability Hull*, is proposed by Deprins, Simar, and Tulkens (1984) and it draws the efficiency frontier that is the closest to the DMU.

Such model already calls for further development because this frontier may not be a neoclassical one, hence preventing economic application. Diving deeper into details, R. D. Banker, Charnes, and Cooper (1984) brings the possibility to conduct analysis in the case of increasing, constant or decreasing returns to scale.

However, the assumption of a convex production possibility set proposed by Charnes et al. is inconsistent with the condition of increasing returns to scale. This assumption may be relaxed and Petersen (1990) provides an alternative approach which invokes convex inputs and outputs sets rather than convexity of the full production possibility set. Several methodologies have been developed subsequently to impose convex isoquants. Specifically, Agrell, Bogetoft, Brock, and Tind (2005a) propose the Convex Pairs Approach, the focus of this thesis, that uses pairs of associated convex input and output sets. Arguments for this new method is that it is consistent with the IRS condition and with micro-economic theory, which is important when considering the application of the models. Also, it circumvent multiple weaknesses inherent to previous models.

1.2 DSO benchmarking: introduction and motivations.

The Distribution System Operators, or DSOs, typically experience large fixed costs and relatively low marginal cost. The fixed cost is incurred when building the infrastructure like transmission lines and the network substations. In addition, if the firms operate in remote or mountainous geographical areas, the low network density increases the cost in comparison to their revenue.

Because of such market characteristics, the DSOs are natural monopolies and considered as legal monopolies. Obviously, monopolies do not try to reduce their costs since they are not subject to competition. In addition, there are no close substitute to electricity supply and the demand is relatively inelastic, which makes the DSOs' position relatively dominant.

For this reason, regulators have been introduced as a "proxy purchaser of services, imposing constraints on the prices and the modalities of the production" (Agrell & Bogetoft, 2016). However, production and price setting depends on a large array of parameters, and the different DSOs evolve in heterogeneous environments. Setting standards and objective is a very difficult task, and the firms usually benefit from a informational rent, which is a prejudice to the society.

In the European energy transmission and distribution regulation landscape, Data Envelopment Analysis has now been used for more than 20 years and this technique is well established. Indeed, a strength of DEA is precisely that it can cope with the diversity in firms and still provide consistent comparison. The quality of the comparison, however, depends on the number of firms we can dispose of for analysis.

1.2.1 Applied case: Swedish DSO.

In Sweden, the market has been reformed in 1996; generation and trading of electricity is open to competition, where electricity transportation is still considered as a natural monopoly (Wallnerström et al., 2016). The Swedish Electricity Inspectorate is therefore responsible to create incentives for fair prices, quality and reliability of supply.

Since 2016, the regulation regime for the DSO is the revenue cap, which defines "the total amount that the DSO may charge their customers" (Wallnerström et al., 2016). The objective of this methodology is to ensure that the DSO do not benefit unfairly from their informational rent and obtain excessive cost coverage and ROI. Depending on the Inspectorate assessment, the revenue cap can be adjusted.

The data we will base our analysis on is already anterior to the current regulation. In addition, the power system is expected to experiment major changes in a near future, since decentralised and renewable energy generation will create new supply patterns. Modifications will also happen on the demand side as demand response and storage will modify the load profile.

1.2.2 Convexity assumption and research question.

Benchmarking is a technique that allows to compare firms between themselves and provide a "mentor unit" they can learn from. Various assumptions, including the convexity assumption, allow to increase the information we can obtain from the actual data and draw conclusions on efficiency at different levels of confidence. This convexity assumption is presumed by all the classical DEA models. However, there are different reasons to challenge this assumption, and interesting opportunities in doing so.

The objective of this thesis is to provide a review of the building blocks of DEA benchmarking, while developing some of the convexity relaxation works that have been done by several authors. The present research focuses on the one proposed by Agrell et al. (2005a) and the research topic is the following; Efficiency Benchmarking of Distribution System Operators using the Convex Pairs approach. This work will be mainly exploratory on the side of other DEA assumptions. The application is on the Swedish DSO since it has already been heavily benchmarked, but is not intended to provide a turnkey solution for regulators.

The outline of the thesis is as follows. In Chapter 2, we introduce the reader to the basics of benchmarking and production analysis and its main assumptions. We also provide an introduction to the relaxation of the convexity assumption and the core characteristics of the new approach. In Chapter 3, we propose the approach we will use, that will focus on 2 classic models and the new technique. It will develop the proposition from Agrell et al. (2005a) to alter production assumptions by neglecting global convexity while still presuming local convexity in the input and output spaces. To illustrate the Convex Pairs approach, we present the empirical data we will use for implementation, including the data preparation and the actual results of the models in Chapter 4. We conclude our research in Chapter 5.

Chapter 2

Methodology

In this chapter, we review the fundamentals of production analysis, benchmarking and Data Envelopment Analysis. The technologies and assumptions developed here are the building blocks necessary to develop further the new model proposed. To give some context to the approach proposed in this thesis, the methodology proposed by several prominent authors is reviewed at the end of this chapter and we will see how the new approach fits into the previous ones.

2.1 Benchmarking.

Benchmarking entities equals to relatively evaluate their performance, or more formally is the "systematic comparison of the performance of one firm against other firms" (Bogetoft & Otto, 2011). A common motivation for benchmarking is, among others, the introduction of regulation in a specific industry by conducting an inter-organisations comparison.

In this chapter, we will define the basics of productivity analysis.

2.1.1 Firm and production plan.

From Bogetoft and Otto (2011), we establish that a firm, that we will call a *Decision-Making Unit* or (*DMU*), transforms inputs into outputs by choosing a *Production Plan*, i.e. a combination of inputs and outputs.

Consider K firms indexed on $k = 1, \dots, K$. To produce n outputs $y^k = (y_1^k, ..., y_n^k) \in \mathbb{R}_+^m$, the n vector, firm k has used m inputs $x^k = (x_1^l, ..., x_m^k) \in \mathbb{R}_+^m$, the m vector.

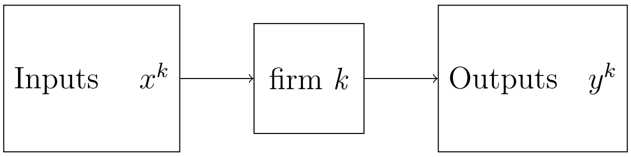


Figure 2.1: A simple firm.

To relatively evaluate firms' production plans, we assume they use a common underlying *technology* or *production possibility set T*. T is unknown as determined by various factors, including the social and technical environment in which firms operate. We can estimate T based on observed input-output combinations. T becomes the smallest set containing actual data, formally defined here:

$$T = \Big\{ (x, y) \in \mathbb{R}_+^m \times \mathbb{R}_+^n \Big| \exists k \in \{1, \dots, K\} : (x, y) = (x^k, y^k) \Big\} \tag{2.1}$$

Moreover, $T = \mathbb{R}_+^m \times \mathbb{R}_+^n$ is the largest possible technology set in which anything can produce anything, which is not realistic.

Therefore, the technology in Equation 2.1 is the starting point of productivity analysis. Assumptions will be added to the model to enlarge the set between those two extremes and allow interpolation and extrapolation of the actual data. Those assumptions will be defined in a later chapter of this section.

2.1.2 Dominance.

Still using Bogetoft and Otto (2011), the concept of *dominance* allows to introduce a partial ranking of firms. It is a weak expression of preferences because almost everyone would prefer a dominating firm than another.

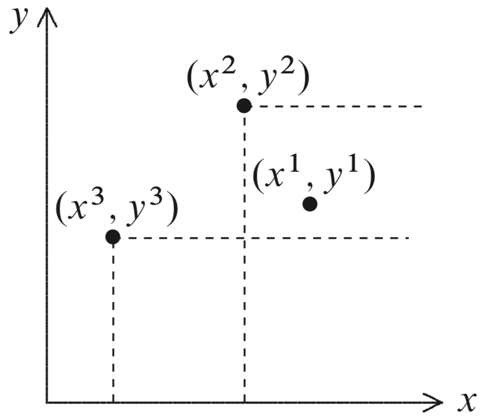


Figure 2.2: Dominance between three firms. (Bogetoft & Otto, 2011).

We consider two firms in the simple configuration of a one-input, one-output production plan (x^1, y^1) and (x^2, y^2). Firm 2 dominates firm 1 if it does strictly better in at least one dimension, formally if and only if $x^2 \leq x^1, y^2 \geq y^1$ and $(x^1, y^1) \neq (x^2, y^2)$, i.e. if firm 2 uses strictly one input less and/or produces strictly one more output than firm 1. A graphical example is provided in Figure 2.2.

The concept of dominance will be particularly useful for the new method. Indeed, we will be creating new pairs of input and output based on actual ones, and the concept of dominance will help us neglect the pairs that are already included in the technology.

2.1.3 Efficiency.

Dominance leads us to the definition of efficiency. It is convenient to gauge performance in a widely accepted way. A firm's production plan $(x, y) \in T$ is efficient in a given technology set T if it cannot be dominated by any other production plan $(x', y') \in T$. The efficient subset of T, T^E is

$$T^E = \{(x, y) \in T | \text{(x,y) is efficient in T}\}.$$

We obtain the efficient subset T^E of T, i.e. the production plans of all the dominant firms.

2.1.4 Farrell Efficiency.

Because we need to go beyond the binarity of (inn-)efficiency, we need a measure of the extend of efficiency. The approach proposed is referred as the Farrell efficiency. It is a multi-input, multi-output setting where the possibility to reduce the input without changing the output is considered.

The *input-based Farell efficiency* of a plan (x,y) relative to a technology T is defined as the maximal proportional contraction of all inputs x necessary to produce y, formally,

$$E = min\{E > 0 | (Ex, y) \in T\}$$

In the same way, the *output-based Farell efficiency* is defined as the maximal proportional expansion of all outputs y produced by x, formally

$$F = max\{F > 0 | (x, Fy) \in T\}$$

.

Though we may build the technology on a basis that is slightly more complex that the one in this section, we will still rely on the Farell efficiency for our analysis.

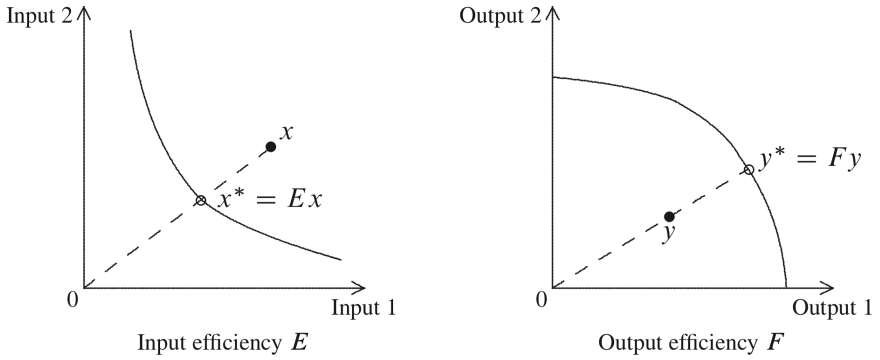


Figure 2.3: Input- and output-based Farell efficiency. (Bogetoft & Otto, 2011).

2.2 Data Envelopment Analysis.

2.2.1 DEA

We can define Data Envelopment Analysis as the method developed by Charnes, Cooper and Rhodes in 1978 as "a methodology for measuring the relative efficiency of units performing similar tasks in a production system that transforms multiple inputs into multiple outputs"(Petersen, 1990). We add with Bogetoft and Otto (2011) that it "provides a mathematical programming method of estimating best practice production frontiers and evaluating the relative efficiency of different entities" to include the notion of frontier, central in the method.

Scholars agree on the main strength of the technique, i.e. that it requires no or little preference, price or technological information and handles multiple inputs and output. Most of all, DEA allows for model flexibility, which is crucial in adapting to the data *mean structure*, or shape, in the context of an industry-wide study (Agrell & Bogetoft, 2007).

2.2.2 The minimal extrapolation principle.

Based on the previous definitions, we know the technology T is not revealed and we need to estimate it based on actual data. The following principle allows the construction of T*, the estimation of the technology. We consider the candidate technology T', the smallest subset of $\mathbb{R}_+^m \times \mathbb{R}_+^n$ that contains the data $(x^k, y^k), k = 1, \ldots, K$ (D) and satisfies the regularity assumptions. Formally, the set of all the technologies is

$$\mathcal{T} = \left\{ T' \subset \mathbb{R}_+^m \times \mathbb{R}_+^m \middle| \text{T' satisfies (D) and (R)} \right\}$$

and the *minimal extrapolation principle* gives us the estimation of the underlying technology as

$$T^* = \bigcap_{T' \in \mathcal{T}} T'$$

The approximation that we will develop will be a subset of the true technology, $T^* \subseteq T$. We refer to this as an *inner approximation* of the technology.

2.2.3 Assumptions of the current approach.

All DEA methods share the minimal extrapolation principle, but differ in the ex-ante assumptions they use to represent technology. The choice of assumptions used is crucial as it will affect the relevance and support of the subsequent DEA analysis.

Traditionally, we assume 4 assumptions, which are the following.

Assumption 1: Free disposability of inputs and outputs.

This first assumption states that "we can always produce fewer outputs with more inputs"(Bogetoft & Otto, 2011): when $(x, y) \in T, x' \geq x$, and $y' \leq y$, then $(x', y') \in$

T i.e.

$$(x, y) \in T, x' \geq x, y' \leq y \Rightarrow (x', y') \in T$$

It also gives an idea of the feasibility of an input-output combination (the dashed area in figure 2.4). Are considered feasible all the input-output combinations that are below and right of an observed combination, i.e. they constitute a technology set T that is called the *free disposable hull.*

We can write the equation of T as in equation 2.1. Formally, an input-output combination $(x, y) \in T$, i.e. is feasible if and only if there exists an observation (x^k, y^k) for $k \in 1, \ldots, K$ such that $x \geq x^k$ and $y \leq y^k$; i.e.,

$$T = \left\{(x, y) \in \mathbb{R}_+^m \times \mathbb{R}_+^n | \exists k \in \{1, \ldots, K\} : x \geq x^k, y \leq y^k\right\} \tag{2.2}$$

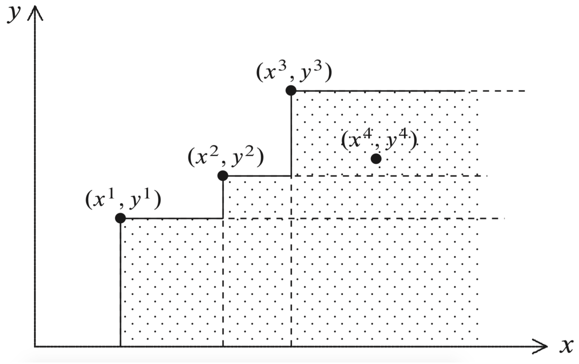


Figure 2.4: Free disposability: 4 firms (Bogetoft & Otto, 2011).

Assumption 2: Convexity.

Particularly useful when, as in the figure above, the number of firms is low, the convexity assumption allows to enlarge the technology set and increase the discriminatory power, i.e. better discriminate between average performance and best practices. Some motivations for convexity are trivial. It includes, in essence, when different internal processes allow the firm to choose among different input-output combinations, or when we need to approximate various aggregations possibilities of subunits (Bogetoft & Otto, 2011).

However, other models also invoke less convexity assumptions, e.g. Agrell, Bogetoft, Brock, and Tind (2005b) that we will use further to develop the mathematical model, but also Bogetoft, Tama, and Tind (2000b), Bogetoft (1996) and Petersen (1990). Since convexity is the assumption challenged by the new approach discussed here, we also consider the applied and theoretical objections to a full convexity assumption. In substance, firms prefer to be evaluated using a weaker convexity assumption as it relies on the possibility of mixed organisations, which does not show them from their best angle. Also, convexity requires divisibility, impossible when a firm considers different idiosyncratic investment plans, and this assumption is not harmless as it does not consider the economies of scale, present in many industries (Bogetoft & Otto, 2011).

Nevertheless, the convexity assumption is used in most benchmarking models. We ought therefore to define it in order to better understand and discuss it later. Intuitively, it means that all the plans on a line between two observed points in the technology T are also part of T. We can say that all weighted averages of two feasible points are also feasible. Moreover, we can also create convex combinations of the convex combinations based on the existing points and so forth. The result is shown in Figure 2.5.

From equation 2.2, we rewrite the equation of the technology set T based on K observations with the convexity assumption as

$$T = \left\{ \Big(\sum_{k=1}^{K} \lambda^k x^k, \sum_{k=1}^{K} \lambda^k y^k \Big) \Big| \sum_{k=1}^{K} \lambda^k = 1, \lambda^k \geq 0, k = 1, \ldots, K \right\} \tag{2.3}$$

for any weight $0 \leq \lambda \leq 1$ and any observed production plan in K. T is therefore the *smallest convex set that contains the K observations* and has, as we can see in

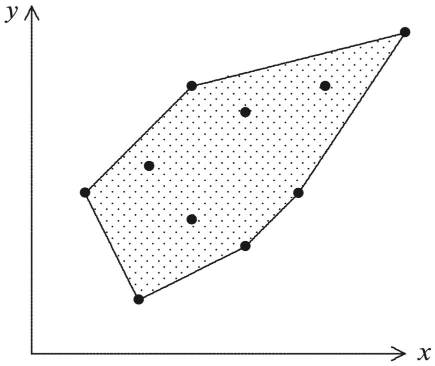


Figure 2.5: Convex hull (Bogetoft & Otto, 2011).

Figure 2.5, the shape of a boat's hull. Hence the name.

Assumption 3: Free disposal and convex.

We can combine assumption 1 and 2 to enlarge the technology, which is currently restricted to exactly match the convex hull. In this new configuration, "we only need weakly more input x and weakly less output y to ensure feasibility" (Bogetoft & Otto, 2011). From equation 2.3, the equation of the technology T based on K observations with the convexity and the free disposability of input and output assumptions is

$$T = \left\{ (x, y) \middle| x \geq \sum_{k=1}^{K} \lambda^k x^k, y \leq \sum_{k=1}^{K} \lambda^k y^k, \sum_{k=1}^{K} \lambda^k = 1, \lambda^k \geq 0, k = 1, \ldots, K \right\}. \quad (2.4)$$

We obtain the *convex and free disposal hull technology* T, the shaded area in Figure 2.6.

This technology set T is the smallest set containing all the K observations and that is convex and free disposable.

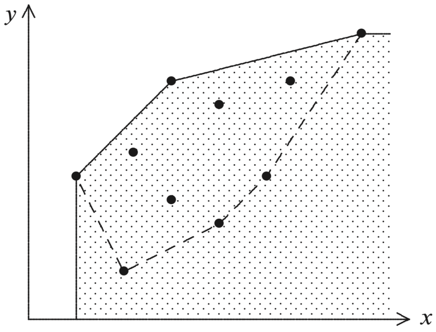


Figure 2.6: Convex technology set with free disposability of inputs and outputs (Bogetoft & Otto, 2011).

Assumption 4: Scaling and Additivity.

1. Scaling.

We can introduce a last set of assumptions that are the options of scaling operations. From the smallest technology set above, those assumptions are crucial since they will enlarge the set in various ways, having substantial consequences on how firms are presented and on their improvement possibilities. A larger set means more improvement possibilities for the firms, but they also look less efficient. Both firms and regulators have specific reasons to prefer an assumption to another. We will discuss those choices in the next section while we quickly introduce the concept here.

The idea of scaling comes from the possibility to produce slightly less outputs with slightly less inputs, and vice versa.

Scaling possibilities basically range from the smallest to the largest technology set, depending on the assumptions we use. At the small end, we have the *Free Disposability Hull* or FDH that we already know where no scaling is possible. With convexity, the set is larger but still no scaling is possible. We call this assumption *Varying Returns to Scale* or VRS. At the other end stands the *Constant Return to Scale* or CRS assumption, the largest set possible. In between are the non-increasing or conveniently *Decreasing Return to Scale* (DRS) and non-decreasing or conveniently *Increasing Return to Scale* (IRS) assumptions.

When scaling is possible; if $(x, y) \in T$, then $\lambda(x, y) \in T$ for λ in the neighbourhood of 1. The value that λ can take depends on the assumption chosen (Bogetoft & Otto, 2011).

DRS.

In any production process, we can arbitrarily decrease the scale of operations. The output increasing slower than the inputs, downscaling is possible, but not upscaling:

$$(x, y) \in T, 0 \leq \lambda \leq 1 \Rightarrow \lambda(x, y) \in T$$

IRS.

In any production process, we can arbitrarily increase the scale of operations. The output increasing faster than the input, upscaling is possible, but not downscaling:

$$(x, y) \in T, \lambda \geq 1 \Rightarrow \lambda(x, y) \in T$$

CRS.

This assumption is the addition of the two previous ones, i.e. when in any production process, we can arbitrarily increase or decrease the scale of operations:

$$(x, y) \in T, \lambda \geq 0 \Rightarrow \lambda(x, y) \in T$$

2. Additivity.

A last commonly used assumption is additivity or replicability, which states that the sum of two feasible production plans is also feasible, formally

$$(x, y) \in T, (x', y') \in T \Rightarrow (x + x', y + y') \in T$$

Though appealing as being able to link the previous assumptions together, the additivity assumption needs Mixed Integer Programming to work with.

2.2.4 DEA technologies.

In this section, we formalise the expression of DEA technologies we will use to construct the models. From the diagram below, we see the links between the various assumptions stated above, which have in common the use of the minimal extrapolation principle.

In Figure 2.7, we can clearly see that the FDH is the smallest technology set possible and that it neglects the global convexity assumption. The second smallest, however, assumes global convexity and the efficiency frontier is composed of convex combinations of existing pairs. Again, we will see where the frontier created using the new method lies, but we can expect it to be somewhere in between the FDH and the VRS technologies.

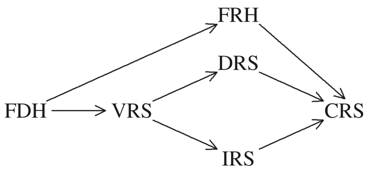


Figure 2.7: Technologies ranking (Bogetoft & Otto, 2011).

Using Equation 2.4 and the different values λ can take, we rewrite once again the minimal extrapolation expression of the technology T^* for each of the 4 models γ we retained,

$$T^*(\gamma) = \left\{ (x, y) \in \mathbb{R}_+^m \times \mathbb{R}_+^n \middle| \exists \lambda \in \Lambda^K(\gamma) : x \geq \sum_{k=1}^{K} \lambda^k x^k, y \leq \sum_{k=1}^{K} \lambda^k y^k \right\} \tag{2.5}$$

where

$$\Lambda^K(fdh) = \left\{ \lambda \in \mathbb{N}_+^K \middle| \sum_{k=1}^{K} \lambda^k = 1 \right\}$$
$$\Lambda^K(vrs) = \left\{ \lambda \in \mathbb{R}_+^K \middle| \sum_{k=1}^{K} \lambda^k = 1 \right\}$$
$$\Lambda^K(drs) = \left\{ \lambda \in \mathbb{R}_+^K \middle| \sum_{k=1}^{K} \lambda^k \leq 1 \right\}$$
$$\Lambda^K(irs) = \left\{ \lambda \in \mathbb{R}_+^K \middle| \sum_{k=1}^{K} \lambda^k \geq 1 \right\}$$
$$\Lambda^K(crs) = \left\{ \lambda \in \mathbb{R}_+^K \middle| \sum_{k=1}^{K} \lambda^k \geq 0 \right\} = \mathbb{R}_+^K$$
$$\Lambda^K(frh) = \left\{ \lambda \in \mathbb{N}_+^K \middle| \sum_{k=1}^{K} \lambda^k \geq 0 \right\} = \mathbb{N}_+^K$$

We recall here that the technology expressed is the estimation of the true technology T, hence the use of the * in the notation. $T^*(\gamma)$ is derived from K observations and built using the minimal extrapolation principle as well as the 4 core assumptions of DEA.

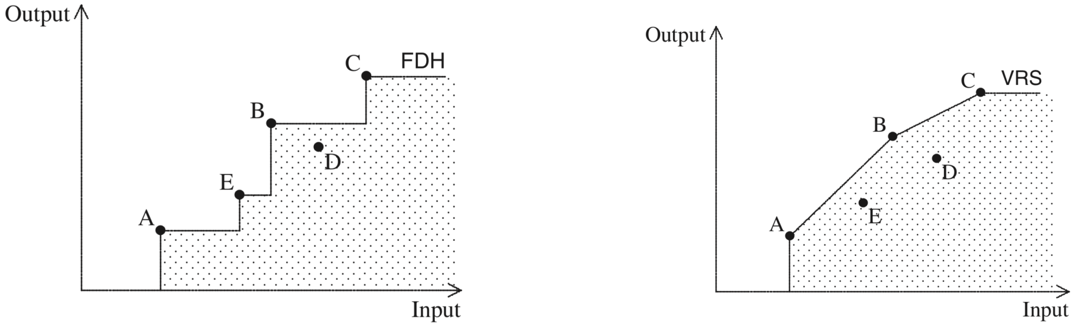


Figure 2.8: DEA technology sets under FDH and VRS assumptions (Bogetoft & Otto, 2011).

2.2.5 DEA programs.

The last basic building bloc of DEA analysis is the combination of the *Farell Efficiency* we defined in the first section and the expression of the technology we carefully built in the previous sections.

On the input side, it results in the following linear programming formulation, for firm 0 (Bogetoft & Otto, 2011);

$$\begin{aligned} \min_{E,\lambda^1,\dots,\lambda^K} \ & E \\ \text{s.t.} \ Ex_i^0 &\geq \sum_{k=1}^{K} \lambda^k x_i^k, \quad i = 1,\dots,m, \\ y_j^0 &\leq \sum_{k=1}^{K} \lambda^k y_j^k, \quad j = 1,\dots,n, \\ \lambda &\in \Lambda^K(\gamma) \end{aligned} \tag{2.6}$$

Similarly, on the output side;

$$\begin{aligned} \max_{F,\lambda^1,\dots,\lambda^K} \ & F \\ \text{s.t.} \ x_i^0 &\geq \sum_{k=1}^{K} \lambda^k x_i^k, \quad i = 1,\dots,m, \\ Fy_j^0 &\leq \sum_{k=1}^{K} \lambda^k y_j^k, \quad j = 1,\dots,n, \\ \lambda &\in \Lambda^K(\gamma) \end{aligned} \tag{2.7}$$

It is interesting to note that in the situation of constant return to scale, i.e. when $\gamma = \text{CRS}$, there is an inverse relationship between input and output efficiency: $F = \frac{1}{E}$.

2.2.6 Peer units.

Bogetoft and Otto (2011) simply defines the peer units as "the firms with positive weights in the evaluation of a given firm", i.e. the right hand-side of Equation 2.6:

$$\left(\sum_{k=1}^{K} \lambda^k x^k, \sum_{k=1}^{K} \lambda^k y^k \right)$$

It is indeed a weighted average of actual observations; the projection of the given firm on the production frontier. The fact to rely on convex combination of existing units is actually the main reason the firms complain about the improvements targets proposed by the regulator; firms do not know how to mimic the production of a convex combination.

This is therefore one of the motivations to discuss and develop alternative convexity assumptions. In the following section, we review a few of them.

2.3 Dispensing with convexity.

The choices available at hand to enlarge the technology are many, but researchers focus on new models that satisfy the free disposability assumption while relaxing the convexity assumption. We provide here a modest literature review that details several of the motivations that exist to deviate from the *statu quo*, as well as a few methods proposed by several authors. We see that the methodology that we use in this research is perfectly consistent with previous work and is still valid today.

2.3.1 The convex projection approach.

The method proposed by Bogetoft et al. (2000b) aims at finding the "smallest neoclassical production possibility set that contains a given set of observed (input-output) combinations". The authors introduce a recursive process that produces the possibility set satisfying the free disposability assumption and, more specifically to this methodology, has convex projections in the input and output space. Because we use this methodology in the Convex Pairs approach, we focus here on the motivations for such technique. They are the following:

- As already stated above, comparing actual firms with fictitious production possibilities, i.e. assuming the production possibility set to be convex, is generally not very convincing and exposes the regulator to challenges from the firms. Therefore, this method provides peers units that are either real observations or conservative estimates of such units. This careful approach used with the minimal extrapolation principle allows a more plausible benchmark.

- Natural non-convex possibility set: under the neo-classical assumption, the possibility set is convex. Here, we neglect convexity in the possibility set but assume the isoquants, i.e. "the contour lines drawn through the set of points at which the same quantity of output is produced while changing the quantities of two or more inputs" (Varian, 1992), are convex. This is a very natural assumption that can be explained by the law of diminishing marginal rates of substitution (MRS): in a 2 input context, the more you own a good, the more you are willing to give up some (Varian, c2014). The resulting indifference curve is therefore convex.
- This technique allows the relaxation of the return to scale assumption. Indeed, in previous models, the increasing return to scale assumption (IRS) was conflicting with the convexity assumption of the possibility set. Further explanations of this point are available in Bogetoft et al. (2000b).

2.3.2 Non convex input or output isoquants by Chang and Post.

The FDH model is the most general production possibility set (relying only on the free disposability assumption) but also the most consistent since any other particular assumption can be violated by the production technology. Therefore, we are interested in using the largest panel of assumption to obtain the most efficient and consistent estimator. From previous sections, we recall that convexity in the production possibility set is inconsistent with "increasing returns-to-scale, decreasing marginal rates of transformation for the outputs and decreasing marginal rates of substitution for the inputs" (Post, 2001a). Though the FDH model can deal with such issues, it is obvious that we hope to enlarge the technology set from that smallest possible set. Peterseon (1990), relayed in the article Bogetoft et al. (2000b) we reviewed in the previous section, assumes the convexity of inputs and outputs isoquants while the complete production set remains non-convex. However, this method presents some limitations. Post proposes two models to overcome such limitations: one for the input set and, separately, one for the output set.

Convex input set: again, it is justified by the law of diminishing marginal rates of substitution. If all input bundles (a bundle: input 1, input 2,... combination) of a particular set of input-output can produce at least a particular output bundle, then all convex combinations of these input bundles can produce this particular output bundle (Post, 2001a). This condition, associated with free disposability and containment ("observations in the data set are contained within the production set") gives the following Convex Input Set Approximation for the production set:

$$\begin{aligned} T^*_{CIS} = \Big\{ (x,y) \in \mathbb{R}^{p+m}_+ \Big| & x \geq \lambda^T X; y \leq (\min_{j:\lambda_j>0} y_{1j} \dots \min_{j:\lambda_j>0} y_{sj}); \\ & \lambda^T e = 1; \lambda \in \mathbb{R}^n_+ \Big\} \end{aligned} \tag{2.8}$$

Here, and in the following equations and approaches, e represents the sum operation. This equation means that "an evaluated unit can be compared with all fictitious input-output combinations, in addition to the observations".

Similarly, the model to impose convexity on the output set implies that all output bundles (a bundle: output 1, output 2,... combination) of a particular set of input-output can be produced at least by a particular input bundle, then all convex combinations of these output bundles can be produced by this particular input bundle (Post, 2001a). Again, the Convex Output Set Approximation for the production set:

$$T^*_{COS} = \left\{ (x,y) \in \mathbb{R}^{p+m}_+ \middle| x \geq (\max_{j:\lambda_j>0} x_{1j} \dots \max_{j:\lambda_j>0} x_{sj}); y \leq \lambda^T Y; \right. \\ \left. \lambda^T e = 1; \lambda \in \mathbb{R}^n_+ \right\} \tag{2.9}$$

Therefore, both for the input and output approximation, we take advantage of the convexity within the input and output sets without forcing convexity of the entire production set.

For each approximation, we can compute an efficiency estimator. The authors simply take the minimum between the input and the output estimator as the estimator when considering convexity in both the input space and the output space. The authors also acknowledge that, although statistically consistent and easier to implement, this method produces less good results than the previous one. Chang (1999) uses similar assumptions to deal with quasiconcave production frontiers.

2.3.3 Conditional convexity by Kuosmanen.

The existing observations are usually sufficient to declare a DMU efficient or not. However, finding the degree of inefficiency requires additional properties to extend the set of production possibilities. Kuosmanen, in the same way as the authors mentioned above, acknowledges that "the importance of (monotonicity and) convexity in traditional production analysis lies in their analytic convenience rather than in their economic realism" (Kuosmanen, 2001), hence he proposes to add a weaker convexity property to the model in the form of conditional convexity.

Formally, production set T is what the author calls convex conditional upon C^{EP}, if

$$\forall X, Y \text{ and } \lambda \in \mathbb{R}^n_+, e\lambda = 1 : (X,Y) \in T \wedge C^{EP} \Rightarrow (X\lambda, Y\lambda) \in T$$

where C^{EP}, the condition, is defined as

$$X\lambda x, Y\lambda \ngtr y \quad \forall (x,y) \in \mathrm{WEff}\,T$$

where WEff stands for "Weak Efficiency". What differentiates this approach from the previous ones, is that it states "that if no weak efficient production plan is dominated by a convex combination of feasible production plans contained in the production set, then the convex combination is feasible". It is also important to note that convexity implies so-called *c-convexity*, but the reverse is not true. C-convexity is then a more general property than the convexity assumption.

$$\begin{aligned} RCCMH(X,Y) = \Bigg\{(x,y) \in \mathbb{R}_+^{q+p} \Bigg| \begin{pmatrix} y \\ -x \end{pmatrix} \leq \alpha \begin{pmatrix} Y\lambda \\ -X\lambda \end{pmatrix}; \\ \alpha > 0; e\lambda = 1; \lambda \in \mathbb{R}_+^n; C^{EP} \Bigg\} \end{aligned} \tag{2.10}$$

Where RCCMH is "Ray-unbounded C-Convex Monotone Hull" and describes the minimal set containing all DMU and satisfying the c-convexity assumption, (monotonicity) and constant return to scale. The CCMH version simply ignores α.

Figure 2.9 shows the production frontier for a one-input, one-output configuration. Though the computation of efficiency scores amounts to solving a finite number of LP, for larger data set the computational time becomes heavy, hence the need to develop a more efficient algorithm.

2.3.4 Transconvex functions by Post.

Again, this methodology accounts for non-convex production technologies. Indeed, data distribution can sometimes amount to the violation of the assumptions used

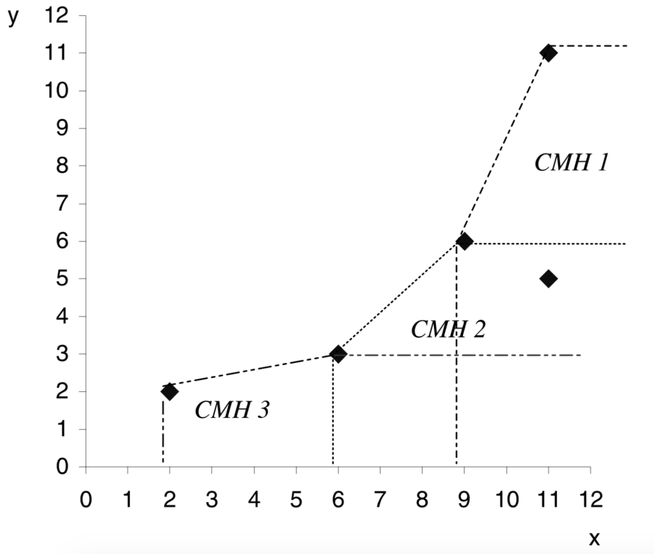


Figure 2.9: The production possibility set T under the CCMH assumption is the union of convex subsets (Kuosmanen, 2001).

in the classic model. For example, when data indicates the property of increasing return to scale, the convexity assumption of the production set is violated as it is indeed concave. This undermines the consistency of the efficiency analysis because it is based on very few observations. In addition, "inefficiency is systematically underestimated if few observations operate near the frontier" (Post, 2001b), meaning that there is no convex combination on the frontier that can serve as the reference point of the evaluated unit. This last limitation is called the small sample error.

The contribution of the author to introduce alternative production assumptions and extend the standard model is by transforming the input-output variables and replacing them in the standard model. When considering the frontier as a production function, it is assumed to be monotonically increasing and convex. However, when the frontier is not convex, like in the event of increasing return to scale, or equivalently when the production function is not concave, one can make the general assumption that it is actually concave "if expressed in terms of transformed input-output variables, or equivalently that the correspondence is concave transformable" (Post, 2001b), or transconcave. Here the definition from Post (2001b):

"f is concave transformable if there exist invertible, continuous and monotonically increasing transformation functions $H : S \to \mathbb{R}_+^{n\times m}$ and $G : I_f(S) \to \mathbb{R}_+^n$ such that $G(f(H^{-1}))$ is concave on $H(S)$. Alternatively, if $G^{-1}(\theta^T G(f(X))) \leq f(H^{-1}(\theta^T H(X)))$
$\forall X \in S, \theta^T e = 1, \theta \geq 0$.

In the standard Banker et al. (1984) model, the output inefficiency is measured by

$$\hat{\varepsilon}_{BCC,0} = \max_{\lambda_0} \Big\{ \lambda_0^T y - y_0 \lambda_0^T e = 1; \\ \lambda_0^T X \leq x_0; \lambda_0 \geq 0 \Big\}$$

.

When transconcavity is introduced, the original variables are replaced by the new inputs-outputs and the benchmarking model becomes

$$\hat{\varepsilon}_{TDEA,0} = \max_{\lambda_0} \Big\{ \lambda_0^T G(y) - G(y_0) \lambda_0^T e = 1; \\ \lambda_0^T H(X)(x_0); \lambda_0 \geq 0 \Big\}$$

The implementation of such method can be found in Post (2001b). The final frontier shown in Figure 3.1 expresses the difficulty to implement the methodology due to "the practically infinite number of alternative range and domain transformations available". Therefore, the author focuses on the exponential domain transformation, the motivation being that "BCC and FDH models represent the limiting cases of the exponential model".

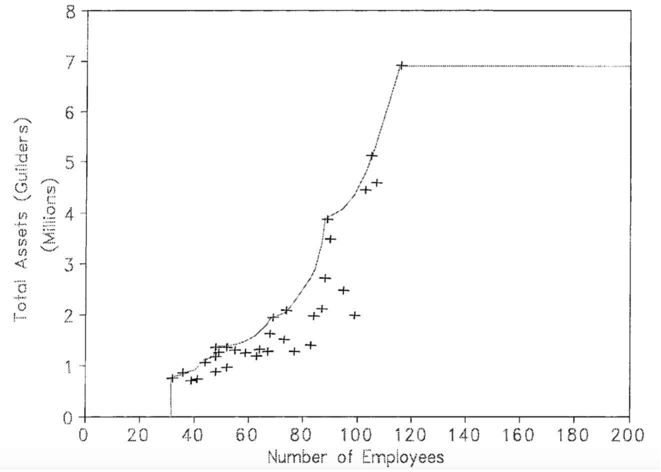


Figure 2.10: Final exponential TDEA frontier for the banking industry (Post, 2001b).

2.4 The convex pairs approach.

This approach provides an interesting exploration of the possibilities at hand when considering different convexity assumptions. In the previous chapter, we already enhanced some of the many discussions various authors propose on the question, and the relevance to investigate new efficiency analysis possibilities, as much for regulators as for firms.

At its core, this approach allows different degrees of convexity in the overall production possibility set T and defines a convex pair, for each DMU, as a "a convex input possibility set and a convex output possibility set such that all input vectors in the inputs set can produce all vectors in the output sets" (Agrell et al., 2005a). The reason for allowing different degrees of convexity is that it takes into consideration the possibility of economies of scale and scope that was neglected by the conventional approach.

2.4.1 Convex pair and set of convex pairs.

We model the full production set as the union of convex pairs defined below. While local convexity within the input and output sets is possible, we dispense with global convexity.

We begin with a formal definition and the first assumptions of the approach. We consider the set $\mathcal{N}$. It is the most aggregated level and contains I subsets, the second level of aggregation that represents the DMU. At this level, and it is the particularity of this approach, convexity is neglected. Each DMU subset contains itself the pair of subset L and P which contains the m and n input and output vectors, respectively. At this third and lowest level in terms of aggregation, convexity is possible and each component can be represented as a polyhedron.

Formally, for each $i \in \mathcal{N}$ let L_i be a set of m dimensional input vectors x and let P_i be a set of n dimensional output vectors y.

Assumption 1.

For the inputs we assume that L_i is nonnegative and satisfies the free disposability assumption in the input space. Moreover, L_i is assumed to be convex. Formally,

$$(L_i + \mathbb{R}^m_+) \cap \mathbb{R}^m_+ = L_i \tag{2.11}$$

For the output we assume that P_i is nonnegative and satisfies the free disposability assumption in the input space. Moreover, P_i is assumed to be convex. Formally,

$$(P_i - \mathbb{R}^n_+) \cap \mathbb{R}^n_+ = P_i \tag{2.12}$$

Assumption 2.

The input and output sets are associated as a pair (L_i, P_i) by the assumption that any input vector $x \in L_i$ can produce any output vector $y \in P_i$.

Finally, consider the full technology T i.e. $T = \Big\{(x, y) \in \mathbb{R}_+^{m+n} \Big| \text{ x can produce y}\Big\}$. T is the union of the feasible pairs, formally

$$T = \cup_{i \in \mathcal{N}}(L_i, P_i) \tag{2.13}$$

We note already the difference with the previous definition of the technology, as here we acknowledge the possibility of local convexity within the input and output subsets, but we dispense with the convexity of the full production set T that was assumption 2 in the previous section.

2.4.2 Natural convexity in the input and output space: the convex projections.

We rely on the convex projections approach proposed in Section 3.2.1 by Bogetoft et al. (2000b) to build the technology set that respects the properties required by the model stated above. This methods allow to enrich the existing technology in a flexible way.

We consider a set I of DMU with x_i the non-negative input vector and y_i the non-negative output vector, for $i \in I$. From Assumption 1 and 2 above, we know any input vector can produce any output vector. We need now to find the smallest T such that the input projections $L(y) = \{x|(x, y) \in T\}$ and the output projections $P(x) = \{y|(x, y) \in T\}$ are convex for all $(x, y) \in T$. In a simple way, for a given output vector, we want to find all the convex combinations of the input vectors that can be used to produce this output, and vice-versa for any input considered.

For arbitrary indices $i, j \in I$, i.e. for two DMU i and j, we define then two pairs of sets:

$$(L_a, P_a) = (L_i \cap L_j, P_i \uplus P_j) \tag{2.14}$$

and

$$(L_b, P_b) = (L_i \uplus L_j, P_i \cap P_j) \tag{2.15}$$

where $\uplus$ represents the convex union operation. Equ 2.14 means that from the smallest convex set containing output of i and j, we build the input set using projections in the input space. Equ 2.15 follows the same process but starts with the input set. Any x, y belonging to the pair (L,P) a or b are feasible and convex because *i.* any $x \in L_i \cap L_j$ and $y \in P_i \cup P_j$ are a feasible input-output due to free disposability and *ii.* the output projection P is required to be convex.

We use the concept of *dominance* detailed above to remove any DMO that would be dominated by another, i.e. when $L_i \subseteq L_j$ and $P_i \subseteq P_j$ for some $i, j \in I$. If it is true for all combinations for i and j, then the DMU i (artificial or not) can be removed.

We use this concept of convex projection when enriching the technology based on the existing data. The procedure is detailed in Section 3.2.1.

In the next chapter, we will cover the algorithm required to construct the enriched technology, and we will see that linear programming is still valid to solve $|N|$ optimisation problems, one for each of the pairs of sets.

2.4.3 Polarity: blocking and antiblocking theory.

Agrell et al. (2005a) proposes two alternatives to describe the input and output sets. Those sets can be called convex polytopes, i.e. a bounded polyhedron that is the convex hull of a finite number of vertices. This is called the $\mathcal{V}$-representation. Alternatively, the $\mathcal{H}$-representation is the intersection of a finite set of closed half-spaces (Fukuda, Liebling, & Lütolf, 2001). What is interesting when doing operations on polyhedra, is to understand the duality of polyhedra.

"Two d-polytopes, P and $\mathcal{P}^*$ are said to be dual to each other provided there exists a 1-1 mapping between the set F of all faces of P, and the set F^* of all faces of $\mathcal{P}^*$, such that the mapping is inclusion-reversing" (Wilde, 2000).

Indeed, the $\mathcal{V}$-representation is relevant when computing the convex union of the polytopes, when the $\mathcal{H}$-representation, or dual formulation, is better when computing the intersection since "the intersection is defined by a set of constraints" (Agrell et al., 2005a).

Both blockers and antiblockers are the dual representation of those polyhedra that are the inputs and output sets. They are based on the theory of polarity, see Wilde (2000) for more details, that is slightly amended by the addition of the non-negativity assumption.

Antiblockers.

The antiblocker of the output set P of dimension p is defined as:

$$\mathcal{A}(P) = \{y^* \in R^p_+ | y^* y \leq 1 \forall y \in P\} \tag{2.16}$$

Blockers.

Similarly to the antiblocker, the blocker of the input set L of dimension r is defined as:

$$\mathcal{B}(L) = \{x^* \in R^r | x^* x \geq 1 \forall x \in L\} \tag{2.17}$$

The various properties and their proof can be found in Agrell et al., and we will use one of the main implications of these properties in the algorithm to build the convex pairs.

Chapter 3

Model specifications

In this chapter, we seek to develop a model based on the theory proposed above on the Convex Pairs approach. In the meantime, we should also develop a methodology to assess the consistency of this method to estimate the production frontier.

In this purpose, in addition to the estimation of efficiencies using the Convex Pairs approach, we will develop the *Free Disposability Hull* efficiency and the efficiency under the *Varying Return to Scale* assumption because it dropped the convexity assumption for the former and is not violated in case of non-convexity for the latter. *Free Disposability* is imposed for the three models because it is a strong assumption. Because the choice of assumptions has a crucial impact on the results of the analysis and consequently on the improvement effort demanded to the regulated firm, those assumptions should be tested. R. Banker (1996) proposes an unified framework of the statistics used to test the hypotheses chosen.

3.1 The control group: the standard approach.

3.1.1 FDH efficicency calculation.

We recall Assumption 1 in Section 2.2.3 to build the *Free Disposability Hull.* With Figure 2.2.3, we recall that FDH is the less restrictive model because it drops the convexity assumption. The method, proposed by Deprins, Simar, and Tulkens (2006), "characterises Y with no other assumption than input and output disposability". We simply measure the horizontal distance between an observed unit and the frontier to assess the input-efficiency of the point. Indeed, for a certain level of output, it is easy, and consistent, to ask a non-efficient firm to use less inputs in its production plan, as does an efficient firm on the frontier. The input-oriented efficiency is the following:

$$\begin{aligned}
&\min_{E,\lambda^1,\ldots,\lambda^K} E \\
&\text{s.t. } Ex_i^o \geq \sum_{k=1}^{K} \lambda^k x_i^k, && i = 1,\ldots,m, \\
&\qquad y_j^o \leq \sum_{k=1}^{K} \lambda^k y_j^k, && j = 1,\ldots,n, \\
&\qquad \lambda \in \Lambda^K(\gamma)
\end{aligned} \tag{3.1}$$

Where

$$\Lambda^K(fdh) = \left\{ \lambda \in \mathbb{N}_+^K \middle| \sum_{k=1}^{K} \lambda^k = 1 \right\}$$

To compute efficiency, the model solves m mixed integer programming problems (MIP) with integer λ variables. However, FDH problems are usually not solved using MIP problems because they can be rewritten as a series of minimax problems (Bogetoft et al., 2000b); here the input efficiency we will use:

$$E^o(fdh) = \min_{k:y^k \geq y^o} \max_{i=1,\ldots,m} \frac{x_i^k}{x_i^o} \tag{3.2}$$

In the outer optimisation; i.e. the min, we look for a peer unit k for evaluated firm o that has the smallest output greater than the output of firm o. In the inner optimisation, we find the input $i = 1, \dots, m$ that "determines the largest proportional reduction that we can make to all inputs at the same time" (Bogetoft et al., 2000b) for firm o.

It is important to note that Banker (1993) proved that the model is statistically consistent for monotonically increasing concave (equivalent to convex possibility set) functions. This is the precise reason we build an alternative model; this last condition being not automatically respected.

3.1.2 VRS efficiency calculation.

The model to compute VRS efficiency is similar. The sum of all λ is still equal to 1, but its domain is now restricted to real numbers (FDH is extended to all natural numbers), therefore introducing convexity. Again, we chose to develop the VRS efficiency model because the frontier is closer to the observations and it can still produce consistent results in the absence of global convexity.

The input efficiency:

$$\begin{aligned} &\min_{E,^1,\dots,\lambda^K} E \\ &\text{s.t. } Ex_i^o \geq \sum_{k=1}^{K} \lambda^k x_i^k, && i = 1, \dots, m, \\ &\qquad y_j^o \leq \sum_{k=1}^{K} \lambda^k y_j^k, && j = 1, \dots, n, \\ &\qquad \lambda \in \Lambda^K(\gamma) \end{aligned} \tag{3.3}$$

Where

$$\Lambda^K(vrs) = \left\{ \lambda \in \mathbb{R}_+^K \middle| \sum_{k=1}^{K} \lambda^k = 1 \right\}$$

In this situation, the model solves as many linear programming problems as there are firms.

The two DEA programs developed here will serve as a comparison with the new program proposed by Agrell et al. (2005a) and detailed below.

3.2 The Convex Pairs approach.

3.2.1 Build the technology based on the convex pairs.

Based on Bogetoft et al. (2000b), Agrell et al. (2005a) enlarges the technology from existing data using the convex pairs approach for the reasons we state in Section . In the algorithm quoted below, we create the new pairs such as in Equation 2.14 and 2.15. This procedure uses all the building blocks of the Convex Pairs approach we define above. It continues incrementally and stops when there is no possibility of creating any new non-dominated pair.

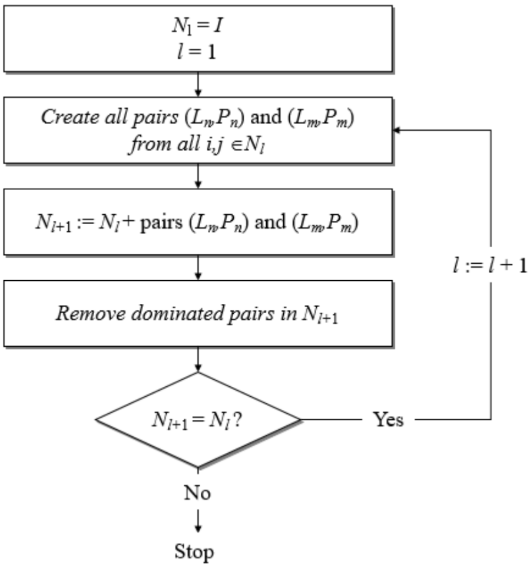


Figure 3.1: Algorithm design (Agrell et al., 2005a).

At the termination of the algorithm presented above, two operations will have been done, i.e. the convex union $\uplus$ and the intersection $\cup$. We use the half space representation proposed above, as well as the blocking and antiblocking theory to conduct those operations.

For the outputs:

For the 2 polyhedral output spaces O and P, $\mathcal{A}$ is the antiblocker as defined above:

$$\begin{aligned} P \cap O &= \mathcal{A}(\mathcal{A}(O) \uplus \mathcal{A}(O)) \\ P \uplus O &= \mathcal{A}(\mathcal{A}(O) \cap \mathcal{A}(O)) \end{aligned} \tag{3.4}$$

For the inputs:

For the 2 polyhedral input spaces L and K, $\mathcal{B}$ is the blocker as defined above:

$$\begin{aligned} L \cap K &= \mathcal{B}(\mathcal{B}(L) \uplus \mathcal{B}(K)) \\ L \uplus K &= \mathcal{B}(\mathcal{B}(L) \cap \mathcal{B}(K)) \end{aligned} \tag{3.5}$$

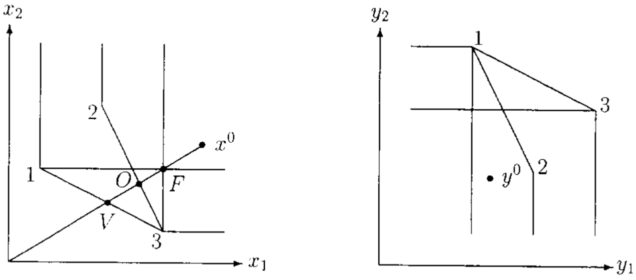
In practice, we will simply take the minimum input value from L and K as the union, and the larger value as the intersection since the input space is 1-dimensional in our application.

The proof for those propositions can be found in Agrell et al. (2005a).

3.2.2 Efficiency index.

With this procedure, we built an approximation of the true technology that we expect to lie between the FDH, i.e. the no-convexity situation, and the VRS frontier which assumes global convexity. On the figure below, the input efficiency of x^0 is found by comparing it with the points F, O and V for the FDH, Convex Pairs and VRS approach, respectively.

We look for the input-oriented efficiency, since the DSOs will *a priori* keep their outputs constant, i.e. the service they provide, and reduce their cost. To find the productivity index for our original pairs, we run the following linear programming



(a) Input Space. (b) Output Space.

Figure 3.2: Example of the representation of the frontier after the last iteration (Bogetoft et al., 2000a).

problem (Agrell et al., 2005a):

$$\begin{aligned} &\min_{\theta_i, z_i} \sum_{i \in \mathcal{N}} \theta_i \\ &\text{s.t. } B_i x_0 \theta_i - z_i \mathbf{1} \geq 0 \qquad \forall i \in \mathcal{N} \\ &\qquad (A_i y_0 - \mathbf{1}) z_i \leq 0 \qquad \forall i \in \mathcal{N} \\ &\qquad \sum_{i \in \mathcal{N}} z_i = 1 \\ &\qquad z_i \geq 0 \end{aligned} \tag{3.6}$$

Basically, the second constraint verifies if the polyhedron considered, i.e. the output space y_0, is included within the polyhedron representing the virtual pair we compare y_0 to. If true, the left hand side of this constraint would be indeed positive. Among the spaces i selected in this manner, z_j would be set to 1 for the virtual pair that would yield the lowest θ in the first constraint, putting z_i for all the other constraints, i.e. $i \in \mathcal{N} \setminus j$ to 0.

Chapter 4

Data and implementation

4.1 Data and variables.

For the application of the methodology developed in the previous section, we will use data provided by the Energy Market Inspectorate (Swedish regulator) and used by Agrell and Niknazar (2014) to discuss outliers detection and robustness in DEA modelling, calculation and interpretation. From the 245 observations in the original data set, the panel post treatment contains 207 different DSO, hence excluding potential negative, zero and missing data, as well as outliers. To cope with possible imbalance in data magnitude, we simply mean normalised the data.

The Swedish distribution market was unbundled and deregulated in 1996, and shows a large number of firms, particularly useful for DEA analysis. Variable selection is crucial in benchmarking because efficiencies can substantially vary depending on the input and output chosen. Here, we use the technical efficiency variables proposed by (Agrell & Niknazar, 2014).

4.1.1 Missing, negative and zero values.

Data Envelopment Analysis is a non parametric method, because the entire data is used to determine a posteriori the parameters of the production function. It is a deterministic method, which means that any variation in the data is supposed to contain valuable information about the efficiency of the firms. In such regard, DEA models do not allow any noise in the data, and we must ensure this when preparing the data set for analysis. In addition, all numbers should preferably be strictly positive, i.e. no zero values, defined as the "positivity requirement" of DEA. Though models relaxing this condition, as well as some scaling or translating techniques exist, we prefer to simply remove those data to ensure minimal transformation of the data. Furthermore, this only removes 9 firms from our data set, leaving a relevant number of firms for analysis.

4.1.2 Outliers removal.

Regulators need certainty in their analysis to prevent the risk that evaluated firms appeal their attributed efficiency score and target. Because DEA models are highly sensitive to extreme data, a lack of robustness of the frontier can cause serious scepticism about the frontier-based regulation (Agrell & Niknazar, 2014). Wilson (1993) simply defines outliers as atypical observations. Heterogeneity can come from a structural perspective, i.e. differences in the use of technology, scale and scope. These incomparabilities are inevitable when performing a benchmarking analysis. Given our focus on mathematical modelling, the behavioural perspective (*collusion* and *maverick reporting*) of heterogeneity is not covered here. More detailed information on the source of heterogeneity, as well as a comparison of 3 different outliers detection methods can be found in Agrell and Niknazar (2014). The methodology proposed by Wilson (1993) suits best our statistical rather than applied research because it is based on the volume of the data cloud. Though upfront removal based on size is not advised, we still prefer this method because it allows to remove outliers prior to computing the efficiency score.

From Agrell and Niknazar (2014), we know that Wilson (1993) defines the volume of the data cloud as the determinant of the corresponding input-output matrix, $[x, y]$. When removing firm i, we look at the ratio between the volume of the data cloud without firm i (D^i) and the volume of the full data cloud (D). A smaller ratio

$$R^i = \frac{D^i}{D}$$

shows the greater importance that firm i only had in the entire data set, relatively to the other firms. Firm i is then defined as an outlier, i.e. such that when it is removed from the data cloud (D), $R^i < \eta$ for a predefined threshold level $\eta > 0$. We have chosen the package Frontier Efficiency Analysis with R, FEAR (Wilson, 2008) to conduct this outliers analysis because it is much faster than the function proposed in the Benchmarking package. We arbitrarily chose to remove up to $p = 12$ firms, and the method looks for the smallest R value for each number of firms deleted to plot

$$\left(p, log\left(\frac{R^p}{R^p_{min}}\right)\right)$$

as shown in Figure 4.1.

The graph peaks when 2 outliers are removed and we identify them as DMUs number 88 and 146 from Table 4.1.

Our data set now consists of $n = 209 - 2 = 207$ observations and the descriptive statistics can be seen in Table 4.2. The outliers removal has been effective as standard deviation has been divided by up to ten depending on the variable.

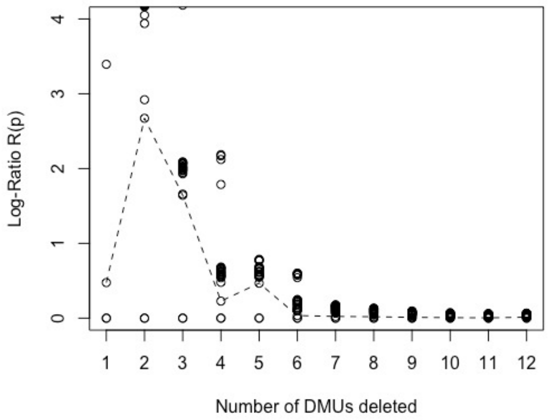


Figure 4.1: Log-ratio plot for outliers detection.

p	Deleted observations												R^p_{min}
1	88												7.9e-03
2	146	88											1.0e-04
3	144	146	88										1.1e-05
4	54	144	146	88									5.2e-06
5	100	54	144	146	88								2.4e-06
6	111	100	54	144	146	88							1.7e-06
7	58	111	100	54	144	146	88						1.2e-06
8	58	111	5	100	54	144	146	88					9.3e-07
9	50	58	111	5	100	54	144	146	88				7.1e-07
10	50	58	174	111	5	100	54	144	146	88			5.4e-07
11	189	50	58	162	111	5	100	54	144	146	88		4.1e-07
12	189	50	58	174	162	111	5	100	54	144	146	88	3.1e-07

Table 4.1: p removed observation with the corresponding minimum determinant ratio.

4.1.3 Mean normalisation.

We follow the advice of Sarkis (2006) and we reduce potential imbalances in data magnitude using this simple technique. Indeed, though it does not seem that it has an impact on our calculations, it may cause problem in the execution of the software. This operation is done in two steps, i.e. finding the mean of each input and output column, and dividing each value by the mean. Therefore, every value is represented by a percentage of the mean.

$$VNorm_n = 1000\frac{V_n}{\bar{V_i}} \qquad \text{where} \qquad \bar{V_i} = \frac{\sum_{n=1}^{N} V_n i}{N}$$

We multiply this value by 1000 and round it to obtain integer values with sufficient differentiation between the observations.

4.1.4 Inputs.

The inputs are resources controllable by the company. We consider here the short term operating expenditures, which are typically used in regulation to incentive short term improvements. For long term applications, which we do not cover here, technical efficiency is used and two more inputs, namely network capital and total energy losses, are added.

Input 1: **Operating expenditures short run (OPEX.sr)**: today, the main element for determining the revenue cap is the TOTEX (divided into CAPEX, Non-controllable OPEX and Controllable OPEX), complementary to incentives for good quality of supply (Council of European Energy Regulators, 2019).

4.1.5 Outputs.

The outputs can be seen as derived from the demand and are therefore out of control of the utility companies (Jamasb & Pollitt, 2003).

Output 1: **Energy delivered low-voltage (DEL.lv)**: energy delivered at low voltage, i.e. for domestic and light commercial use.

Output 2: **Energy delivered high-voltage (DEL.hv)**: is the energy transported between low-voltage networks and to heavy industry.

Output 3: **Total connections low-voltage (CONX.lv)**: is the number of connected clients to the low-voltage grid.

Output 4: **Total connections high-voltage (CONX.hv)**: is the number of connected clients to the high-voltage grid.

Output 5: **Total system peak load (PEAK)**: peak in demand that excess the planned delivery.

Variable	Code	Units	Min	Average	Max	Sd
OPEX SR	OPEX.sr	SEK	1502.11	32837.15	523150.64	53533.78
Energy Delivered LV	DEL.lv	MWh	10	281855.88	4320286	481032.7
Energy Delivered HV	DEL.hv	MWh	1	141210.61	4203252	381524.91
Total connections LV	CONX.lv	#	768	22151.51	451398	43859.34
Total connections HV	CONX.hv	#	1	27.54	453	48.28
System peak load	PEAK	MW	2	99.21	3360	264.75

Table 4.2: Descriptive statistics, electricity DSO, Sweden.
Reference year: 2000, n=207.

4.2 The classic approach: implementation and results.

After the implementation of our algorithm, we acknowledged that the computation time was reasonably high for a normal personal computer, and obviously increasing at each iteration of the algorithm in Section 3.2.1. We decided to run it on our personal machine, so we reduced the data set to 40. However, we kept all the one input variable and five output variables, and it is still representative of the rest of the data set and relevant since we could reach the termination condition of the algorithm, i.e. create pairs until no additional pair can be non-dominated.

4.2.1 FDH.

Unsurprisingly, many of the units are deemed efficient under this method, i.e. 22 out of 40. The estimated production function sticks to all dominating units, which are many because it does not rely on composite peers to evaluate efficiency. This method is the simplest and the most restricted method to compute efficiency.

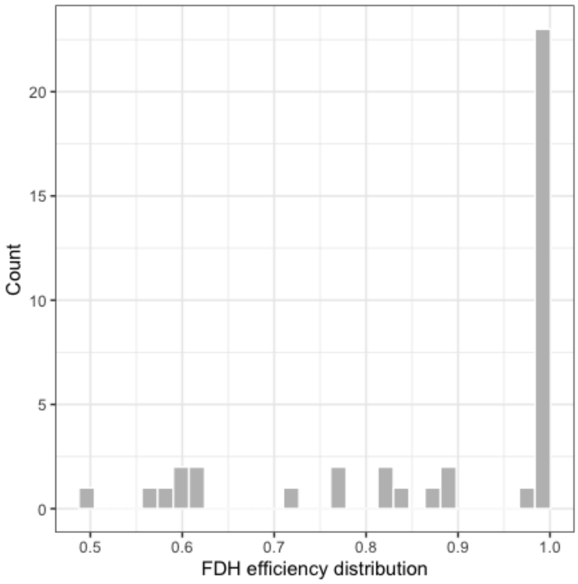


Figure 4.2: Frequency histogram of FDH efficiencies.

However it is very useful and still relevant in an applied perspective if we cannot justify the convexity assumption, such as it may be the case in the real world.

4.2.2 VRS.

"In the short run model, VRS is assumed to be the most relevant specification because DSOs have limited opportunities to reorganise in the short run" (Bogetoft & Otto, 2011).

Here, only 13 DSO are deemed efficient by the model. Though this figure is much lower than for the FDH model, the VRS estimated technology is the second largest, see Figure 2.7.

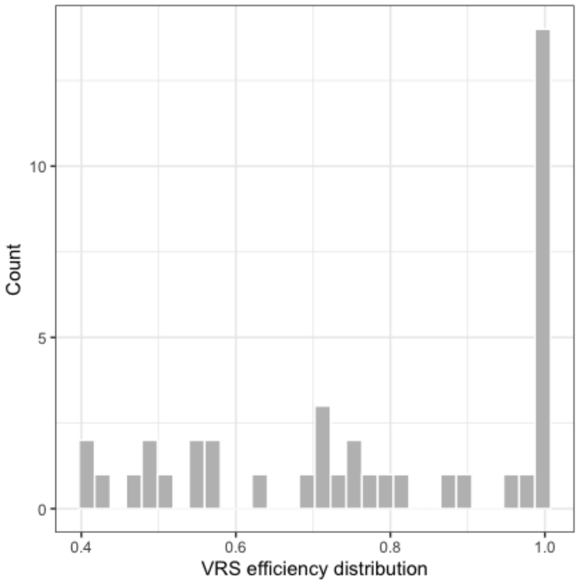


Figure 4.3: Frequency histogram of VRS efficiencies.

4.3 Convex Pairs approach: implementation and results.

The algorithm used for computing the convex pairs relies on the polyhedron, polar and blocking theory developed in the first chapters of this work.

To compute efficiencies using the new method, we implemented the algorithm proposed in Section 3.2.1 in R. The code can be found in Appendix 1. Though costly in computational time, it can still be run on a regular computer for a limited set of observations and variables.

Practically, we first conducted a dominance test on the original data set to remove any dominated pairs. Indeed, there is no utility to form pairs with such observations, since it would be immediately rejected by the dominance filter. From our 40 original data, 16 are already removed at this step.

For the building part itself, we run the program as long as there are observations that are removed at the end of an iteration. It works with two indexes. One contains the IDs of the original observations at the beginning of the iteration, while the second contains the original IDs, as well as those of all the non-dominated pairs that are created along the procedure. This allows, during one iteration, to create new pairs that are the combination of the original observations, and to verify dominance with all the existing pairs, including those created just before. The dominance check here is two-side, i.e. when creating a new pair, it is excluded as soon as it is dominated by an existing one, but the IDs of the pairs that are subsequently dominated are also stored for deleting at the end of the iteration. This is thought as the most efficient accumulation method.

The operations functions, all different for input, output, intersection and convex union, are based on the polytope theory developed above. Though it calls the exact same concepts as the blocking theory and its inequalities, the vertices enumeration and other $\mathcal{V}$- and $\mathcal{H}$-representation from the R package scdd are better for a visual representation of the polytopes. The operations on inputs, however, are very simple since there is only one input variable. A similar function than the one used for the outputs can be implemented for the input polyhedron, provided that assumptions are made to close the polyhedron, which enables vertices enumeration.

The dominance function also calls these space representation of the problem, and, for instance, there is dominance in the output space when the vertices of a specific polytope are included, or "made redundant", within another. If there is no domination because the polytopes are identical, the newest is removed. Finally, if both the polytopes have one of their vertices included in the other, there is no domination in the output space.

The productivity index is computed by implementing the problem proposed by Agrell et al. (2005a) using the Rglpk package, a classic linear programming problem solver.

Figure 4.4 shows the distribution of the Convex Pairs approach efficiency. 13 are deemed efficient.

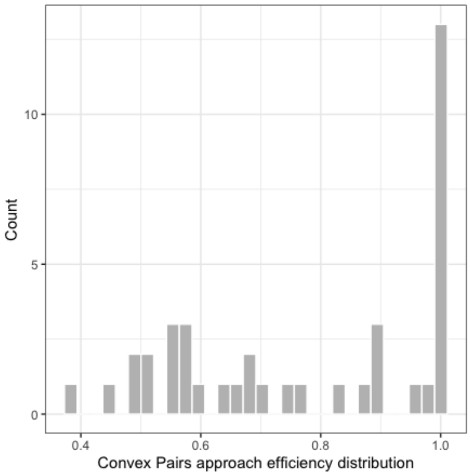


Figure 4.4: Frequency histogram of the Convex Pairs approach efficiency.

Chapter 5

Comparison, limits and conclusion

5.1 Illustrative example.

To better understand the behaviour of the code and the potential reasons for difference between the methodologies, we will take three pairs as illustrative examples. The first and second are respectively non-efficient and efficient in all the 3 models. For the last example, we will focus on a pair that is efficient under VRS and FDH, but that is not in our model.

5.1.1 Example 1: pair 23.

The efficiencies are respectively 0.61, 0.58 and 0.48 for FDH, CP and VRS.

	#	inputs	outputs				
Pair	23	493	227	379	264	453	309
VRS	5	1081	715	5562	835	1008	1352
	6	150	143	110	147	101	124
	33	578	990	1472	826	1058	1082
	37	302	456	438	352	1108	448
FDH	37	302	456	438	352	1108	448
CP	66	286	328	347	345	453	386
			329	98	336	302	311

We see here that the pair 23 is compared in our model to the pair 66, a combination of 2 original pairs.

5.1.2 Example 2: pair 21.

The efficiencies are 1 for FDH, CP and VRS.

	#	inputs	outputs				
Pair	21	1098	1738	2223	1713	2418	1746
VRS	21	1098	1738	2223	1713	2418	1746
FDH	21	1098	1738	2223	1713	2418	1746
CP	59	1098	1738	2223	1713	2418	1746
			715	5562	835	1008	1352

For the CP efficiency, the pair 21 is compared to the new pair 1098, which is the combination of itself and another one.

5.1.3 Example 3: pair 39.

The efficiencies are respectively 1, 0.83 and 1 for FDH, CP and VRS.

	#	inputs	outputs				
Pair	39	1997	3192	904	2674	1209	2735
VRS	39	1997	3192	904	2674	1209	2735
FDH	39	1997	3192	904	2674	1209	2735
CP	102	1654	2343	3724	2431	2166	2658
			1548	1820	1879	2620	1344

Here, the difference in efficiency may be explained by the very large size of pair 39 in comparison to the rest of the data.

5.2 Comparison between the methodologies.

In this section, we will compare the results obtained using the different methodologies, and attempt to draw a conclusion about whether the new approach can be used as a substitute to the methodology under the VRS assumption with the advantage of neglecting the global convexity assumption. The choice of assumption should indeed not be taken lightly, since it can have a great impact on the results and an incorrect specification could "produce biased efficiency scores" (Niknazar

& Bourgault, 2018).

The expectation from the new method is that the results are sufficiently close to the VRS methodology such that they can be assumed relatively identical, and used equivalently depending on the requirement in global convexity.

Comparing the results of these three models reveals how the convexity axiom affects the efficiency score. However, we try to find here the extend of the difference between the new method and the others, with the clear aim of taking the best of both worlds.

First, we observe in Table 5.4 that the majority of the DMUs are deemed inefficient in all the 3 models, while some are efficient in 1, 2 or 3 models.

n=0	n=1	n=2	n=3
1	21	22	23
2	34	26	25
4	35	28	32
5	39	30	33
6	42	31	36
7		44	37
8		45	38
9		46	40
10			43
11			
12			
13			
15			
16			
17			
18			
19			
20			

In Figure 5.1, we can observe that most of the results computed using the convex pairs approach are as expected bounded between the VRS and the FDH method. The horizontal lines represent the median value for each method.

There are some exception where our score is below the VRS score. This can be

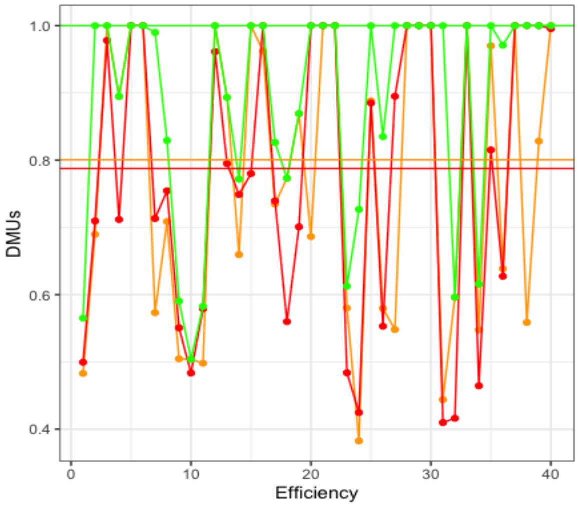


Figure 5.1: VRS (red), Convex Pairs approach (orange) and FDH (green) efficiencies.

explained by a remaining heterogeneity in the data, despite our data cleaning and outliers removal. Indeed, Figure 5.2 shows that the CP efficiency is below VRS efficiency for DMUs that are large relatively to the rest of the data.

5.3 Conclusion and limitations.

In this thesis, we were interested in an alternative method to represent technology in DEA analysis. We began by reviewing the different components of benchmarking, and in particular the convexity assumption, which has a significant impact on the efficiency score finally assigned to the firm studied. We have also discussed the motivations inherent to this questioning of the convexity assumption, and detailed some of the different studies that already exist on this topic.

On this basis, we were able to develop the hypotheses of a new approach proposed by Agrell et al. (2005a), which is mainly based on the rejection of the global convexity assumption (at the technology level), while assuming that it does exist in the space of inputs and outputs.

Although exploratory in nature, we have applied the new approach to the actual

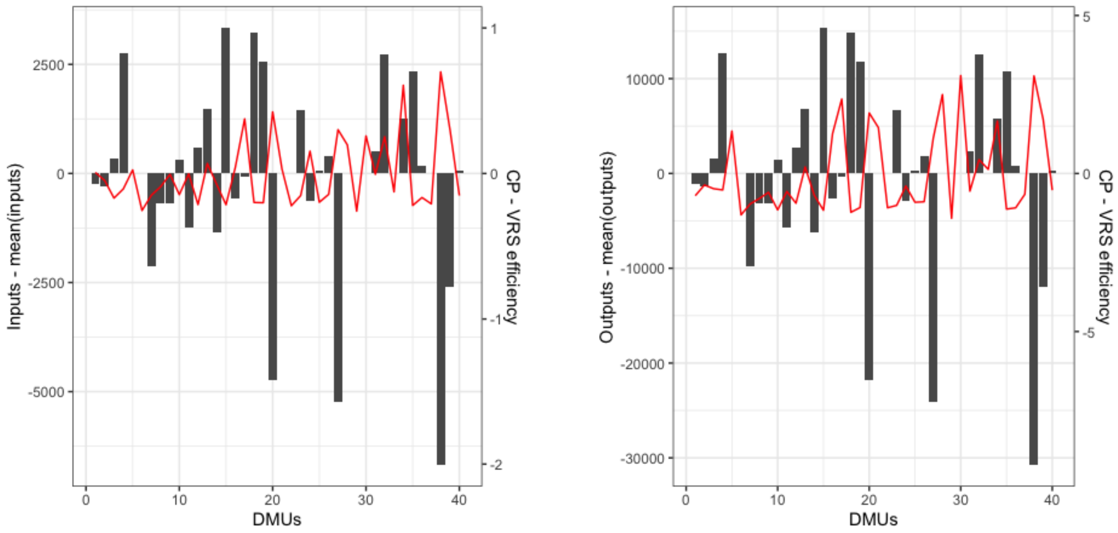


Figure 5.2: Comparison between relative size of the DMU (red) and difference VRS-CP efficiency (bars).

data of Swedish electricity distributors of the year 2000 because they constitute a large number of firms, which is what our analysis requires, and are subject to the same regulatory regime, which allows the data to be consistent across the firms.

We used and implemented the model to construct the Convex Pairs based on the convex union and intersection of the extreme points of the existing pairs, and called the concept of dominance to remove pairs already included in others. We then calculated an efficiency index for each of the original pair.

We conclude that the difference between the results is on average too large to assume that the VRS and the convex pairs approach can be used equivalently to circumvent the convexity assumption.

Further comparison could be done using Monte-Carlo simulations to better understand the impact of the methodological choices on the efficiency measure. In this thesis, the comparison is indeed limited to the observation of the difference between the methods.

References

Agrell, P., & Bogetoft, P. (2007, 01). Development of benchmarking models for german electricity and gas distribution.

Agrell, P., & Bogetoft, P. (2016, 07). Regulatory benchmarking: Models, analyses and applications. *Data Envelopment Analysis Journal*, *3*, 49-91. doi: 10.1561/103.00000017

Agrell, P., Bogetoft, P., Brock, M., & Tind, J. (2005a). Efficiency evaluation with convex pairs. *Advanced Modelling and Optimization*, *7*(2), 211–237.

Agrell, P., Bogetoft, P., Brock, M., & Tind, J. (2005b). *Efficiency evaluation with convex pairs* (CORE Discussion Papers RP No. 1828). Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Agrell, P., & Niknazar, P. (2014). Structural and behavioral robustness in applied best-practice regulation. *Socio-Economic Planning Sciences*, *48*(1), 89 - 103. (Data Envelopment Analysis in the public sector) doi: https://doi.org/10.1016/j.seps.2013.12.004

Banker, R. (1996, Jul 01). Hypothesis tests using data envelopment analysis. *Journal of Productivity Analysis*, *7*(2), 139–159. doi: 10.1007/BF00157038

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*(9), 1078-1092. Retrieved from `https://doi.org/10.1287/mnsc.30.9.1078` doi:

10.1287/mnsc.30.9.1078

Bogetoft, P. (1996, 03). Dea on relaxed convexity assumptions. *Management Science*, *42*(3), 457.

Bogetoft, P., & Otto, L. (2011). *Benchmarking with dea, sfa and r* (Vol. 157). Springer. doi: 10.1007/978-1-4419-7961-2

Bogetoft, P., Tama, J. M., & Tind, J. (2000a). Convex input and output projections of nonconvex production possibility sets. *Management Science*, *46*(6), 858–869. Retrieved from `http://www.jstor.org/stable/2661490`

Bogetoft, P., Tama, J. M., & Tind, J. (2000b, 06). Convex input and output projections of nonconvex production possibility sets. *Management Science*, *46*(6), 858-869.

Chang, K.-P. (1999). Measuring efficiency with quasiconcave production frontiers. *European Journal of Operational Research*, *115*(3), 497 - 506. doi: https://doi.org/10.1016/S0377-2217(98)00043-5

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429 - 444. Retrieved from `http://www.sciencedirect.com/science/article/pii/0377221778901388` doi: https://doi.org/10.1016/0377-2217(78)90138-8

Council of European Energy Regulators. (2019). *Report on regulatory frameworks for european energy networks* (Tech. Rep.). Author.

Deprins, D., Simar, L., & Tulkens, H. (1984). *Measuring labor-efficiency in post offices* (CORE Discussion Papers RP No. 571). Université catholique de Louvain, Center for Operations Research and Econometrics (CORE). Retrieved from `https://EconPapers.repec.org/RePEc:cor:louvrp:571`

Deprins, D., Simar, L., & Tulkens, H. (2006). Measuring labor-efficiency in post offices. In P. Chander, J. Drèze, C. K. Lovell, & J. Mintz (Eds.), *Public goods,* *environmental externalities and fiscal competition* (pp. 285–309). Boston, MA: Springer US. doi: 10.1007/978-0-387-25534-7_16

Encyclopædia Britannica. (n.d.-a). Determinant. Retrieved 11-07-2019, from `https://www.britannica.com/science/determinant-mathematics`

Encyclopædia Britannica. (n.d.-b). Polyhedron. Retrieved 11-07-2019, from `https://www.britannica.com/science/polyhedron`

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General)*, *120*(3), 253-281. Retrieved from `https://rss.onlinelibrary.wiley.com/doi/abs/10.2307/2343100` doi: 10.2307/2343100

Fukuda, K., Liebling, T. M., & Lütolf, C. (2001). Extended convex hull. *Computational Geometry*, *20*(1), 13 - 23. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0925772101000323` (Selected papers from the 12th Annual Canadian Conference on) doi: https://doi.org/10.1016/S0925-7721(01)00032-3

Hosten, S. (2003, 09). *Extreme points, vertices and basic feasible solutions.*

Hui, C. W. (2001). *Polyhedra.*

Jamasb, T., & Pollitt, M. (2003). International benchmarking and regulation: an application to european electricity distribution utilities. *Energy Policy*, *31*(15), 1609 - 1622. doi: https://doi.org/10.1016/S0301-4215(02)00226-4

Kuosmanen, T. (2001). Dea with efficiency classification preserving conditional convexity. *European Journal of Operational Research*, *132*(2), 326 - 342. (Data Envelopment Analysis) doi: https://doi.org/10.1016/S0377-2217(00)00155-7

Niknazar, P., & Bourgault, M. (2018, 01). Analysis of axioms and assumptions of data envelopment analysis: Application for efficiency measurement in project management contexts. *International Journal of Productivity and Quality Management*, *23*, 446. doi: 10.1504/IJPQM.2018.090259

Penenberg, D. N. (2016). Mathematical statistics: Basic ideas and selected topics, 2nd edn, vols i and ii p. j. bickel and k. a. doksum, 2015 boca raton, chapman and hall–crc xxii + 548 pp., $99.95 (vol. i); 438 pp., $99.95 (vol. ii) isbn 978-1-498-72380-0. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *179*(4), 1128-1129. Retrieved from `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssa.12217` doi: 10.1111/rssa.12217

Petersen, N. C. (1990, 03). Data envelopment analysis on a relaxed set of assumptions. *Management Science*, *36*(3), 305.

Post, T. (2001a). Estimating non-convex production sets –imposing convex input sets and output sets in data envelopment analysis. *European Journal of Operational Research*, *131*(1), 132–142. doi: https://doi.org/10.1016/S0377-2217(00)00053-9

Post, T. (2001b). Transconcave data envelopment analysis. *European Journal of Operational Research*, *132*(2), 374 - 389. (Data Envelopment Analysis) doi: https://doi.org/10.1016/S0377-2217(00)00158-2

Richardson, A. (2015, 04). Nonparametric statistics: A step-by-step approach. *International Statistical Review*, *83*. doi: 10.1111/insr.12095_3

Rockafellar, R. T. (1970). *Convex analysis.* Princeton University Press.

Royden, H. L., & Fitzpatrick, P. (2018). *Real analysis.* Pearson.

Sarkis, J. (2006). Preparing your data for dea. In N. Avkiran (Ed.), *Productivity analysis in the service sector with data envelopment analysis* (p. 115-123). The University of Queensland.

Varian, H. (1992). *Microeconomic analysis* (Third ed.). New York: Norton.

Varian, H. (c2014). *Intermediate microeconomics : a modern approach.* Ninth edition. New York : W.W. Norton & Company.

Wallnerström, C., Grahn, E., Wigenborg, G., Wether Öhling, L., Bobadilla Robles, H., Alvehag, K., & Johansson, T. (2016, 08). The regulation of electricity network tariffs in sweden from 2016..

Wilde, D. (2000). A library for doing polyhedral operations..

Wilson, P. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business Economic Statistics*, *11*(3), 319–323. Retrieved from `http://www.jstor.org/stable/1391956`

Wilson, P. (2008). Fear 1.0: A software package for frontier efficiency analysis with r. *Socio-Economic Planning Sciences*, *42*(4), 247 - 254.

LOUVAIN-LA-NEUVE | BRUXELLES | MONS | TOURNAI | CHARLEROI | NAMUR
Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/lsm**