Louvain School of Management

and

Katholieke Universiteit Leuven

Benchmarking the efficiency of healthcare systems among the OECD countries between the years 2000 and 2015 using Data Envelopment Analysis with a focus on Belgium

Research Master's Thesis submitted by
Sébastien Vanderaa

with a view of getting the degrees
Katholieke Universiteit Leuven — **Master 120 credits in Business Engineering**

Supervisor at KUL
Erik Schokkaert

Supervisor at LSM
Per J. Agrell

Academic Year 2017-2018

ABSTRACT

The objective of this study is to benchmark the healthcare systems of the OECD member countries between the years 2000-2015 using Data Envelopment Analysis (DEA). Two output-oriented models have been constructed under the assumption of variable returns to scale, each representing a subpart of healthcare systems. First, a "Medical Care" model has as inputs the number of doctors, hospital beds and the health expenditure per capita. Second, a "Lifestyle" model consists of the following inputs: the consumptions of tobacco and alcohol, the supply of fruits and vegetables and the health expenditure per capita. Both models share the same outputs, including the life expectancy at birth, the infant survival rate and the perceived health status. Results present the technical efficiencies, rankings for Belgium, and the Malmquist index. Australia, Canada, Chile, Spain, Iceland, South Korea, Mexico and New Zealand are consistently efficient in both models. Belgium is not efficient relatively to the other OECD members. Besides, rich countries do not appear to be necessarily more efficient.

ACKNOWLEDGMENTS

I would like to express my gratitude to Professor Per Joakim Agrell and Professor Erik Schokkaert. I am thankful for their aspiring guidance, constructive criticism and flexibility during the research.

I feel lucky to have had the opportunity to work with recognized experts in their respective fields.

I would like also to acknowledge the support of Christophe Vanderaa, assistant researcher at the KULeuven, for his help with regards to the use of the software R.

TABLE OF CONTENTS

1. INTRODUCTION

The Organization for Economic Cooperation and Development (OECD) consists of 36 countries and strives to improve the economy of the member states and the social status of their population. During the last decades, the OECD faced many challenges regarding the healthcare as demand for health has enhanced up to the point that it surpasses the capacity of healthcare (HC) systems. For instance, one of the most significant factors is the aging population according to Yaya and Danhoundo (2015). Besides, one of the goals of the OECD is to keep improving HC performances while keeping the costs as low as possible.

Over the years, the share of GDP spent in HC all over the world underwent an increase of 17%. This is illustrated in Figure 1.1 between the years 2000 and 2015 (source: Worldbank). It is expected that public spending on health and social care will increase faster than the economic growth over the next 50 years, in line with recent trends across the OECD (Rumbold et al. (2015)).

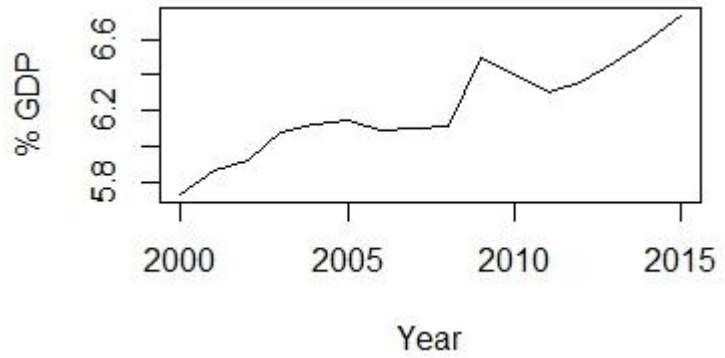


Figure 1.1 Health expenditure in the world from 2000 to 2015 as a % of the GDP

The healthcare (HC) system plays a vital role in any economy as discussed in a study by Bloom et al. (2003). Indeed, the population health status has a significant impact on the prediction of the economic growth. Having a population with good health seems to increase the productivity which is of high interest for governments and companies. Besides, the most important determinant of welfare of the Belgian is health according to the research conducted by le Bureau fédéral du Plan (Bureau fédéral du Plan (2017)). Between the years 1983 and 2008, the life expectancy rose of 6 years in average (OECD (2011)). More generally, one can state that the health status is getting better.

With the augmentation of HC expenses and its key role on a country's welfare, it is crucial for governments to be able to benchmark the performances of their HC systems given their spending policy. Indeed, public policies are crucial for the development of health. In constrast to a classic competitive market, a country's HC system does not undergo high pressure from competitors. Consequently, inefficient systems are not naturally identified by the "market".

Therefore, it is necessary to compare the performances between different HC systems. It leads to a double learning process. First, a country learns about the current performance of its HC system compared to the others. It enables a country to assess if the HC system is doing relatively good or bad. Arising from this, the second purpose is to identify the sources of (in)efficiencies by observing how the other systems are functioning and identify the gaps in outcomes or excesses in resource utilization. It supports decision-making processes of public policies and investments.

1.1 Research Question

As just mentioned above, the ultimate goal of this thesis is to assess the performance of HC systems of the OECD member through benchmarking between the years 2000 and 2015. The method used is known as Data Envelopment Analysis and will be later described in this work.

Throughout the study, a government's perspective is taken. In other words, the idea is to conduct the research to offer new insights to the Belgian government. Note that the role of the state in HC systems will be defined. Hence, the current efficiency situation of Belgium HC system is presented relatively to the others HC systems of the OECD members. An analysis of the efficiency's evolution over time is also computed. This might be useful to evaluate passed policy reforms. Finally, sources of inefficiencies are identified as well as other efficient OECD members that could inspire the Belgian decision makers by looking at the policy practices of efficient countries.

The research is organized as follows. As opening, the production theory is covered as it forms the basis of efficiency evaluations. Next, efficiency measures are presented to conduct our analysis and briefly discuss effectiveness before dealing with some developments of DEA. Then, a summary of the literature regarding the HC systems performance evaluations of OECD countries is presented. Finally, the specifications of our models are described from which the results are computed and analysed.

1.2 Limitations

Though the used method is very popular, many critics exist regarding its utilization. Indeed, DEA, and more generally efficiency frontier approaches, are often adopted as efficiency evaluation tool even though it comprises some drawbacks. The validity of the method outcomes highly depends on the quality and the choice of the data. Consequently, the analysis is limited by a few factors.

Quality aspects regarding HC are partly neglected due to data unavailability or incompleteness. For instance, the number of doctors is used in our model, but no distinction is made between the competences or expertises of doctors across countries. Moreover, other quality aspects are not included. These may concern the accessibility, costs or the general curing process that patients experienced (level of suffering, time to cure, etc.). Lastly, specific country characteristics are not part of the study. Indeed, geographical and social differences exist between countries such as the physical isolation of population or the work conditions that may affect the access to HC or the development of health issues respectively.

Some of these limitations are discussed in Lefebvre et al. (2015). According to them, frontier approaches only englobe part of the efficiency, known as productive efficiency. Furthermore, they highlight the issue of missing data. In an ideal situation, one can construct a model to gauge efficiency using any variable. In reality, a gap exists between ideal and available data. This comes back to the example of the data regarding the doctor's expertise.

Furthermore, our modelling does not consider input and output costs or prices by focusing on technical efficiency.

2. PRODUCTION THEORY

The purpose of this section is to define and describe the basis of the so-called production theory, which forms the basis of the method used in this study.

2.1 Setting definition

Production theory can be applied to different kinds of units, e.g. firms, NGO's, hospitals or governments. For the sake of clarity, we refer these units by what they could be commonly described as: *Decision Making Units*, noted as DMU.

Assume there are K DMUs indexed by $k = 1, \dots, K$. Rasmussen (2011) establishes that factors of production, called inputs, are part of a production plan and generate end products or services, called outputs. Let's consider this idea in the following setting for a simple DMU as described in Figure 2.1.

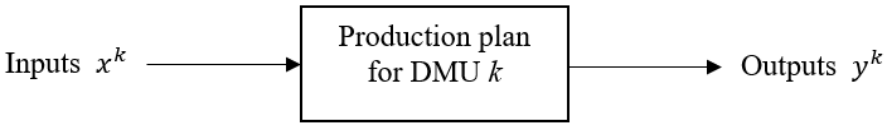


Figure 2.1 General production process

Each DMU uses x^k inputs that are processed to produce y^k outputs. The subscript *k* denotes the DMU to which the inputs or outputs belong.

Suppose there are *m* inputs for the DMU *k*, we can denote the *inputs* for DMU *k* as the m-vector

$$x^k = \left(x_1^k, \dots, x_m^k\right) \epsilon \, \mathbb{R}_+^m$$

Likewise, if we have *n* outputs for DMU *k*, we have that the *outputs* for DMU *k* is the n-vector

$$y^k = \left(y_1^k, \dots, y_n^k\right) \epsilon \, \mathbb{R}_+^n$$

Here, the inputs and outputs are defined as non-negative real numbers. This is done by assuming that the resources needed and produced by DMUs are all positive in value or quantity (or equal to 0). Another note is that superscripts are used to designate the DMU and subscripts for the type of inputs and outputs.

The *production plan* of DMU *k* is thus a point in the space resulting from the pair of input and output vectors. It can be defined as follows:

$$Production\ plan^{\,k} = \left(x^k, y^k\right) \epsilon \, \mathbb{R}_+^m \times \mathbb{R}_+^n$$

An input vector x^k is associated with a unique output vector y^k. This production plan is based on observed, and thus past data. Further, no output exists with associated inputs equal to zero. This principle is known as the *no free lunch* assumption.

2.2 Technology set

In a benchmarking application, the measured efficiency of a DMU is always a *relative* efficiency to the top performers in its industry. The following part is based on Chambers (1988) and Rasmussen (2011). It is assumed that the agents in an industry share a common technology. This means that all DMUs can be compared to the same technology set. The technology *T* can be defined more formally as

$$T = \{ (x, y) \in \mathbb{R}_+^m \times \mathbb{R}_+^n \mid x \; can \; produce \; y \}$$

Note that there are no superscripts for the inputs *x* and the outputs *y* as *T* is shared across all DMUs. The conditional part of the definition stating that *x can produce y* must be known, or at least approximated. Two approaches can be applied to learn what possibilities exist in terms of amount of inputs *x* needed to produce an output level *y*. Experts or engineers could estimate a theoretical production function. This approach seems at first glance the most appropriate as it will result in a technology that incorporates the perfect theoretical efficiency. However, this is mostly true for simple production processes or systems. Benchmarking DMUs in a complex setting like any kind of industry (e.g. healthcare) using this method is obviously not appropriate. Indeed, the setting is too complex to evaluate a theoretical production plan and identifying the perfect efficiencies. The second approach constructs the technology based on empirical observations. Hence, the resulting technology set is guaranteed to be feasible knowing that another agent in the same industry can perform at the given level. More assumptions must be established to claim that the technology set feasible (see later).

Suppose that all DMUs are part of the technology set, i.e.

$$(x^k, y^k) \in T \qquad \forall k$$

The smallest technology set is only composed of the observed points. This may not be interesting as a DMU can only improve its efficiency by exactly copying what a better performer is doing. Nonetheless, it gives us a lower bound on the technology set. To enlarge the production possibilities, the technology set is approximated thanks to several assumptions and properties. In the following sections, these are discussed one by one.

2.3 Free disposability

Let's assume that it is feasible to produce a certain amount of outputs *y* with *x* inputs. What will happen if the resources are increased, e.g. the inputs? Our intuition tells us that more outputs could be produced, or at least the same quantity. Presented in Chambers (1988), this corresponds to the idea of *free disposability*. According to McFadden (1978), when a quantity *y* of outputs can be produced given a quantity *x* of inputs, then the same quantity of outputs is producible with less or equal level of resources. To be more precise, the author defines it as the *free disposability of inputs*. More formally, it gives

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

By ignoring existing costs to dispose of inputs and outputs, free disposability allows to freely dispose of unwanted resources and outputs.

Likewise, the *free disposability of outputs* corresponds to the idea that a firm or organisation can produce less or equal outputs, given the same amount of inputs, i.e.

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

By combining both assumptions, we get the *free disposability of inputs and outputs* (Bogetoft (2011)). A production setting that uses more inputs to produce less outputs must be part of the technology set, i.e.

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

This reasoning is quite straightforward. However, when considering prices or preferences for certain inputs or outputs (see allocation efficiency), this becomes more delicate.

Let's illustrate the principle using a single-input and single-output setting. Consider four DMUs using x^k inputs to generate y^koutputs where *k* denotes the DMU *k* to which it corresponds. In Figure 2.2, one can observe the effect of free disposability on the production possibilities. The result of the function has a 'stair' shape. It can be seen that every possibility that is at the right side of *x* ($x' \geq x$) and below *y* ($y \geq y'$) is part of the technology set *T*.

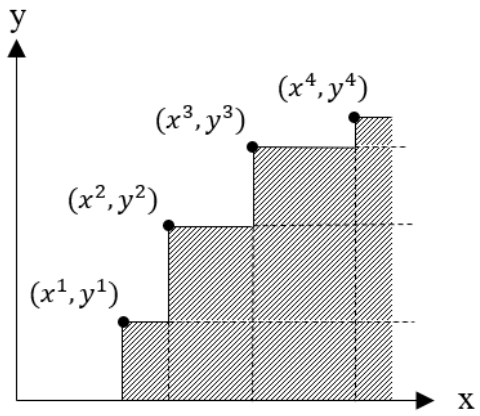


Figure 2.2 Free disposability

This type of technology is often called *Free disposable hull (FHD)* as it is constructed from the free disposability assumption. Thanks to the new assumption, the definition of the technology set *T* can be rewritten as

$$T = \{(x, y)\, \epsilon\, \mathbb{R}_+^m \times\ \mathbb{R}_+^n \mid \exists k\, \epsilon\, \{1, \dots, K\} :\ x^k \leq x,\ y^k \geq y\}$$

In other words, the production plan (x, y) is part of *T* if there exists at least one DMU *k* such that it produces more or equal quantity of outputs with less or equal quantity of inputs.

2.4 Convexity

The free disposability assumption discussed above is quite limited in terms of taking advantage of the best practices. In order to shift to the best performers of the industry, a DMU can move its production plan to the left (decrease in inputs *x* for a same level of outputs *y*), upwards (increase in outputs *y* for a same level of inputs *x*) or both (decrease inputs and increase in outputs). In the FDH technology, the 'stair' shape does not allow DMU to look for an improvement that lies in between 2 observed production plans.

Discussed in Banker et al. (1984) and Chambers (1988), the assumption of *convexity* enlarges the technology set T by taking the weighted average of observed productions plans. Every production plan resulting from the convexity assumption is only theoretical in some sense as it is not observed empirically. These new points are generated from observed data, but do not exist yet in the studied industry.

To understand the concept, the property is illustrated in Figure 2.3 that is constructed from the original points in Figure 2.2, to which convexity is applied. The hashed areas correspond to the additional surface covered by the technology using convexity on the top of the FDH. Informally, one can state that the steps of the stairs from the FDH are filled using convexity.

More formally, T is *convex* if for any two points (x^i, y^i) and (x^j, y^j) that are in the technology set and weights $0 \leq \lambda \leq 1$, the convex combination $(1-\lambda)(x^i, y^i) + \lambda (x^j, y^j)$ must also be in the technology. I.e.

$$(x^i, y^i) \epsilon T, (x^j, y^j) \epsilon T, \ 0 \leq \lambda \leq 1 \Rightarrow (1-\lambda)(x^i, y^i) + \lambda (x^j, y^j) \epsilon T$$

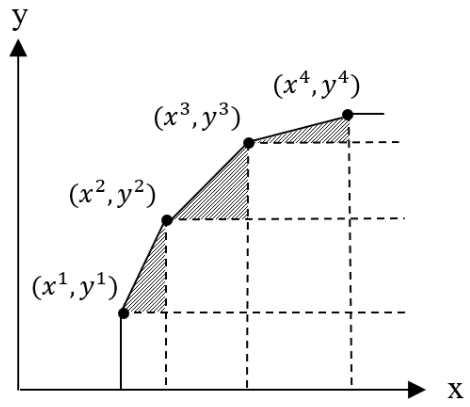


Figure 2.3 Convexity

For certain situations, convexity appears naturally, especially where organizations control the allocation of resources for different processes. Moreover, convexity gives us a good approximation for non-observed data. Taking the average of two existing points is a reasonable estimation method.

Convexity allows us to interpolate from observed DMUs to new fictive DMUs that lies in between the existing ones. As it enlarges the technology *T*, benchmarking can now rely on fewer observations which is essential when large volumes of data are not available (when the number of DMUs is small).

2.5 Scaling

In the previous section, constant returns to scale are implicitly assumed by ignoring any change in proportion of outputs produced per input. Now, consider a restaurant hiring 10 waiters and serving 100 customers per night on average. The owner decides to hire an additional employee. As a result, 120 customers are now served per night. Having as input the waiters and as outputs the number of customers served, an increase of 10% in input leads to an increase of 20% in outputs. Thus, the returns to scale are not constant. Four types of returns to scale exist and are defined in Chambers (1988) and Bogetoft (2011). When incorporating returns to scale, the empirical technology can be defined as follow

$$T(\gamma) = \{(x, y)\, \epsilon\, \mathbb{R}_+^m \times\ \mathbb{R}_+^n \mid \exists \lambda\, \epsilon\, \Lambda^K(\gamma) : \sum_{k=1}^{K} x^k \lambda^k\ \leq x, \sum_{k=1}^{K} y^k \lambda^k\ \geq y\ \}$$

where γ is a list of assumptions regarding the scaling.

The simplest one is the *constant returns to scale* (CRS) that assumes a constant increase or decrease in outputs when the input quantity is increased or decreased respectively. Any proportional change in inputs results in the same proportional change in outputs, such that it remains in the technology set. To put it more formally

$$\Lambda^K(CRS) = \{\ \lambda\, \epsilon\, \mathbb{R}_+^K \mid \sum_{k=1}^{K} \lambda^k\ \geq 0\ \}$$

where the λ are said to be free.

The second one corresponds to the illustration of the restaurant. I*ncreasing returns to scale* (IRS) describes the case where the outputs grow faster than the inputs. It also implies that any production can be upscaled while remaining in *T*, i.e.

$$\Lambda^K(IRS) = \{\ \lambda\, \epsilon\, \mathbb{R}_+^K \mid \sum_{k=1}^{K} \lambda^k\ \geq 1\ \}$$

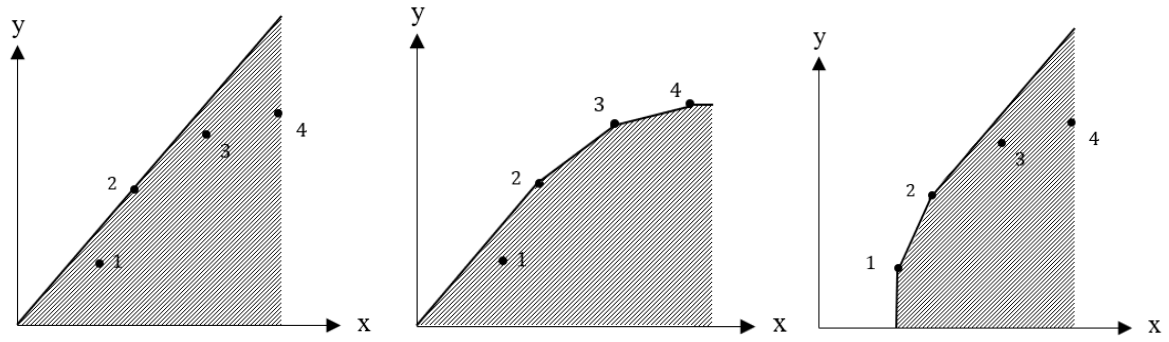
In practise, this increase may be due to a gain in experience or higher efficiency in terms of equipment use (non-exhaustive causes).

Decreasing returns to scale (DRS) is the opposite of IRS. Now, the λ lies in between 0 and 1, meaning that the scale is reduced as the input are augmented while still being part of the set *T*. Mathematically, it results in

$$\Lambda^K(DRS) = \{\ \lambda\, \epsilon\, \mathbb{R}_+^K \mid 0 \leq \sum_{k=1}^{K} \lambda^k\ \leq 1\ \}$$

Under DRS, the output will grow proportionally less than the input. Norman and Stoke (1991) show in their book how these properties affect visually the technology set. From Figure 2.4 and the definitions given above, one can see that the DRS and IRS production possibilities (shaded area) form a subset of the CRS production possibilities since the λ are more restricted in DRS and IRS than in CRS. Under CRS (Figure 2.4a), the technology set is made of a straight line from the origin while respecting the properties of free disposability and convexity discussed

above. Here, the outputs of DMU 2 can be extended proportionally to the inputs through a straight convex combination due to the fact that λ is not restricted. In a DRS environment, a straight line is drawn from the origin and then curved by DMU 3 and 4. It clearly illustrates the idea that increasing inputs *x* will end in an increase that is proportionally less in outputs *y*. With IRS, *T* does not start at the origin, but vertically from the level of inputs of DMU 1.



(a) CRS ***(b) DRS*** ***(c) IRS***

Figure 2.4 Returns to scale

Lastly, the *variable returns to scale (VRS)* (Ozcan (2008); Norman and Stoke (1991)) can be conceptualized as the weakest assumption as no rescaling is possible. Visually, it can be represented as in Figure 2.5 where it can be seen that the efficient production frontier encompasses the observed points to form the smallest subset possible. Formally, we have

$$\Lambda^K(VRS) = \{\, \lambda \,\epsilon\, \mathbb{R}_+^K \mid \sum_{k=1}^{K} \lambda^k \;= 1 \,\}$$

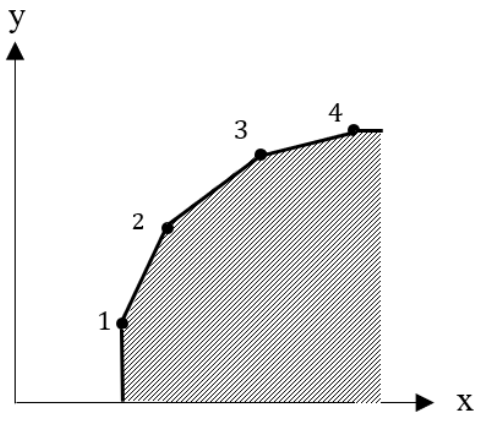


Figure 2.5 Variable returns to scale (VRS)

2.6 Additivity

Finally, *additivity*, also called *replicability* assumes that if two production plans lie within the technology set, then the production plan resulting from the sum of the two initial plans is feasible as well. It is defined as

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

The property is illustrated in Figure 2.6 retrieved from Bogetoft (2011).

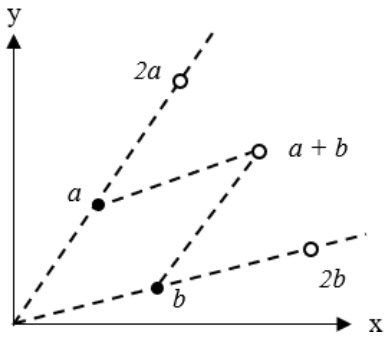


Figure 2.6 Additivity

Note that not only the sum of two different production plans is outlined, but also the sum of two identical plans. If a convex technology exists and that additivity is applied, a *T* with constant returns to scale is obtained. Indeed, the various properties that we have just covered are linked.

2.7 Minimal Extrapolation Principle

In practise, the technology *T* is unknown and must be estimated. Therefore, empirical observations of production plans must be extrapolated to build a complete technology set that includes all observed DMUs. Hence, the theoretical technology *T* is estimated by the empirical extrapolated technology *T**. According to the *minimal extrapolation principal* presented in Banker et al. (1984), *T** is the smallest set of $\mathbb{R}_+^m \times \mathbb{R}_+^n$ that contains all observations and satisfies some assumptions discusses previously such as free disposability, convexity or some specified returns to scale.

3. EFFICIENCY MEASURES

The idea of efficiency often relies on a ratio of inputs used over the number of outputs produced from theses inputs. Obviously, using as few inputs as possible to generate a maximum quantity of outputs is what we are aiming for. Various types of efficiency may be considered when benchmarking DMUs. These are discussed in the following sections.

3.1 Dominance

The concept of *dominance* allows us to rank DMUs based on their efficiency. It is the simplest form of efficiency comparison. Consider the production plans (x^1, y^1) and (x^2, y^2), one can state that DMU 1 dominates DMU 2 when it produces equal or larger quantities of outputs *y* using equal or less inputs *x* while both vectors cannot be the same. More formally defined by Koopmans (1951), it gives us the following definition

$$(x^1, y^1) \succ (x^2, y^2) \; iff \; x^1 \leq x^2 \; and \; y^1 \geq y^2 \; and \; (x^1, y^1) \neq (x^2, y^2)$$

where $\succ$ stands for "is dominating". Remember that x^k and y^k both represent vectors (inputs and outputs respectively) for DMU *k*. Given the definition of dominance above, it implies that the dominating DMU should be at least better in one dimension as the two DMUs cannot be similar. The principle of dominance is easy to illustrate and is represented in Figure 3.1, where DMU 1 dominates DMU 2. In addition, the production plan of DMU 3 is neither dominating or dominated by any of the two other DMUs.

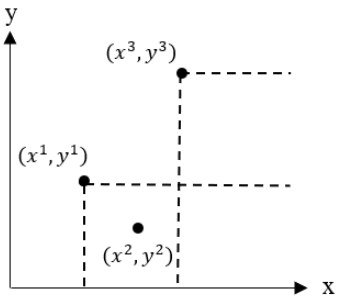


Figure 3.1 Dominance

Now, how to rank the DMU according to their efficiency and the concept of dominance? In general, efficient DMUs are defined as the one that cannot be dominated by any other DMU. This implies that a perfectly efficient performer does not have to dominate all the others, but must not be dominated by any other competitor. Koopmans defines efficiency by identifying the dominating and dominated DMUs. *Koopmans-efficiency* (also referred as the *Pareto-Koopmans efficiency*) can now be defined as

$$(x, y)\ is\ efficient\ in\ T\ iff\ it\ cannot\ be\ dominated\ by\ some\ (x', y') \epsilon\ T$$

Note that the other DMUs compared to the evaluated DMU are part of the technology set. Otherwise, there would always be a better performer if there is no restriction on the set of DMUs to compare with. Moreover, this way of defining efficient DMUs does only allow us to say which DMUs are (in)efficient but it is not possible to quantify the efficiency or to rank efficient DMUs. In Figure 3.1, two efficient DMUs are identified according to the Pareto-Koopmans efficiency's definition: DMUs 2 and 3 as none of them are dominated by any other DMU.

3.2 Efficient production frontier

In section 2, the technology set and its properties have been discussed. We learned that in a single-input and single-output setting, the DMUs that lie on the most top-left of the graph of *T* are the best performers. We can adapt Koopmans definition to make this idea more obvious by defining the *efficient technology set*

$$T^E = \{(x, y)\ \epsilon\ T \mid (x, y)\ is\ efficient\ in\ T\}$$

This is often called the *efficient production frontier* (Farrell (1957); Chambers (1988)). One of the goals in efficiency evaluation methods is to construct this frontier of efficient DMUs as we will see later. In Figure 2.3, the frontier is illustrated by the full line.

3.3 Farrell efficiency – Orientation

From Koopmans's definition, inefficient DMUs can be highlighted. However, we are limited to discriminating the efficient from the inefficient DMUs. It should be interesting to look for way to quantify the level of inefficiency. This is done by one of the most popular approach to measure efficiency nowadays. In a multi-input and multi-output setting, Farrell (1957) proposes a method to measure efficiency, commonly known as the *Farrell efficiency*. In his paper, two approaches are described to study efficiency. On one hand, an *input-oriented efficiency* that evaluate the possible proportion of inputs reduction to become efficient. The idea is to compare a DMU to a perfectly efficient one producing the same level of outputs *y*. Then, by taking the ratio of the input of the evaluated DMU over the efficient level of inputs, we get a value of the efficiency *E*. To put it more formally

$$E = \min\{\ E > 0 \mid (Ex, y)\ \epsilon\ T\}$$

In other words, the value *E* represents the maximum reduction of inputs such that the production plan still lies in the technology set, i.e. on the efficient production border.

Likewise, the *output-oriented efficiency* measures the proportion of outputs that could be increased to reach the best performing DMUs.

$$F = \max\{ F > 0 \mid (x, Fy) \epsilon T\}$$

It represents the maximum expansion in outputs while remaining in the technology set. In a single-input and single-output setting, the direction of improvement is parallel to the output y axis.

Farrell efficiency results in continuous values that can be ranked. Note that the input-oriented efficiency is equal or smaller 1 whereas the output-oriented efficiency is equal or larger than 1. This is due to the respective definitions given above.

3.4 Technical and Allocative Efficiency

Up to now, we have worked without any information about the input costs or output prices. In that respect, the importance of some resources regarding others was not taken into account. From Farrell (1957) and Debreu (1951), one could say that the first way of working enables to assess the *technical radial efficiency (TE)* that splits into the *technical radial input efficiency* and the *technical radial output efficiency* (the *radial* denomination is ignored in the rest of this work). To understand these concepts, the following diagram retrieved from Farrell's work is used.

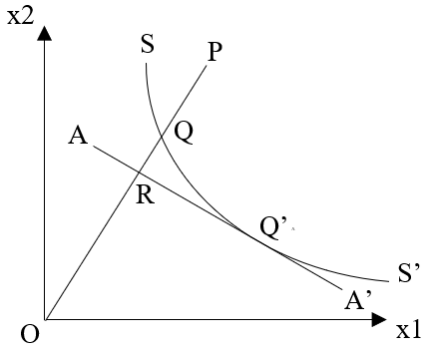


Figure 3.2 Technical and allocative efficiency

For the sake of simplicity, only an input perspective is taken (all what is discussed can similarly be applied to an output perspective). In Figure 3.2 (retrieved from Farrell (1957)) a two-inputs setting is drawn (x_1 and x_2) where the isoquant SS' is represented. Assume that we want to evaluate DMU P. According to Farrell (1957), the technical input efficiency corresponds to

$$TE = \frac{OQ}{OP}$$

Additionally, the isocost line AA' is drawn corresponding to an equal ratio of the prices of the two inputs. This allows one to compute new measure called *allocative efficiency* (Farrell (1957); Färe and Primont (2011)) that evaluates how well a DMU allocates its production resources. Here, the studied DMU P (OP) is compared to the point Q' (OQ'). But this goes beyond the scope of our research that only deals with technical efficiency.

3.5 Super Efficiency

As an industry may end up with different DMUs being efficient, how can we compare them? Efficient DMUs have an efficiency score of 1, and we could wonder who is the best? This problem can be handled using the so-called *super efficiency*, originally stated by Anderson and Peterson (1993). It has also been used for identifying outliers in Banker, Change and Zheng (2017).

It works by removing an efficient DMU, referred as DMU 0, from the technology *T* (the feasible production set). Then, the efficiency measures are recalculated for all DMUs based on the technology that does not incorporate DMU 0. From an input-orientation, super efficiency scores larger or equal to one are obtained for efficient DMUs. This process should be iteratively repeated for each efficient DMU. The new efficiency scores are called *super efficiencies*, allowing us to rank the efficient DMUs or even detect outliers.

3.6 Effectiveness

So far, efficiency estimation only has been discussed as a tool to benchmark DMUs. The concept of efficiency is about minimizing the waste or, in the same idea, getting the most out of the less. Why is effectiveness not measured instead of efficiency as a benchmarking tool?

Being *efficient* is about doing the things right whereas being effective means doing the right things. Norman and Stoker (1991) define *effectiveness* as the attainment of pre-determined goals while efficiency is the use made of resources in the attainment of output quantities, in the context of environmental factors. According to Ozcan (2008), efficiency relates the purpose of minimizing the resources to achieve a certain level of output while effectiveness evaluates the outputs, by asking if we are producing the right outputs using the current inputs. You can be efficient and not effective or effective but not efficient. Ideally, a DMU is both of them.

How can we know if we are doing the right things (effectiveness)? This can be done by studying the preferences of an agent or the utility of a production plan. Ozcan (2008) also states that *e*ffectiveness in health care measured by outcomes or quality is of prime importance to many components including patients, clinicians, administrators, and policy makers, but that measuring them is more problematic than efficiency measures. Therefore, our models try to incorporate part of the notion of effectiveness through some quality aspects, under the assumption that higher quality levels are always preferred.

4. DEA DEVELOPMENTS

The Data Envelopment Analysis, known as DEA, is a mathematical programming method that estimates best practices (resulting in an efficient production frontier) and benchmarks different DMUs through relative efficiency against these best practices. Thanks to the minimal extrapolation principle (see section 2.7), DEA constructs the smallest set that "envelops" all observed data (Bogetoft 2011). This activity analysis has diverse properties such as the free disposability, convexity, scaling and additivity described in section 2.

4.1 Origins

Farrell (1957) proposed a method (LP formulation) to assess technical efficiency, also valid for a multiple-input or -output case. It is worth to note that the future of computers was uncertain at that time. It is part of Charnes et al. (1978) method's success to have anticipated the growing power of modern computers. The development of DEA is made possible thanks to the preliminary work of Debreu (1951), Koopmans (1951), Shepard (1953 and 1970), Farrell (1957) and the work on linear fractional transformation of Charnes and Cooper (1962).

The DEA originated in Charnes et al. (1978) providing a scalar measure of technical efficiency for multiple input and multiple output. By then, only constant returns to scale (CRS) could be applied. This will be later addressed by further developments. The first applications were mainly aiming the public sector, emphasizing on the efficiency of organization (Seiford (1996)).

4.2 Assumptions

DEA models rely on two major assumptions that must be specified when constructing the model. Firstly, the model orientation must be choose as described in section 3.3 (input- or output-oriented models). Secondly, the returns to scale discussed in section 2.5 are specified. These will be covered in section 4.6 within the DEA LP formulation.

4.3 Primal formulation

In this section, the basic DEA models are developed. The following setting and notation are used:

DMUs $j = \{1, \ldots, n\}$

Inputs x_{ij} $i = \{1, \ldots, m\}$

Outputs y_{rj} $r = \{1, \ldots, s\}$

Weights u_r, v_i

The primal formulation of this linear program can be expressed in two ways. First, an input-orientation maximizes the numerator of e_0 (the level of outputs). The problem can be expressed as the following LP, known as the *Multiplier Model*:

$$\max e_0 = \sum_{r=1}^{s} u_r \; y_{r0} \qquad (1)$$

Subject to:

$$\sum_{i=1}^{m} v_i \; x_{ij} - \sum_{r=1}^{s} u_r \; y_{rj} \geq 0; \qquad j = 1, \dots, n$$

$$\sum_{i=1}^{m} v_i \; x_{i0} = 1$$

$$u_r, v_i \geq 0; \quad \text{r} = 1, \dots, \text{s}; \quad \text{i} = 1, \dots, \text{m}$$

As DMUs are rated relative to the other, the subscript '0' designs the tested DMU (denoted as DMU 0). Inputs and outputs are given (from past data) and the weights are computed by an LP. Hence, the model solves a fractional problem (ratio of outputs to inputs).

The idea is that we look for a scalar measure e_0 representing the minimal proportion of current inputs used by the DMU 0 to produce a certain amount of outputs. The level of inputs is fixed by the constraint equalizing the sum of inputs for the tested DMU to one. The same exercise could have been done for an output-oriented setting by minimizing the input weighted combination given a certain output level.

4.4 Dual formulation

In LP theory, each primal formulation is associated to its dual formulation. Called the *Envelopment Model*, the dual of problem (1) is

$$\min \; z_0 \qquad (2)$$

Subject to:

$$-\sum_{j=1}^{n} \boldsymbol{\lambda}_{0j} \; x_{ij} + z_0 \; x_{i0} \geq 0; \qquad i \; = 1, \dots, m$$

$$\sum_{j=1}^{n} \boldsymbol{\lambda}_{oj} \; y_{rj} \geq \; y_{r0}; \quad \text{r} = 1, \dots, \text{s}$$

Problem (2) can be decomposed to ease the interpretation. The value z_0 (equivalent to *E* in section 3.3) is the minimal reduction of inputs for DMU 0 while staying in the technology set. The first constraint states that the weighted combination of inputs ($\boldsymbol{\lambda}_{0j} \; x_{ij}$) should not exceed the proportion of the input of the tested DMU ($z_0 \; x_{i0}$), for each input. The second one states that for each output, the weighted combination of outputs is at least as great as the output level

of DMU 0. Note that z_0 is smaller or equal to one in any case and represents a "measure of how much we can reduce all inputs of DMU 0 in the same proportion to produce a performance in line with the weighted combination" (Norman and Stoke (1991)).

4.5 Slacks variables

In problem (2), it is evident that some constraints may not achieve equality for some inputs or outputs for the tested DMU. Inequality means that an opportunity exists for improvement. To measure the level of relative excess utilization of inputs or potential improvement of outputs, slack variables are introduced. Adding these extra variables, problem (2) becomes

$$\min \ z_0 \qquad (3)$$

Subject to:

$$z_0 \, x_{i0} - \sum_{j=1}^{n} \boldsymbol{\lambda}_{oj} \, x_{ij} - \ s_i^- = 0; \quad i \ = 1, \dots, m$$

$$\sum_{j=1}^{n} \boldsymbol{\lambda}_{oj} \, y_{rj} - \ s_r^+ = y_{r0}; \quad \text{r} = 1, \dots, \text{s}$$

$$\boldsymbol{\lambda}_{0j}, \, s_i^+, s_r^- \geq 0$$

The model enables to eliminate all inefficiencies through the slacks. The reduction in inputs required to become efficient is equal to the value of s_i^-. Similarly, s_r^+ stands for the gap's value between the efficiency border and the DMU's actual output level.

Charnes et al. (1978) refine problem (3) into a formulation that is more popular when using the dual DEA LP. The slack values are maximized as they are added to the objective function (OF). As a result, we have as OF

$$\min z_0 - \ \epsilon[\sum_{i=1}^{m} s_i^+ + \ \sum_{r=1}^{s} s_r^-]$$

where ϵ is a non-Archimedean value (an infinitely small value). Hence, the slacks are directly maximized in the OF, but does not affect significantly the result (z_0) thanks to ϵ.

4.6 Returns to scale in DEA

The returns to scale problem is tackled by using observed data and will be tested in section 6.6. The incorporation of returns to scale in the DEA LP is done through an intercept in the OF. As a result, the following LP is obtained

$$\max e_0 = \sum_{r=1}^{s} u_r \, y_{r0} + c_0 \qquad (4)$$

Subject to:

$$\sum_{i=1}^{m} v_i \, x_{ij} - \sum_{r=1}^{s} u_r \, y_{rj} + c_0 \geq 0; \qquad j = 1, \ldots, n$$

$$\sum_{i=1}^{m} v_i \, x_{i0} = 1$$

$$u_r, v_i \geq 0; \quad r = 1, \ldots, s; \quad i = 1, \ldots, m$$

c_0 to be defined depending on the returns to scale option (CRS, IRS or DRS)

where c_0 represents the intercept. It characterizes the returns to scale depending on whether it is less than, equal to, or greater than zero. See Banker et al (1984) for the development of the following result: when c_0 is strictly positive, we have increasing returns to scale. When c_0 is strictly negative, we have decreasing returns to scale. When c_0 is equal to zero, constant returns to scale is applied.

The associated dual LP is

$$\min \; z_0 \qquad (5)$$

Subject to:

$$-\sum_{j=1}^{n} \boldsymbol{\lambda}_{0j} \, x_{ij} + z_0 \, x_{i0} \geq 0; \quad i = 1, \ldots, m$$

$$\sum_{j=1}^{n} \boldsymbol{\lambda}_{oj} \, y_{rj} \geq \; y_{r0}; \quad r = 1, \ldots, s$$

$$\sum_{j=1}^{n} \boldsymbol{\lambda}_{oj} = 1$$

which is identical to (2), except for an additional constraint where the $\boldsymbol{\lambda}_{oj}$ are summed to one (assuming here that VRS should be applied). In problem (2), no weighted average constraint is defined, meaning that we are facing constant returns to scale (CRS). If $\sum_{j=1}^{n} \boldsymbol{\lambda}_{oj} \leq 1$, DRS is assumed and if $\sum_{j=1}^{n} \boldsymbol{\lambda}_{oj} \geq 1$, IRS is assumed as explained in section 2.5.

4.7 Malmquist index

As the purpose of this work is to benchmark the healthcare systems over several years, it will most probably happen that the performance of DMUs varies over time. With units changing over time, the empirical technology set based on the unit's production plans will also change. A new efficiency measure is needed to grasp this effect. In our application, we will be dealing with time series data, making the following efficiency measure relevant to consider.

$E^i(s,t)$ assesses the efficiency of a DMU *i* in period *s* against the technology in period *t*. Named after Malmquist (1953), the *Malmquist index* evaluates how much a DMU improves from period *s* to period *t* (where period *t* comes after period *s*). To compute the index, two ratios are needed. First,

$$M^s = \frac{E(t,s)}{E(s,s)}$$

evaluates the improvement of the DMU in period *t* compared to period *s*. Note that period *s* is used as benchmark. If there is indeed an improvement, the ratio will be larger than 1. Here, M^s measure the improvement regarding technology *s*. Likewise, we could be interested in measuring the improvement compared to technology *t*. In that case, the index becomes

$$M^t = \frac{E(t,t)}{E(s,t)}$$

To get the Malmquist index, the geometric average of the two previous ratios is computed. Hence, it is defined as

$$M(s,t) = \sqrt{\frac{E(t,t)}{E(s,t)}\,\frac{E(t,s)}{E(s,s)}}$$

If one is interested in studying the change or progress in *T*, the following measure can be used, known as the *technical change index* (technology change) derived from Farrell (1957) and developed in Färe et al. (1992).

$$TC(s,t) = \sqrt{\frac{E(t,s)}{E(t,t)}\frac{E(s,s)}{E(s,t)}}$$

$M(s,t)$ is considered as a broader measure of efficiency change whereas $TC(s,t)$ focuses on the technology variations. If the goals is to know if a DMU is closer to the efficiency frontier than in the previous period, we would look at the *efficiency change index* (also known as the catch-up effect), defined as

$$EC(s,t) = \frac{E(t,t)}{E(s,s)}$$

5. DEA APPLICATIONS TO HEALTHCARE

It is fundamental to understand what a healthcare (HC) system incorporates. The purpose of this section is to get a brief overview of what a HC system is, what it does and its components. Besides, a review of the literature of is presented regarding the evaluation of HC systems.

5.1 Healthcare systems

In its annual report (WHO (2000)), the World Health Organisation (WHO) defines a HC system as something that includes all the activities whose primary purpose is to promote, restore or maintain health. In parallel to this definition, the WHO identifies 3 intrinsic objectives linked to HC systems. First, it must improve the population health. It must also meet the notion of responsiveness that can be understood as the capacity to meet the expectations of the population regarding health systems. It can be seen as the perceived quality of service of the patients. Lastly, a health system should be fair in the financing and risk protection.

Murray and Frenk (1999) developed a framework to assess the performance of HC systems according to the definitions of the WHO. Moreover, each dimension is divided into two categories: the level and the distribution. Their result is present in the form of a framework (Table 5.1).

	Level	Distribution	
Health	✔	✔	Efficiency
Responsiveness	✔	✔	
Fairness in financing	X	✔	
	Quality	Equity	

Table 5.1 Health System Goals

According to their research, the efficiency should be measured on five components and in the three dimensions, i.e., all except for the level of fairness in financing which makes no sense as fairness is intrinsically linked to the distribution. This must be kept in mind when selecting the inputs and outputs of our model.

Our research seeks to perform a cross-country performance evaluation. In case of inefficiencies, who should be held responsible? In other words, who is monitoring healthcare systems? Rothgang et al. (2005) studied in their paper the role of the state in HC systems, which is probably what most would think as a responsible for HC performance in a country. While many

studies focus on the role of state as budgeting HC expenses (*Financing* role), they point out that there are other roles too. One of them is to be a *Service Provider*. The share of public services (opposed to private services) is a good indicator of the importance of role of a state. It has also a *Regulation* role between the service providers (public/private), the financing agencies (public/private) and the potential users of the system (patients). These roles are represented in Figure 5.1 (OECD Health Data (2002)). At the basement, the goals, values and perceptions of HC are to be found. Two pillars representation the *Financing* and *Service provision* role of the state, and on top, as a roof, the *regulation* role between the 3 types of actors.

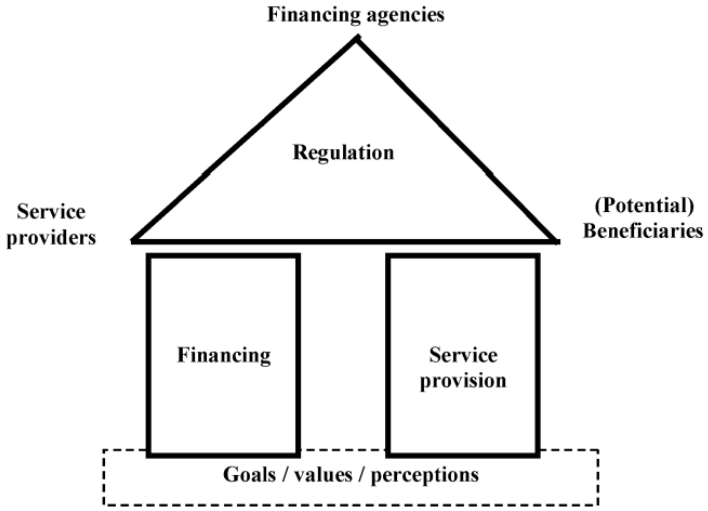


Figure 5.1 Roles of the state in HC systems

Two types of studies can be considered when analysing a healthcare system. On one hand, a micro-level provides a narrower angle on the study of efficiency in healthcare. For instance, it could be an evaluation of hospitals or clinics efficiencies within a unique country. On the other hand, the macro-level takes a broader view and tries to identify the overall healthcare system's performance. The literature presents more micro-level studies (Hadad et al. (2013)).

The purpose of this work is to evaluate national healthcare systems in their entirety. One of key issue is to select the right indicators to represent the inputs and outputs of the system and make the appropriatex assumptions. Notwithstanding, it is crucial to define the boundaries of our DMUs to create an evaluation frame on which HC systems will be benchmarked. Murray and Frenk (1999) raised the question of finding a correct definition for healthcare boundaries. From their research, it is suggested that most definitions consider individual health services provided by clinic or hospitals are included in all existing definitions in the literature. Nonetheless, education is not often part of these definitions while it is a health determinant. Moreover, they claim that the boundaries are mainly arbitrary chosen.

As explained above, the state is at the forefront of a country's HC system organization and performance. Therefore, the objective is to evaluate the efficiency of these systems within a range of significative influence of the state. As 'significative' may be interpreted in various ways, the boundaries must be delimited. In our benchmark study, an HC system consists of all individual health services that patients get from hospitals, clinics, medical cabinet or physician's offices. Healthcare personal is also considered, including nurses, pharmacies, dentists or even medical pedicure. Finally, social factors such as consumption habits are within our definition of the HC system boundaries. Regarding the last components, governments may influence population consumptions through taxes or prohibition (set limits and standards) on certain product (tobacco, sugar, etc.). It is to note that intercultural differences also affect consumption habits. Elements that are not kept in our evaluation frame are the access to medical care or the fairness in financing. If possible, the "quality" of services is incorporated but this seems more delicate to achieve.

5.2 Previous studies measuring HC efficiency

To evaluate a system's efficiency, significative indicators (inputs and outputs) must be selected and considered in the study. The purpose of this section is not to select the features for the DEA, but rather to get an overview of what has already been done and what should be considered when evaluating the HC system. Furthermore, methods, assumptions used (returns to scale and orientation when available) and results are presented.

In the literature, many papers relate to the evaluation of performance of HC systems on a micro-level (national level at maximum). Even though, our research has a macro-level, the covered studies give interesting insights for our research, especially regarding the choice of variables to consider. Grosskopf and Valdmanis (1987) assess the relative performance of public and not-for-profit (NFP) hospitals in California. According to them, health status is too difficult to capture. Therefore, they focus on the hospitals production and select four measures of outputs: acute care, intensive care, surgeries, ambulatory care and emergency care. The resources selected as inputs consist of the number of physicians, non-physician employment, admissions and the net plant assets. Public hospital appears to be more efficient than NFP hospitals. Hence, ownership plays a key role in the efficiency level. In their results, each type of hospital has its own best practise frontier.

In his attempt to evaluate Romania's public healthcare system (for hospitals), Nistor (2017) sets as inputs the number of beds available, the number of doctors and the non-salary operating

expenditures. An input-oriented model is constructed under VRS using as outputs the total operating revenues, the number of cases and hospitalization days. Out of the twenty hospitals, five are identified as inefficient. A Tobit regression indicates that the number of cases and the total operating revenues have a significant positive influence on the efficiency of a hospital. A second result is that non-salary operating expenditures and the number of doctors significantly affects negatively the efficiency score. More recently, Olesen, Petersen and Podinovski (2015) applied DEA with probabilistic assurance regions for measuring hospital efficiency (70 Danish hospitals). The observed costs are used for each hospital and the number of discharges as outputs. The idea of assurance regions is to restrict the possible solution set of weights for inputs and outputs, avoiding ending in unrealistic or undesired results.

Many macro-level (cross-country) studies are conducted in the context of HC systems' evaluation, which is the ultimate goal of this work. In their cross-country comparison of efficiency between 191 countries, Evans et al. (2000) used a parametric approach with the DALE (Disability Adjusted Life Expectancy) variable as output and the health expenditure per capita with the education level as inputs. DALE is defined as a "measure of the equivalent number of years of life expected to be lived in full health" in a report of the WHO (WHO (2000); Mathers et al. (2000)). It is one of the first study to estimate efficiency on such a large sample of countries. According to their efficiency ranking for the OECD countries, Italy and France are classified as having the best HC systems, whereas Latvia and Estonia end up in the last places.

In Cetina and Bahceb (2016), the efficiency of health systems of OECD countries is measured by DEA. The inputs used in their model are the number of doctors and patient beds per thousand people as well as the health expenditure per capita. As outputs, life expectancy (LE) and infant mortality rate (IMR) are selected, which are popular outputs in the evaluation of efficiency of HC systems in the literature. A DEA model using the assumption of CRS is applied on the data in a two-stage approach. In the first stage, the model is applied on the 34 members of the OECD from which 8 are identified as outliers. From the remaining 26 countries, Canada, Czech Republic, Iceland, Ireland, Israel, Japan, Korea, Poland, Slovenia, Sweden and the UK are identified as efficient. Besides, the results point out that some countries with high efficiency may have bad HC indicators. This is due to the low level of inputs used, making the ratio of outputs over inputs high. So, this reminds us that efficient HC systems (like in Mexico) does

not mean that the health provided to the population achieves high standards of quality, accessibility or low costs for the patients.

Afonso and St Aubyn (2005) conducted non-parametric models (FDH and DEA) to assess the efficiency of expenditures in education and health in OECD countries. In the model for health, the number of hospital beds and doctors are used, in parallel with the number of nurses. Regarding the outputs, the authors used the same outputs as the previous paper, namely LE and the infant survival rate (ISR). An output-oriented model is adopted where the outputs must be maximized. It implies that the increase of outputs must have a positive impact on both the objective function and the HC system. Therefore, it may happen that some features should be transformed. Here, the infant mortality rate is converted into a survival rate. The following formula is applied: ISR = (1000-IMR)/IMR), where ISR stands for Infant Survival Rate. In their paper, results for two models (FDH and DEA) are presented. The FDH model identifies eleven OECD countries as efficient, i.e. Canada, Denmark, France, Japan, Korea, Norway, Portugal, Spain, Sweden, the UK and the US. Seven efficient countries result from the DEA model: Canada, Japan, Korea, Portugal, Spain, Sweden, the UK and the US. Note the overlap between the two methods, increasing the confidence of the results.

Recently, Ozcan and Khushalani (2016) constructed a Dynamic Network DEA to conduct a cross-country (34 OECD countries) analysis on health efficiency between the years 2000 and 2012. Two dimensions of health are explored and aggregated into an overall estimate of efficiency. On one hand, the public health includes all population-based services and on the other hand the medical care is based on individual services. For the public health evaluation, the inputs set are the smoking and alcohol consumption, body weight and expenses on public health. Note the socio-economic indicators used here. LE is the unique output for the public health model. Regarding the medical care assessment, employees in HC, the number of hospital beds and the medical technology are considered while hospital discharges and consultations are used as outputs. A Malmquist index is also computed to analyse the evolution in efficiency over time. The study shows that countries that underwent health reform had their HC systems improved, especially regarding the public health dimension. 13 out of the 34 DMUs are identified as inefficient with the UK and the US ends up being the less efficient.

A two-stage approach is implemented in the work of Samut and Cafri (2016) where the purpose is to analyse the efficiency determinants of HC in OECD countries. DEA is applied in the first stage setting as inputs the number of beds, MRI (magnetic resonance imaging), CT

(computerized tomography), nurses and physicians. The outcomes considered are the number of discharges and the infant survival rate. In the second stage, a Tobit panel regression is performed to identify significant environmental factors. GDP, income, education and the number of private hospitals contributes positively to efficiency. Moreover, GDP has been identified as a determinant in the efficiency of HC systems (Hadad et al. (2013); Joumard et al. (2008)). Inversely, a negative relationship exists between the efficiency and the spending in health. The DEA model covers the period 2000-2010 and identified Mexico, the UK and Turkey as fully efficient during this time interval. Some rich countries are below the average, i.e. Japan, Iceland, France and Belgium. The Malmquist index indicated that Greece had the highest increase in efficiency with its +2% while Ireland faced the biggest decrease. Considering the total factor of production change (TFP change) of DMUs, the biggest decrease lies between 2000 and 2001 (because of the technology that did not progress measure by the Malmquist index decomposition). No increase in the TFP is detected during the 10-year period.

Another cross-country (OECD) comparison is performed that measures the technical efficiency of hospitals over the period 2000-2009 in Varabyova and Schreyögg (2013) by a non-parametric DEA and a parametric Stochastic Frontier Analysis (SFA). A two-stage analysis is performed where the second stage regresses the explanatory variables by a truncated regression. Moreover, a resampling method (bootstrapping) is applied and three countries are identified as outliers. The authors choose as inputs the number of beds and the hospital employment and as outputs, the number of discharges and the mortality rate in-hospital. Results lead to the conclusion that higher HC expenditures per capita implies higher technical efficiency scores. Moreover, a country like Japan that has a good HC system presents a weak technical efficiency. This can be explained by the same reasoning as in Cetina and Bahceb (2016). Finally, inequality in income and length of stay in a hospital have a negative relationship with technical efficiency.

In the research of Afonso et al. (2003), the objective is to evaluate the performance of the public sector across 23 OECD countries. Total government spending as a percentage of the GDP is set as input. Taking the share of GDP as a HC spending measure allows one to avoid biases due to differences between the costs of living across countries (e.g. doctor wages). A public sector performance (PSP) is created that is based on 7 indicators reflecting the outputs of government policies. In the health dimension, LE and IMR are used in the evaluation. As presented in section 2.3, the FDH is also an envelopment technique but is more restrictive than DEA as the frontier has a stair-shape because it cannot use a weighted average of two observed

points like in the traditional DEA method. In Afonso et al. (2003), FDH is used to rank DMUs according to their efficiency scores. It appears that small governments have a more efficient spending policy than large ones that could reduce on average 35% of their spending to attain similar results. Moreover, 15 EU countries show lower efficiency scores than the US or the average of the OECD members. The best performers in overall efficiency are Australia, the US, Japan, Switzerland and Luxembourg.

To evaluate the efficiency of the public sector in new member states of the European Union compared to the ones in emerging markets by DEA, Afonso et al. (2006) used the same inputs and outputs as in Afonso et al. (2003). Singapore, Thailand, Cyprus, Korea, and Ireland lie on the efficiency frontier or are very close to.

Gupta and Verhoeven (2001) assess the efficiency of government expenditures experiences from 37 countries in Africa in comparison with countries from Asia and the Western Hemisphere. In their FDH model, education and health spending per capita are set as inputs while the outputs are represented by social indicators such as LE, IMR and protections against specific diseases.

The following papers have contributed to the choice of relevant variables for modelling HC systems. Socio-economic factors are less common but can be found in Hadad et al. (2013) where diverse DEA models are constructed to analyse what are the efficiency determinants. One of them includes the consumption of fruits and vegetables per capita as input. Furthermore, fat intake appears to play a significant role in efficiency. In like manner, smoking and alcohol consumptions along with the body weight are adopted as inputs in Ozcan and Khushalani (2016) as explained above. In section 3.5, super-efficiency has been presented as a ranking tool for efficient DMUs. It has been applied in Hadad et al. (2013) where two DEA models are implemented. Both models use the same outputs: LE and ISR. Inputs differ between the models as one of them focuses on the physical and financial resources whereas the other one also takes into account the financial aspect but with more socio-economic factors in addition. Hence, health expenditure per capita, physician's density and inpatient bed density are included in the first model, while health expenditure per capita, GDP per capita and consumption of fruits and vegetables per capita are in the second one. The two models identified the following countries as efficient: Czech Republic, Estonia, Iceland, Japan, Korea, Poland, Portugal, and Slovenia. According to the super-efficiency ranking, Japan and Iceland have the highest performance

score in both models. Besides, it has been shown that the consumption of fat plays as significant role.

A consensus exists on the inputs contributing to the performance HC system (Joumard et al. (2008)). Most of the inputs included in this list are already considered in some of the studies mentioned previously. Three vectors define these inputs according to the consensus. First, the resources of HC includes the number of doctors, the country's GDP or the money spend (private or public) in HC (non-exhaustive). Lifestyle seems to play a determinant role in the population health status through the consumption of fat, alcohol, smoking, sugar, calories, fruits and vegetables. Finally, socio-economic factors must be considered, including income per capita, pollution or the education level. The OECD report (OECD (2013)) comes approximately to the same conclusions for inputs contributing to HC status.

In their paper, Jourmard et al. (2008) identified various indicators of outputs, e.g. DALE, the Potential Years of life lost (PYLL). Health-adjusted Life Expectancy (HALE) and Disability Free Life Expectancy (DFLE) reflect both mortality and morbidity. The author also considered sick leave and public satisfaction. Unfortunately, the availability of data sets narrow limits to the utilization of certain outputs.

We have shown evidence that LE and IMR are popular indicators of HC system performance. Mohan and Mirmirani (2008) assess the significance of factors contributing to LE and IMR between the years 1990 and 2002 for 25 OECD countries. For the LE, empirical results show that the alcohol consumption does have a significant impact. On the contrary, the size of HC's workforce, the acute care bed days number per capita as well as the education level have a serious effect on a population's LE. Regarding the IMR, the results about the impact of alcohol coincide with those of the LE. In addition, the total expenditures on health per capita does not affect the IMR, while it extends the LE.

As a last remark, the OECD report (OECD (2013)) describes relevant indicators to represent the health status. First, two indicators relate to the LE (at birth, by sex and education level). Next, several features linked to the mortality and thirdly, more disease- and satisfaction-oriented indicators are described. Again, LE and mortality are here chosen as output measures, as well as more socio-economic indicators in a second instance.

5.3 Contribution to literature

The previous section presents parts of a rich literature regarding the gauging of HC systems. The scope of previous works includes micro-level studies (e.g. hospitals at national level) but also macro-level (e.g. cross-country evaluations of HC systems). Few studies have been conducted considering intertemporal effects on the efficiency in HC. Hence, this work will contribute to the literature by evaluating cross-country HC systems among OECD countries using recent data (until 2015) and applying the Malmquist productivity index.

In our work, two models are constructed. From the definition of HC system (section 5.1), two subdivisions can be deducted. Each model will represent one of these HC part. On one side, we have the *medical care* division that incorporates the curative role of a HC system. This holds to heal the diseased part of the population. On the other hand, the *public health* part tries to prevent diseases and improve the well-being of the population through a good health status. Appropriate variables will be chosen in the next section. This distinction will allow us to get a specific performance evaluation for each division while in most studies, an overall HC system score is computed. Besides, an attempt is made to incorporate quality aspects within our models.

Lastly, the research is conducted with a focus on Belgium. As mentioned in the beginning, the objective here is to support the Belgian decision makers in their work. Hence, the analysis for Belgium's results goes more in depth.

6. MODEL SPECIFICATIONS

The following section describes the model build to benchmark the HC systems of the OECD members. Some comments are made about data pre-processing and the DEA assumptions including the choice of model orientation (input-oriented or output-oriented) and the returns to scales.

6.1 Data sources

Most of the data originate from the OECD databases (www.stats.oecd.org and www.data.oecd.org) that collect and measure numerous variables across multiple topics, including a topic dedicated to health. Few data were retrieved from the World Bank database (www.data.worldbank.org/topic/health) and the World Health Organisation (WHO) database (www.who.int). These databases have the benefit to be reliable and to offer a wide range of variables to exploit.

6.2 Missing values

Regarding missing values in the datasets, it is essential to treat them carefully as DEA is data-driven. In Figure 6.1, data about the number of consultations for Denmark is plotted and a continuous decline appears. One can observe similar trends (continuous growth or decline) for the other countries and variables. Therefore, missing data are imputed by applying a standard interpolation (spline). As a result, the obtained data for Denmark for the consultation variables are shown Figure 6.1 where the interpolated values are shown in orange. This technique is applied on all data.

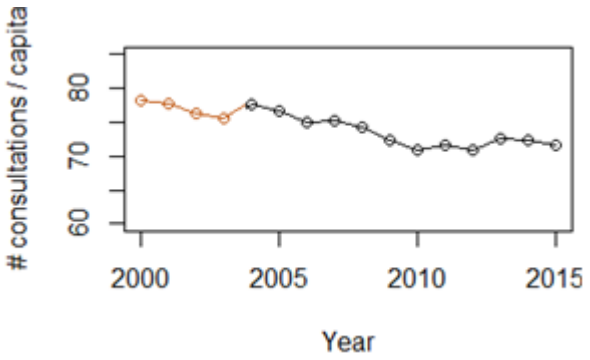


Figure 6.1 Missing data interpolation

6.3 Variable selection

This part relates to the choice of inputs and outputs. The variables shown in Table 6.1 stem from the analysis of the literature (see section 5.2). Several resources and performance indicators are presented that could be used in the model. We will go through them one by one.

Table 6.1 Candidates variables for inputs and outputs

Input variables	Output variables
Doctors	Life Expectancy
Nurses	Infant mortality rate (or survival rate)
Hospital beds	Hospital discharges
Healthcare expenditure per capita	Consultations
Education level	DALE
Non-salary operating expenditures	Cancers
Body weight	Perceived health status
Smoking consumption	
Alcohol consumption	
Food and supply consumption	
Medical technology	

As seen in the previous section, the number of **doctors** and **nurses** are popular variables among efficiency estimation studies for HC systems. The importance of these variables in the performance of HC systems is straightforward. Nevertheless, let us point out that there is an implicit assumption stating that no difference exists between doctors and nurses across countries. In other words, the expertise and competences of doctors and nurses is assumed to be similar. In reality, disparities exist but are ignored for the sake of simplicity.

Another commonly used input is the number of **hospital beds**. It represents the total number of beds in hospitals of a country. The main benefit of this variable choice is its association to the capacity level of a country's HC system. Indeed, one can see the number of beds in hospital as the level of a country's HC capacity.

Next, the **HC expenditure** per capita is often selected as input. The consideration for this variable is all the more significant for this study as one of the goals is to inform the state about its national HC system. Likewise, non-salary operating expenditures could be interesting, but because of unavailable data, it won't be retained.

The level of **education** may play a role in the outcome of a healthcare system. An educated population is probably more conscious about what could harm their health and make wiser decisions. Besides, the education level is often positively correlated with the welfare of an individual. Despite that, it won't be retained in our study as it may be more difficult from a government perspective to act in the short run. Some of its effects are to be found in other selected variables described in the next paragraph more related to the way of living.

The next considered inputs are closely linked to the population's lifestyle. The consumption habits and population characteristics are frequently added as inputs to gauge a HC system.

Hence, the following variables are retained in our study: **smoking** and **alcohol consumption**, as well as the food consumption are included using the amount of **supply of fruits** and **vegetables**. The fat and sugar consumption are ignored as the interpretation can be difficult. For instance, the supply of sugar or fat is vital to some point but may become excessive and unhealthy. Regarding the body weight, data about obesity are too sparse and cannot be exploited in our case.

As last input variable, a country's **medical technology** seems interesting to consider. Nevertheless, the related data present a non-negligible number of missing values and a universal definition for medical technology is difficult to construct. Incorporating expertise of HC employment or the technology level of medical machines is quite complex. Therefore, it is not retained for our models.

On the output side, two variables appear as essential in the evaluation of a HC system performance, the **LE** and **IMR** (or equivalently, **ISR** through the following transformation $\mathrm{ISR} = (1000 - \mathrm{IMR})/\mathrm{IMR})$). In our application, ISR is used and represents a ratio of surviving infants over dying ones before the age of 1. In almost every study, one can find at least one of these variables, but often both are used. These outputs have the benefit not only to incorporate the ability of a HC to heal the population, but also reflect the more general health status. Besides, the **DALE** (Disability Adjusted Life Expectancy) is very similar to the measure of LE, but is not retained as its definition may be more ambiguous.

The capacity of the system is represented by the number of hospital beds input. Likewise, the amount of **consultations** and **hospital discharges** can play the role of capacity possibility of a HC system. It would be more accurate to define it as a throughput rate rather to associate it directly to the notion of capacity. Yet, it is more considered as intermediate variables that consumes inputs resources and help achieve HC objectives. Therefore, it is not retained in our study.

The following output indicator is the **perceived health status** (PHS) that refers to the proportion of the population (15 years old or more) that considers their health to be "good or better". The measures resulted from a standardized survey in which individuals are asked about their perception of their own health (OECD Health Statistics (2018)). The difference between this variable and those seen so far is that it incorporates some quality aspect retrieved from the population which makes it an interesting variable to work with.

Tables 6.2 Correlation between inputs and output variables

		Input variables							Outputs variables		
		Doctors	Nurses	Hospital beds	Healthcare expenditure	Fruits & vegetables	Tobacco consumption	Alcohol consumption	LE at birth	ISR	PHS
		x1	x2	x3	x4	x5	x6	x7	y1	y2	y3
Inputs	x1	1	0.292	-0.219	0.347	0.200	-0.065	0.236	0.342	0.402	0.197
	x2	0.292	1	0.043	**0.772**	-0.225	-0.381	0.256	0.498	0.495	0.465
	x3	-0.219	0.043	1	0.109	-0.118	-0.034	-0.056	0.178	0.064	-0.315
	x4	0.347	**0.772**	0.109	1	0.059	-0.415	0.190	0.687	0.495	0.478
	x5	0.200	-0.225	-0.118	0.059	1	0.134	-0.274	0.209	-0.033	0.210
	x6	-0.065	-0.381	-0.034	-0.415	0.134	1	0.159	-0.350	-0.287	-0.304
	x7	0.236	0.256	-0.056	0.190	-0.274	0.159	1	0.005	0.130	-0.104
Outputs	y1	0.342	0.498	0.178	0.687	0.209	-0.350	0.005	1	0.632	0.413
	y2	0.402	0.495	0.064	0.495	-0.033	-0.287	0.130	0.632	1	0.023
	y3	0.197	0.465	-0.315	0.478	0.210	-0.304	-0.104	0.413	0.023	1

The last covered output variable revolves around the frequency of **cancers**. Here, one does not deal with cancer healing but only the number of incidences. Unfortunately, no recent dataset exists that could be exploited.

In fine, correlations for the selected variables are computed in Table 6.2. The correlation between the nurse variable and the health expenditure per capita is the highest with a value of 0,77. A significance test is applied of the correlation coefficient (alpha = 0.05) to check if there is a significant linear relationship between the variables. It appears that the correlation result is statistically significant. Therefore, the nurse variable is eventually not included in our DEA models.

A summary table of the chosen variables for our DEA models according to the discussion above is presented in Table 6.3. One can already see large disparities between DMUs just by looking to the minimum and maximum of each variable (over the period 2000-2016). Accordingly, outlier detection will be applied further.

Table 6.3 Selected variables description

		Units	Min	Max	Mean
Inputs	Doctors	Number / capita	0.9	6.260	3.019
	Hospital beds	Number / capita	0.00152	0.0147	0.005
	Healthcare expenditure	US $ / capita	219.03	5644.11	2087.175
	Tobacco consumption	Number of daily smokers	36 057.2	53 893 018	7 781 441
	Alcohol consumption	Liters / capita	1.2	14.8	9.369
	Fruits & vegetables supply	Kg / capita	118.6	368.5	218.3
Outputs	LE at birth	Years	70.1	83.9	78.892
	ISR	Ratio	32.78	1110.11	255.093
	PHS	% of pop "good/very good"	29.97	91.4	67.660

As having higher values for the tobacco and alcohol consumptions variables must have a negative effect on the HC system, they must be transformed. An inversion function (1 / tobacco consumption) is applied such that higher values can be interpreted as lower consumptions, and thus better impact on a population health. All other variables remain unchanged. Even though Lithuania is today officially a member of the OECD, the database of the OECD has not collected enough data to incorporate the country in our study. Therefore only 35 countries are included in our research.

6.4 Models

This section's objective is to describe our conceptual modelling which is composed of two distinct models. The purpose of computing results from two models is to enable the possibility to compare or to validate some findings as each model targets a specific part of HC systems.

The first model is referred as the "**Medical Care**" model, noted as *MC*. Its purpose is to represent the curative activities of a HC system. It includes hospital capacity and the healthcare employment. Each DMU is defined as a HC system of country *c* at year *t*, and this holds for the second model too. The MC model is sketched in Figure 6.2. Inputs from time *t* can be seen as resources and include the number of doctors, hospital beds and the health expenditure. Hence, the model reflects the "production" part of a HC system.

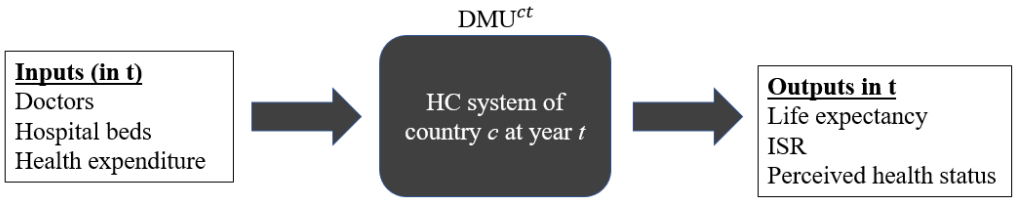


Figure 6.2 Model 1: Medical care (MC)

The second model encompasses the way of living and habits of populations. So, this model is referred as the "**Lifestyle**" model (LF). It does not model the ability of HC system to cure the sick patients, but rather deal with non-medical determinants of health. Therefore, the inputs at period *t* are the health expenditure, the consumptions of tobacco and alcohol and the supply of fruits and vegetables. This has the benefit to outline the non-medical aspect of health and is illustrated in Figure 6.3.

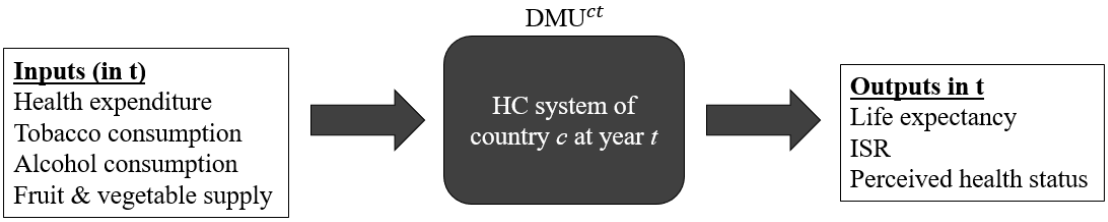


Figure 6.3 Model 2: Lifestyle (LF)

6.5 Model orientation selection

In DEA, a decisive assumption to make is the model orientation. As covered in section 3.3, one can choose between an input-oriented model where an inefficient DMU can become efficient by reducing its inputs, or an output-oriented model that takes the level of inputs as given (they

remain unchanged) and evaluate the efficiency through a potential increase of its outputs. The decision of the model's orientation must be discussed and clearly argued. Both orientations are discussed followed by the decision of the model's orientation.

On one hand, the input-oriented method is preferred since the objectives of OECD member countries converge as they are shaped by approaches of international institutions such as OECD and the World Health Organization (Cetina et al. (2016)). For instance, an input-orientation is chosen in the study of Nistor (2017).

On the other hand, output-oriented models are preferred according to Spinks and Hollingsworth (2009). They claim that the objective is to maximize the health outcomes and minimize inputs to attain a certain level of outputs. An example of output-oriented application has been adopted in the modelling of Afonso and St Aubyn (2005).

One of the goals of this study is to identify the sources of inefficiencies from a government perspective. The objective is to set targets given the amount of inputs. And as explained in the previous paragraph, the state must squeeze the most out of the inputs they have. Hence, output-oriented models are chosen for our application.

6.6 Testing returns to scale

Another assumption to introduce in our modelling relates to the returns to scale of a HC system's technology set. The Banker (1996) test is performed to check the returns to scale. To do so, the following hypothesis and statistical test (goodness of fit) are computed

$$H_0: \theta^{CRS} = \theta^{VRS}$$

$$H_a: \theta^{CRS} > \theta^{VRS}$$

$$t = \frac{\sum_{t=1}^{T}\sum_{k=1}^{K}(\theta_{kt}^{CRS} - 1)^2}{\sum_{t=1}^{T}\sum_{k=1}^{K}(\theta_{kt}^{VRS} - 1)^2}$$

where θ represents the efficiency score. The test follows a F distribution that has degrees of freedom equal to 512 (32 countries for 16 years). A p-value lower than 10e-16 is obtained for both models, meaning that we reject the null hypothesis with high certainty. Consequently, the models are computed under the assumption of **VRS**.

6.7 Number of DMUS

To get a valid DEA model, a minimum number of DMUs are required. Numerous rules of thumb exist to determine validity thresholds. The number of inputs and outputs is noted as *m* and *n* respectively. One can observe different rules of thumb used in DEA procedures. For instance, the number of DMUs should be at least *m*n* (Boussofiane, Dyson and Thanassoulis

(1991)), *2(m*n)* (Golany and Roll (1989), Dyson et al. (2001)) or *3(m*n)* (Dyson et al. (2001)). Nevertheless, as a general rule of thumb in DEA, the number of DMUs should be larger or equal then the maximum between *m*n* and *3(m+n)* (Banker et al. (1989)). In both our models, these conditions are all satisfied.

6.8 Form homogeneous group – Outlier detection

To increase the robustness of the models, it is necessary to work with a homogeneous group of DMUs. Outlier detection is performed on the data thanks to the log-ratio method (Agrell and Niknazar (2014)). A data cloud *D* is the determinant of the matrix composed of the inputs and outputs and can be interpreted as the volume of the data. The method removes a DMU from the data and look at the variations in the data cloud. A high decrease in the data cloud indicates that the DMU is an outlier. The following ratio is used

$$R(i) = \frac{D(i)}{D}$$

where $D(i)$ is the data cloud after removing DMU *i* (note that multiple DMUs can be removed). The number of DMUs to identify as outliers is defined by the peaks from the dotted lines in Figure 6.4. Indeed, a peak indicates a high difference in the data cloud volume, meaning that an outlier still subsists.

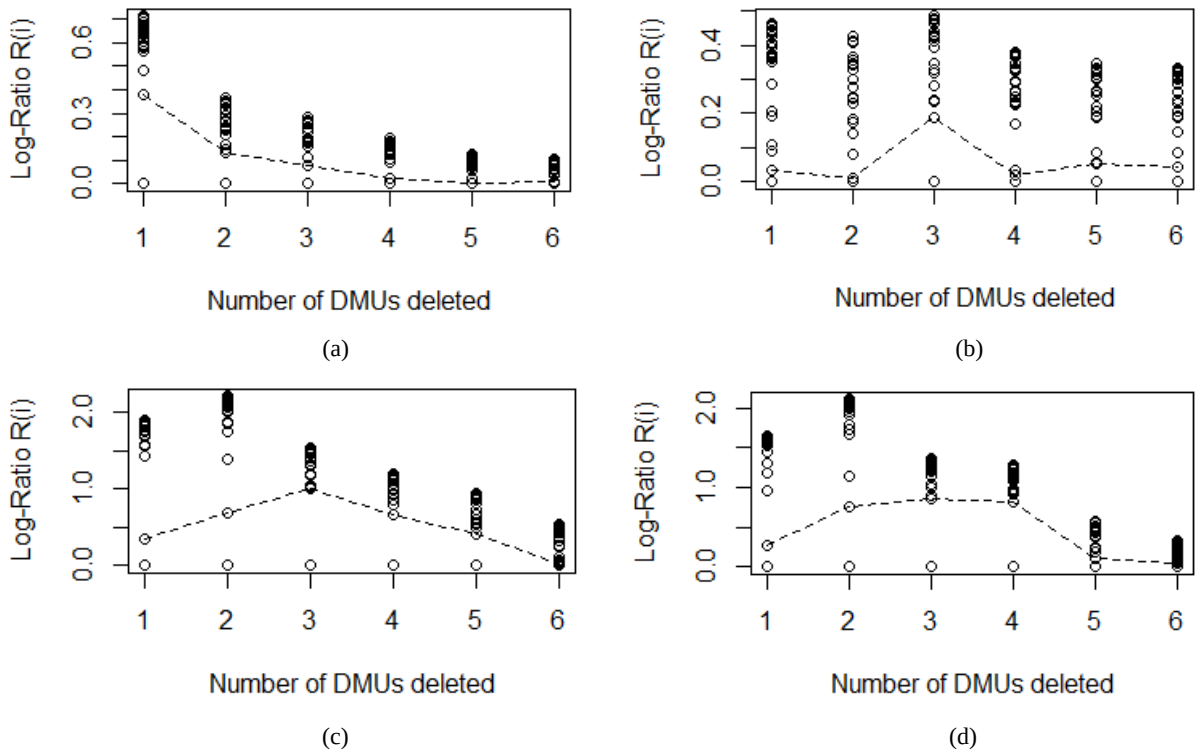


Figure 6.4 Log-Ratio per number of DMUs removed for (a) model MC in 2000 (b) model MC in 2015 (c) model LF in 2000 (d) model LF in 2015

Hence, three countries are removed from each model as three peaks from the dotted lines are clearly highlighted in Figure 6.4b and 6.4c, and in a less extend in Figure 6.4d. For those deleted DMUs, no efficiency score is computed, and they cannot become peers. For the model MC, Greece, Japan and Portugal are removed while the model LF does not include Iceland, Luxembourg and Turkey. The same DMUs are identified as outliers in 2000 and 2015 for each model, confirming the outlier detection results.

7. RESULTS AND ANALYSIS

7.1 Technical efficiency

The technical radial output efficiencies (referred as technical efficiency for the sake of simplicity) are computed for both models for the years 2000 until 2015. As explained above, the corresponding DMUs identified as outliers are not part of the computation. The results are shown in Table 7.1.

Regarding the model MC, multiple DMUs are identified as continuously efficient. These include Australia, Canada, Chile, Spain, Iceland, Israel, South Korea, Mexico, New Zealand and Sweden. Let us remind the reader that this does not mean that the HC systems are better, but that they achieve higher efficiencies. To illustrate this, consider Mexico's and Sweden's life expectancies. A Mexican citizen lives on average 75 years when born in 2015, whereas a Swedish baby is expected to live more than 82 years. One can conclude that the health status of the Swedish population is better, but both countries are performing efficiently according to our results. Besides, the Belgian HC system does not reach efficiency at any point in time. In fact, it presents some low performance results relatively to the other DMUs across the whole time interval.

In the lifestyle model, more DMUs are identified as continuously efficient, including Australia, Switzerland, Canada, Czech Republic, Chile, Spain, France, Germany, Greece, Hungary, Ireland, Iceland, Japan, Mexico, South Korea, Latvia, New Zealand, Poland and the United States. Here, Belgium presents better efficiency measures and even reaches efficiency in 2004, 2005 and 2014. Nevertheless, it will be shown that Belgium is performing badly relatively to the others in both models (see the ranking in section 7.5).

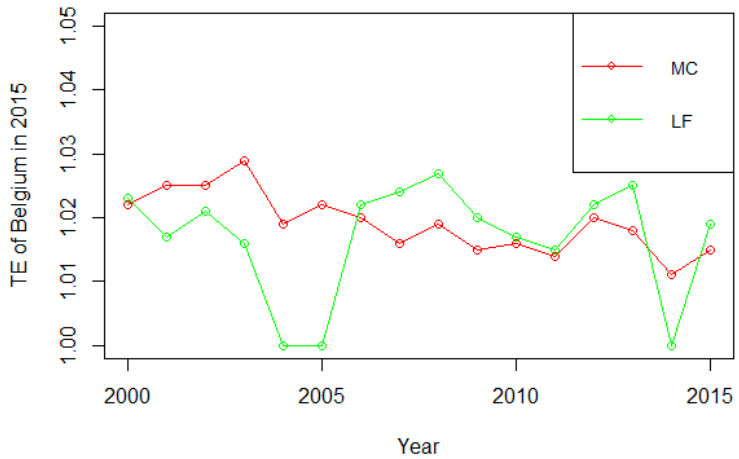


Figure 7.1 Technical efficiency of Belgium for the model MC and model LF

Table 7.1 Technical efficiencies for model MC

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
AUS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AUT	1,022	1,025	1,023	1,026	1,024	1,026	1,021	1,021	1,019	1,022	1,023	1,02	1,025	1,025	1,021	1,021
BEL	1,022	1,025	1,025	1,029	1,019	1,022	1,02	1,016	1,019	1,015	1,016	1,014	1,02	1,018	1,011	1,015
CAN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CHE	1	1	1	1,001	1	1	1	1	1	1	1	1	1	1	1	1
CHL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CZE	1	1	1,021	1	1	1	1,044	1,003	1,003	1,048	1	1,051	1,049	1,044	1,045	1,047
DEU	1,02	1,027	1,025	1,028	1,022	1,025	1,02	1,02	1,021	1,022	1,024	1,028	1,03	1,032	1,026	1,029
DNK	1,034	1,041	1,041	1,042	1,039	1,039	1,036	1,038	1,035	1,036	1,037	1,032	1,032	1,032	1,021	1,017
ESP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
EST	1,08	1,05	1	1,038	1,065	1	1	1,076	1,079	1	1	1	1,042	1	1,02	1,021
FIN	1	1	1	1	1	1	1,012	1,01	1,004	1,012	1	1,016	1,02	1	1,013	1
FRA	1,007	1,016	1,013	1,017	1,008	1,012	1,005	1,005	1,004	1,003	1,002	1	1,007	1,004	1	1,003
GBR	1,007	1,008	1,011	1,014	1,011	1,014	1,013	1,011	1,013	1,008	1,008	1,006	1,007	1,01	1,007	1,007
HUN	1,078	1,073	1,08	1,089	1,086	1,092	1,09	1,08	1,08	1,081	1,077	1,074	1,066	1,065	1,061	1,066
IRL	1,03	1,03	1,028	1,024	1,023	1,02	1,019	1,012	1,007	1	1	1	1	1	1	1,004
ISL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ISR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ITA	1	1,001	1	1,002	1	1	1	1	1	1,003	1,003	1,004	1,006	1,005	1,001	1,005
KOR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LUX	1,012	1,018	1,019	1,027	1,003	1	1,012	1	1	1	1,008	1,007	1,005	1,002	1	1
LVA	1	1	1	1	1,09	1	1,106	1,098	1,075	1,051	1	1,033	1	1	1	1
MEX	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NLD	1,013	1,017	1,021	1,02	1,014	1,016	1,014	1,01	1,01	1,007	1,008	1,009	1,017	1,015	1,009	1,011
NOR	1,011	1,017	1,018	1,015	1,011	1,016	1,012	1,015	1,012	1,014	1,013	1,016	1,016	1,01	1,006	1
NZL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
POL	1,026	1,027	1,029	1,008	1,006	1	1,015	1,027	1,046	1,051	1,039	1,036	1,028	1,038	1,023	1,029
SVK	1,055	1,048	1,058	1,06	1,061	1,067	1,071	1,075	1,08	1,081	1,082	1,075	1,074	1,074	1,069	1,071
SVN	1,027	1,023	1,023	1,03	1	1,028	1,02	1	1	1	1	1,011	1	1,012	1	1
SWE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
TUR	1	1,05	1,046	1,045	1	1	1	1	1	1,002	1	1	1	1	1	1
USA	1,01	1,013	1,016	1,016	1,017	1,015	1,017	1,02	1,02	1,013	1,009	1,01	1,013	1,012	1	1

Table 7.2 Technical efficiencies for model LF

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
AUS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AUT	1,017	1,015	1,008	1,016	1,011	1,011	1,007	1,008	1,007	1,017	1	1	1	1	1	1,013
BEL	1,023	1,017	1,021	1,016	1	1	1,022	1,024	1,027	1,02	1,017	1,015	1,022	1,025	1	1,019
CAN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CHE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CHL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CZE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
DEU	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
DNK	1,026	1,011	1,022	1,021	1,026	1,021	1,01	1,02	1,037	1,026	1,039	1,034	1,034	1,033	1,031	1,027
ESP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
EST	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1,008	1,003
FIN	1	1	1	1	1	1	1	1	1	1	1	1	1,014	1	1,003	1
FRA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
GBR	1	1	1	1	1	1	1	1	1	1	1	1	1	1,015	1,02	1,022
GRC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
HUN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
IRL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ISR	1,002	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ITA	1,001	1,001	1,003	1,005	1	1	1,002	1	1,002	1,004	1	1,001	1	1	1	1
JPN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KOR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LVA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MEX	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NLD	1,014	1,018	1,021	1,021	1,017	1,019	1,017	1,016	1,014	1,012	1,012	1,017	1,018	1,019	1,015	1,016
NOR	1,006	1	1,013	1,01	1	1,01	1,009	1,013	1,008	1,011	1,006	1,01	1	1	1,003	1
NZL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
POL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PRT	1,032	1	1	1	1	1	1	1	1	1	1	1	1	1	1,002	1
SVK	1	1	1	1	1	1	1	1,023	1,03	1	1,022	1,007	1	1	1	1
SVN	1,034	1,01	1,01	1,024	1	1,035	1,004	1	1	1	1	1	1	1,022	1	1
SWE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1,003
USA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Note that eight countries are identified as efficient in both models: Australia, Canada, Chile, Spain, Iceland, South Korea, Mexico and New Zealand. In Figure 7.1, the technical efficiency for Belgium between the years 2000 and 2015 for each model is outlined in a linear graph. In the model MC, the technical efficiency remains between an interval from 1,01 to 1,029 whereas the interval varies from 1 to 1,027 in the second model.

7.2 Sources of inefficiency

So far, we have looked at the efficiency scores. In the two following sections, the black box is opened to get a better understanding of what lies behind these scores. The slack variables represent the amount of inputs to reduce or, likewise, the outputs to increase to become efficient. The computation for the slacks is described in section 4.5. As we have computed the technical efficiency of 2015, the same year is analysed for the slack variables. The results are presented in Table 7.2.

In the model MC, it appears that the ISR and PHS outcomes could be seriously improved. Having less hospitals beds could also improve the efficiency score, but in a less significant extend. From a policy viewpoint, it may be necessary to clarify the ambiguous situation where there is a null slack for the LE and 1,14 for the PHS. The perception of health is influenced by different parameters (with some overlapping) than the LE. Several factors must be considered when improving the PHS. In the scope of PHS, multiple elements are at play as described in section 5.1, e.g. the accessibility, quality standards, equality in healthcare, etc.

Table 7.2 Input and output slacks for Belgium in 2015 for the models MC and LF

	Model MC		Model LF	
	Slack variable	Value	Slack variable	Value
Input slacks	Health expenditure	0	Health expenditure	243,640
	Doctors	0	Tobacco consumption	1,95e-06
	Hospitals beds	0,0019	Alcohol consumption	0
			Fruits & Vegetables supply	12,068
Output slacks	LE	0	LE	0
	ISR	41,066	ISR	0
	PHS	1,14	PHS	0

The lifestyle model results show that three inputs are excessive, namely the health expenditure, tobacco consumption and the fruits and vegetables supply. Note that the tobacco variable has

been inverted and the excess value remains significant as it corresponds to approximately 500.000 Belgian smokers less.

In Figure 7.2, the efficiency frontier is plotted using as input the hospital beds, and ISR as output for the model MC in 2015. Belgium turns out to be quite far from the technology frontier as expected from the slacks. Besides, each country is represented in a colour associated to a region of the world. It is interesting to note that countries of the same region or more likely to end up closer one from the other, or having similar level of a certain input or output.

For instance, Northern European countries are near the efficiency border on the top left corner. Countries from South and Central America such as Mexico and Chile can be found on the bottom left of the graph. Close to the efficiency frontier, they present low ISR and hospital beds values. Regarding the Western European countries, the same level of ISR is shared among them (around the values 300-400) while it shows dissimilarities in the number of hospital beds.

In Figure 7.3, the same reasoning and analyse can be applied. For example, the gap between Western and Eastern European countries regarding the health expenditure is clear to see (when looking at the graph, keep in mind that the units are expressed in US dollars). Likewise, one can observe Southern European countries spending less than Western European countries while achieving approximately the same level of ISR. As in the previous figure, Mexico and Chile position themselves in the bottom left part of the graph, close to the efficiency frontier. This type of exercise can be performed on any dimension of inputs and outputs to have a better grasp of the characteristics of national HC systems.

7.3 Output sensitivity analysis

In Figure 7.4, a sensitivity analysis is conducted for each output variable in both models for Belgium in 2015. The values are enhanced up to the point where an efficiency score of 1 is reached. The efficiency scores resulting from the observed data of Belgium in 2015 is highlighted by a red point. Together, the present findings revealed similar patterns among the two models for each output variable. The LE variable is gradually decreasing by converging to an efficiency score of 1. Here, ceteris paribus, a LE of 83 years guarantees an efficiency score of 1 for Belgium. Concerning the ISR and PHS variables, a gentle decrease is to be seen in the beginning, followed by a substantial drop that ends up in a perfect efficiency score.

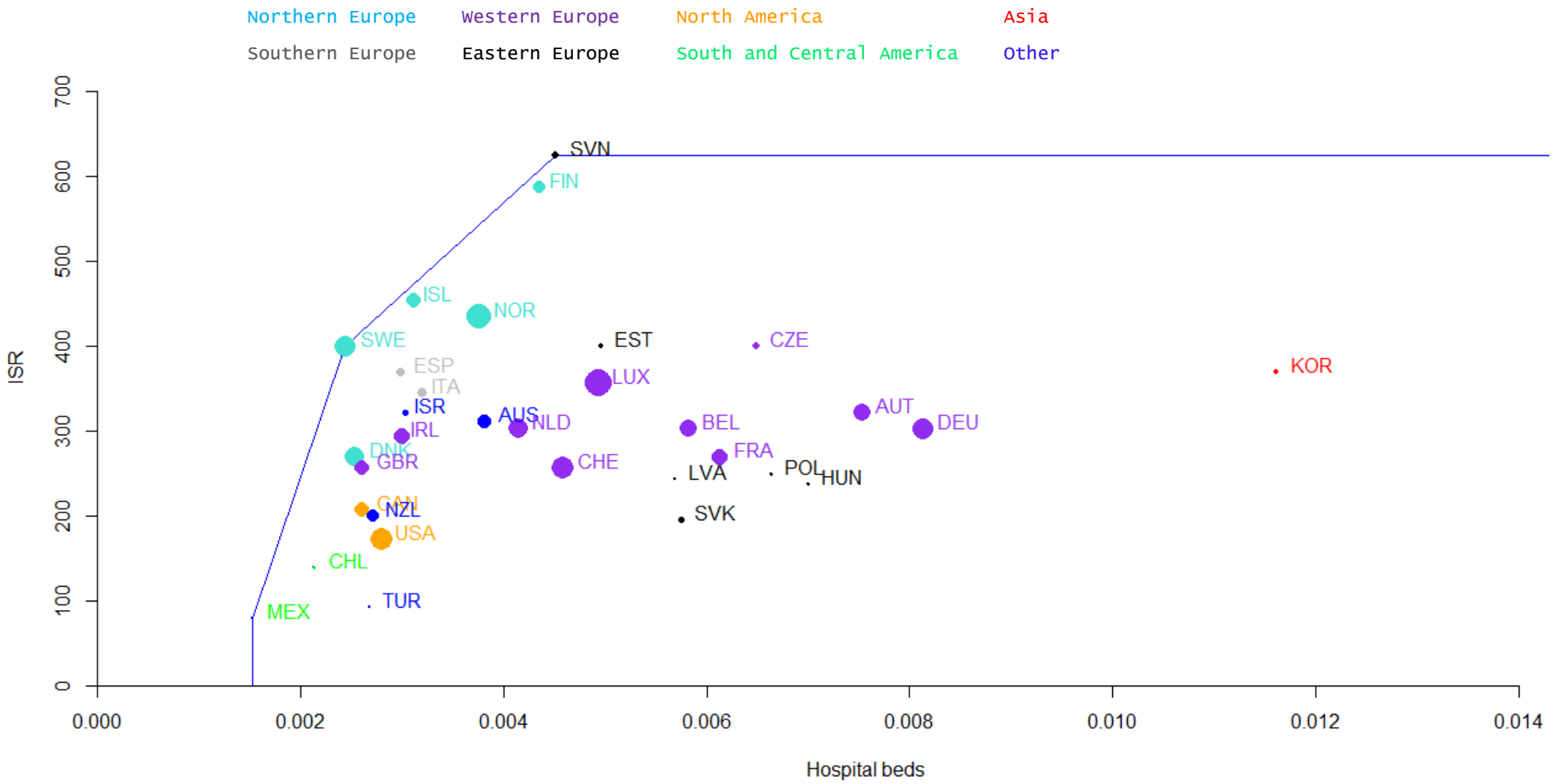


Figure 7.2 DEA technology for hospital beds (input) and ISR (output) with point size proportional to the healthcare expenses for model MC

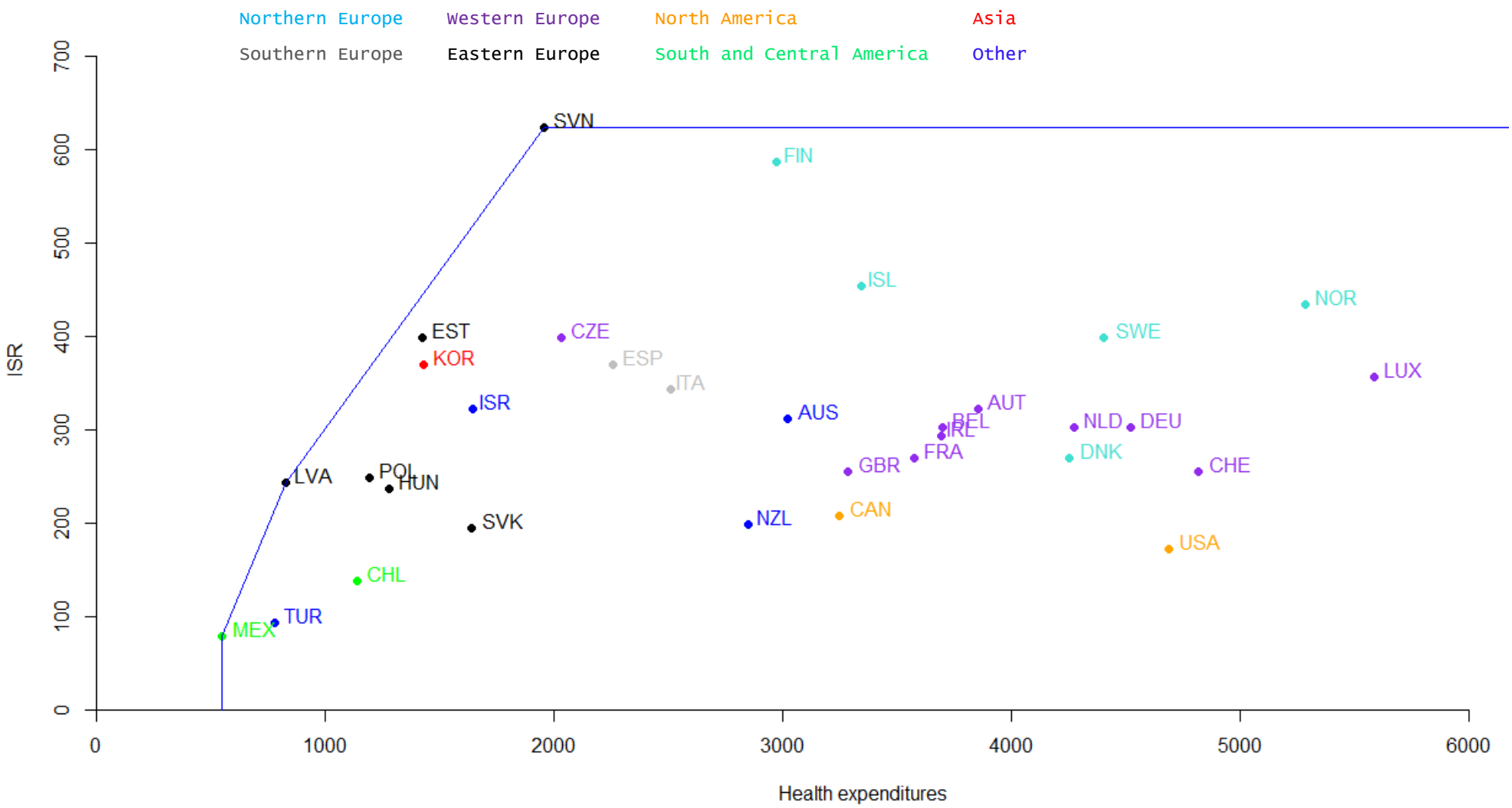


Figure 7.3 DEA technology for health expenditure (input) and ISR (output) for model MC

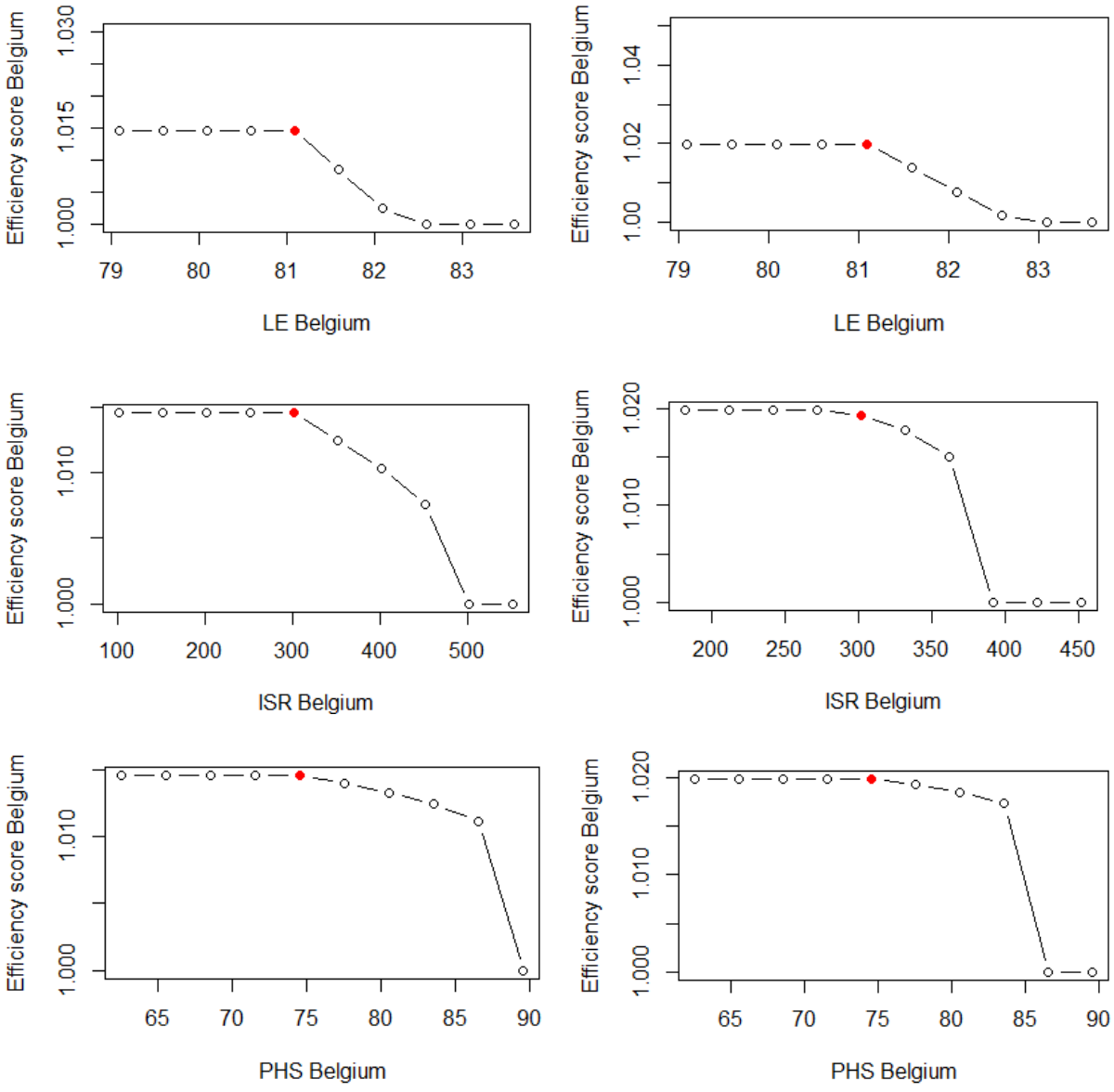


Figure 7.4 Sensitivity analysis for (a) the model MC and (b) the model LF for Belgium in 2015

Diminishing output values does not appear to lead to compelling worst scores. This entails the idea that Belgium has already low relative efficiencies and cannot become significantly worse.

7.4 Best practices for Belgium

The efficiency scores discussed above resulted from a comparison between a limited number of DMUs. The DMUs compared to compute the efficiency score are called *peers* and have an associated lambda which represents a kind of weight, enabling the expansion of observed technology to a weighted average one (see section 2.4 about convexity).

The resulting Belgian peers and their weights (lambdas) are shown in Table 7.3 for both models. In the model MC, as can be seen, Belgium has 5 peers. The two most influential DMUs (highest

lambdas) are Canada and Luxembourg, followed by Spain, South Korea and, lastly, Australia as the less significant peer. In a same way, the model LF has 4 peers, namely Australia, Switzerland, France and Japan where Australia plays the largest part in the score of Belgium.

With these results available, the Belgian decision makers have now valuable information at their disposable. The next step would be to examine what makes those countries more efficient than Belgium and how their HC system works. As different peers are identified in the models, each analyse can concentrate on the specific division of a HC system that is illustrated in each model. In other words, the peers of the model MC should be used for the medical care part of a HC system while the peers in the model LF should be analysed for the way of living and consumption habits that relates to the population's health.

Table 7.3 Peers of Belgium and the associated lambdas for the year 2015

Model MC		Model LF	
Peers	Lambdas	Peers	Lambdas
AUS	0,005	AUS	0,540
CAN	0,359	CHE	0,135
ESP	0,246	FRA	0,239
KOR	0,052	JPN	0,086
LUX	0,338		

From the weights (lambdas) above, specific targets for Belgium are computed in the table below using weighted average. Realistic and attainable targets are obtained for both modes and are displayed in Table 7.4.

Table 7.4 Output targets of Belgium based on peers for the year 2015

Output variable	BEL in 2015	Target model MC	Target Model LF
LE	81,1	82,28	82,66
ISR	302,03	306,44	307,91
PHS	74,6	75,67	76,02

7.5 Ranking

To rank the efficiency of HC systems among the DMUs, the super-efficiency described in section 2.7 is computed. As a reminder, the idea is to assign an efficiency score to the DMUs while removing the target DMU from the potential peer set. The super-efficiency may have a value larger than 1, which allow us to rank inefficient and efficient DMUs.

In Figure 7.5, the ranking derived from the super-efficiency of Belgium is plotted for both model MC and LF during the years 2000 to 2015. In the model MC, it can be seen that the

Belgian medical care system is a bad performer relatively to the other members of the OECD. The Belgian rank remains around the 25th position. Le lifestyle model shows slightly worst results for Belgium. Here, the ranking remains around the 30th position in the ranking.

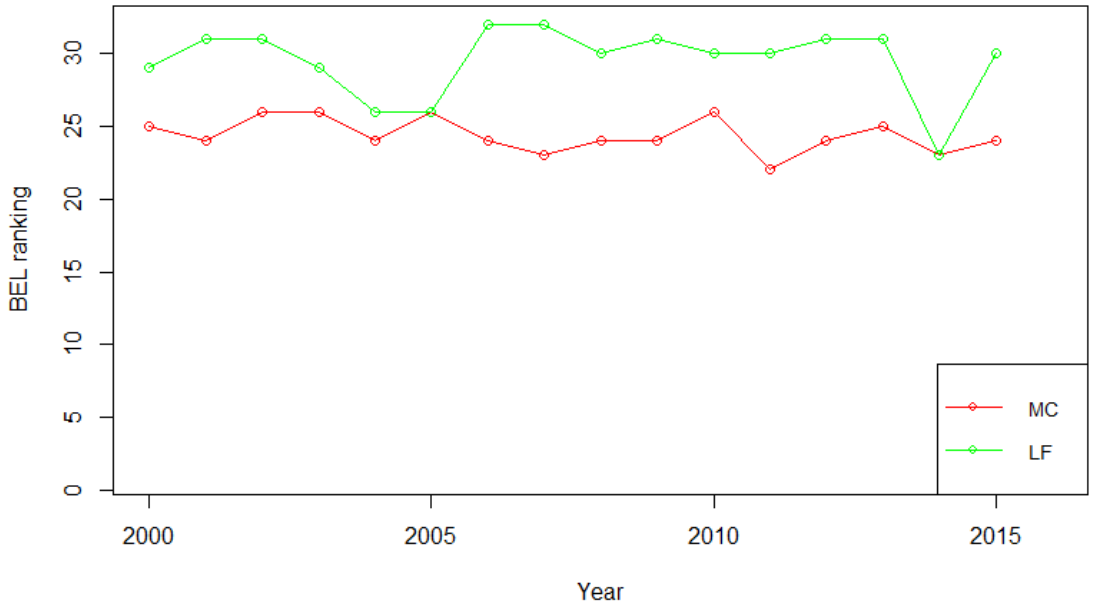


Figure 7.5 Ranking (super-efficiency) of Belgium for the model MC and the model LF between the years 2000 and 2015

7.6 Malmquist Index

The method behind the results computed in this section is described in section 5.1. For the sake of clarity, the means are computed per country to avoid being overloaded with information. If the mean index equal 1, no change happened between the different periods on average. If it is higher than 1, than the HC systems of the country improved. Likewise, it decreased if the value is lower than 1. A significativity test (Wilcoxon test) is performed to asses if the value is significantly different from the others with a type I error rate of 0,05. Significantly different values are denoted with an asterisk.

The second columns of Table 7.5 and 7.6 present the Malmquist productivity index for the model MC and LF respectively. For the model MC, Ireland presents the best results with an average MPI of 1,06. On the other extreme, Latvia appears as having worst index showing a MPI of 0,95. In between lies Belgium that experienced a slight increase in its HC performance having a MPI value of 1,02. The effect of this value is to be seen in Figure 7.1 where the technical radial output efficiency shows a declining trend (and is thus improving as getting closer to 1). When looking at the results for the second model, Latvia is again showing poor results along with Great Britain whereas Slovenia's MPI is the highest. On average, the Belgian

HC systems does not undergo any changes in the model LF. Figure 7.1 outlines the Lifestyle model that appears at the same level on average, confirming the Malmquist index results.

As explained before, the Malmquist index can be decomposed into two subcomponents. One of these is the efficiency change (catch-up effect) that tells us if a DMU is closer to the efficiency frontier. The second measure captures the changes in the technology between two periods and is known as the technology change. These measures are computed, and the results can be found in the columns 3 and 4 of Table 7.5 and Table 7.6.

Both measures (efficiency and technology changes) can be used to explain the Malmquist index results. If we reconsider our remark about the low value of Latvia in the model MC, the corresponding efficiency change remains the same as it equals 1 while the technology change has a value of 0,95. Hence, one can conclude that the low Malmquist value of the Latvian HC system is due to a change in the technology. The catch-up effect (efficiency change) is not high enough to compensate the change in the technology. This reasoning can be applied to any other DMU and may be valuable to assess previous health policies or decisions. For Belgium, in the model MC, the MPI resulted both from an improvement in the efficiency and technological changes as shown in Table 7.5.

Table 7.5 Malmquist productivity index for the model MC

Country	MPI	Efficiency change	Technology change
AUS	1	0,99	1,01
AUT	1	1	1
BEL	1,02	1,01	1,01
CAN	1,01	1	1,01
CHE	1,02	1	1,02
CHL	0,98	1	0,98
CZE	0,99	1,02	0,97
DEU	1	1,01	1
DNK	1,04	1,02	1,02
ESP	1,01	1	1,01
EST	1	1,02	0,98
FIN	1,01	1,01	1
FRA	1,02	1,01	1,02
GBR	1,03	1,01	1,02
HUN	0,97	1,03	0,94
IRL	1,06*****	1,04*****	1,03
ISL	1,05	1	1,05*****
ISR	1,01	1,01	1
ITA	1,02	1,01	1,02
KOR	0,96	1	0,96
LUX	1,03	1,02	1,01
LVA	0,95*****	1	0,94
MEX	0,98	1	0,98
NLD	1,01	1	1,01
NOR	1,02	1	1,03
NZL	1	0,99	1,01
POL	0,96	1,01	0,95
SVK	1,03	1,02	1,04
SVN	1,03	1,02	1,02
SWE	1,04	1	1,04
TUR	0,98	1	0,98
USA	1,01	1	1,01

***** Significantly different from the other means according to the Wilcoxon Test (non-parametric unpaired two-sample test) with α = 5%.

Table 7.6 Malmquist productivity index for model LF

Country	MPI	Efficiency change	Technology change
AUS	1	1	1
AUT	1	1	0,99
BEL	1	1	1
CAN	0,99	0,99	0,98
CHE	0,99	0,99	0,99
CHL	0,99	0,99	0,99
CZE	1	1	1
DEU	1	1	1
DNK	0,98	0,98	0,99
ESP	1	1	1
EST	1	1	1
FIN	1,02	1,02	1,02
FRA	1	1	1
GBR	0,97	0,97	0,98
GRC	0,98	0,98	0,98
HUN	0,99	0,99	0,99
IRL	0,98	0,98	0,98
ISR	1,01	1,01	0,98
ITA	0,99	0,99	0,98
JPN	1	1	1
KOR	0,98	0,98	0,98
LVA	0,97	0,97	0,97*****
MEX	0,98	0,98	0,98
NLD	0,99	0,99	1
NOR	0,99	0,99	1,01
NZL	0,98	0,98	0,98
POL	0,99	0,99	0,99
PRT	0,99	0,99	0,99
SVK	0,99	0,99	0,99
SVN	1,04*****	1,04*****	1,03*****
SWE	1	1	1,01
USA	0,99	0,99	0,99

***** Significantly different from the other means according to the Wilcoxon Test (non-parametric unpaired two-sample test) with α = 5%.

8. CONCLUSION

In this study, the HC systems of the OECD members are assessed between the years 2000 and 2015 by two output-oriented DEA models using variable returns to scale, each representing a specific subdivision of a HC system. One is referred as the Medical Care model (MC) and the other as the Lifestyle model (LF) with appropriate input and output variables. Three countries are removed in both models as they are identified as outliers. Australia, Canada, Chile, Spain, Iceland, South Korea, Mexico and New Zealand have HC systems that are efficient during the whole period in both models.

Belgium seems to be a poor performer in this evaluation compared to the others. In both models, the country ends up in the top 10 of worst HC systems in every period based on a super-efficiency ranking. The countries from which Belgium should look at when trying to improve its efficiency have been highlighted, where the most important peers include Australia, Canada and Luxembourg.

In our analysis, the Malmquist productivity index is computed. It evaluates the changes of performances over time due to technology or efficiency changes. Ireland experienced the most significant improvement during the time interval whereas Latvia has the most decreasing HC system efficiency overtime. The changes in efficiency and technology are also presented in a way that allows to understand part of the Malmquist results. The Belgian HC system shows a slight increase in the model MC over the years while the index for the model LF indicates that no changes occurred on average.

Furthermore, it appears that rich countries like Belgium, Austria, the Netherlands, Norway or Denmark have inefficient HC systems, even though their HC is known to be good compared to other efficient countries such as Mexico, Turkey or Chile. That being said, efficiency may not be an intrinsic goal for governments of rich countries per se. Indeed, one can suppose that they emphasize on the value delivered by the HC while evaluating the associated costs. These points should need further research and go beyond the scope of our study.

Future interesting research could include more quality aspects into the modelling, and thus the variables. The “Health Care Quality Indicators” is a project developed by the OECD that aims to construct reliable and objective indicators of quality in HC (for different types of care). Unfortunately, so far, only spare datasets are available. We can hope that the OECD will continue collecting data and developing the project.

REFERENCES

Afonso, A. et al. (2003). Public Sector Efficiency: an International Comparison. *European Central Bank, working paper 242*.

Afonso, A. et al. (2006). Public Sector Efficiency: Evidence for New EU Member States and Emerging Markets. *European Central Bank, working paper 581*.

Afonso, A., and St-Aubyn, M. (2005). Non-parametric approaches to education and health efficiency in OECD countries. *Journal of Applied Economics, 8*(2), 227-246.

Agrell, P. J. and Niknazar, P. (2014). Structural and behavioral robustness in applied best-practice regulation. *Socio-Economic Planning Science, 48*, 89-103.

Al-Refaie, A. et al. (2014). Applying simulation and DEA to improve performance of emergency department in a Jordanian hospital. *Simulation Modelling Practice and Theory, 41*, 59-72.

Andersen, P., and Peterson, N. (1993). A Procedure for Ranking Efficient Units in Data Envelopment Analysis. *Management Science, 39*(10), 1261-1264.

Asmild, M. et al. (2004). Combining DEA Window Analysis with the Malmquist Index Approach in a Study of the Canadian Banking Industry. *Journal of Productivity Analysis, 21*, 67-89.

Asmild, M., et al. (2003). Measuring Inefficiencies via Potential Improvements. *Journal of Productivity Analysis, 19*, 59-76.

Banker, R., Chang, H., Zheng, Z. (2017). On the use of super-efficiency procedure for ranking efficient units and identify outliers. *Annals of Operations Research, 250*(1), 21-35.

Banker, R., Charnes, A., & Cooper, W.W. (1984). Some Models for Estimating Technical and Scale Inefficiencies using Data Envelopment Analysis. *Management Science, 30*(9), 1078-1092.

Banker, R., Charnes, A., Cooper, W., Starts J. & Thomas D. (1989). An introduction to data envelopment analysis with some of its models and their uses. *Research in Government and Non-Profit Accounting, 5*, 125-163.

Banker, R. (1996). Hypothesis testing using Data Envelopment Analysis. *Journal of Productivity Analysis, 7*(2), 139-159.

Bloom, D. et al. (2003). The Demographic Dividend - A New Perspective on the Economic Consequences of Population Change. *Research and Development Corporation.*

Bogetoft, P. & Otto, L. (2011). *Benchmarking with DEA, SFA, and R*. London: Springer. doi:10.1007/978-1-4419-7961-2.

Bogetoft, P., and Hougaard, J. S. (1999). Efficiency Evaluation Based on Potential (Non-Proportional) Improvements). *Journal of Productivity Analysis, 12*, 233-247.

Boussofiane, A., Dyson, R. G. and Thanassoulis, E. (1991). Applied Data Envelopment Analysis. *European Journal of Operation Research, 52*(1), 1-15.

Cetina, V. R. and Bahceb, S. (2016). Measuring the efficiency of health systems of OECD countries by data envelopment analysis. *Journal of Applied Economics, 48*(37), 3497-3507.

Chambers, R. (1988). *Applied Production Analysis: A Dual Approach*. Cambridge University Press.

Chambers, R. G., Chung, Y., and Färe, R. (1998). Profit, Directional Distance Functions, and Nerlovian Efficiency. *Journal of Optimization Theory and Applications, 98*(2), 351-364.

Charnes, A., and Cooper, W. W. (1962). Programming with liner fractional functions. *Naval Research Logistic Quarterly banner, 9*, 181-186.

Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operation Research, 2*(6), 429-444.

Cook, W., & Seiford, L. (2009). Data Envelopment Analysis (DEA) – Thirty years on. *European Journal of Operation Research, 192*, 1-17.

Cook, W., Tone, K., & Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *International Journal of Management Science, 44*, 1-4.

Culliname, K. and Wang, T. (2010). The efficiency analysis of container port production using DEA panel data approaches. *Operation Research Spectrum,* 32, 717-738. doi: 10.1007/s00291-010-0202-7

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

Dyson, R. G. et al. (2001). Pitfalls and protocols in DEA. *European Journal of Operation Research, 132*(2), 245-259.

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

Evans et al. (2000). The comparative efficiency of national health systems in producing health: an analysis of 191 countries. *WHO publishing.*

Färe, R. et al. (1992). Productivity Changes in Swedish Pharamacies 1980-1989: A Non-Parametric Malmquist Approach. *Journal of Productivity Analysis, 3*, 85-101.

Färe, R., and Primont, D. (2012). Dual allocative efficiency parameters. *Journal of Productivity Analysis, 37*, 233-238.

Farrell, M. J. (1957). The measurement of Productive Efficiency. *Journal of the Royal Statistical Society, 120*(3), 253-290.

Golany, B. and Roll, Y. (1989). An application procedure for DEA. *Omega, 17*(3), 237-250. doi: 10.1016/0305-0483(89)90029-7

Grosskopf, S. and Valdmanis, V. (1987). Measuring Hospital Performance. *Journal of Health Economics, 6*, 89-107.

Gupta, S. and Verhoeven, M. (2001). The efficiency of government expenditure: Experiences from Africa. *Journal of Policy Modelling, 23*, 433-467.

Hadad, S., Hadad, Y. and Simon-Tuval, T. (2013). Determinants of healthcare system's efficiency in OECD countries. *European Journal of Health Economics, 14*(2), 253-265.

Joskin, A. (2017). Mesurer le bien-être en Belgique - Construction d'un indicateur composite pour mesurer le bien-être actuel des Belges. *Bureau fédéral du Plan.*

Joumard, I. et al. (2008). Health Status Determinants: Lifestyle, Environment, Health Care Resources and Efficiency. *OECD Publishing, 627*. doi: 10.1787/240858500130

Koopmans, T. (1951). Activity Analysis of Production and Allocation. *John Wiley and Sons Inc*., 33-97.

Lefebvre, M. et al. (2015). Productivity and performance in the public sector. *Bureau d'économie théorique et appliquée.*

Liu, J., Lu, L., Lu, W., & Lin, B. (2013). Data envelopment analysis 1978 –2010: A citation-based literature survey. *International Journal of Management Science, 41*, 3-15.

Malmquist, S. (1953). Index Numbers and Indifference Surfaces. *Trabajos de Estatistica, 4*, 209-242.

Mathers, D. et al. (2000). Estimates of DALE for 191 countries: methods and results. *WHO publishing.*

McFadden, D. (1978). *Cost, Revenue, and Profit Functions*. University of Berkley .

Mohan, R. and Mirmirani, S. (2008). An Assessment of OECD Health Care System Using Panel Data Analysis. *Southwest Business and Economics Journal*.

Murray, C., and Frank, J. (1999). A WHO Framework for Health System Performance Assessment.

Nistor, C. (2017). Performance through efficiency in the public health system – a DEA approach in an emergent country. *Studia Universitatis Babes-Bolyai Oeconomica, 62*(1), 31-49. doi: 10.1515/subboec-2017-0003

Norman, M. and Stoke, B. (1991). *Data Envelopment Analysis: the Assessment of Performance.* New York: Wiley.

OECD (2002). OECD Health Data 2002: a Comparative Analysis of 30 Countries. *OECD publishing*.

OECD (2011). Health at a glance 2011: OECD indicators. *OECD Publishing.*

OECD (2013). Health at a Glance 2013: OECD indicators. *OECD publishing.*

OECD (2018). OECD Health Statistics 2018: Definitions, Sources and Methods. *OECD publishing.*

Olesen, O., Petersen, N. and Podinovski, V. (2015). Efficiency analysis with ratio measures. *European Journal of Operation Research, 245*, 446-462.

Ozcan, Y. (2008). *Health Care Benchmarking and Performance Evaluation: An Assessment Using Data Envelopment Analysis*. International Series in Operation Research & Management Science. London: Springer. doi: 10.1007/978-1-4899-7472-3

Ozcan, Y. and Khushalani, J. (2016). Assessing efficiency of public health and medical care provision in OECD countries after a decade of reform. *Central European Journal of Operation Research, 25*, 325-343.

Phillips, F. (2005). 25 Years of Data Envelopment Analysis. *International Journal of Information Technology and Decision Making, 4*(3), 317-323.

Rasmussen, S. (2011). *Production Economics: The Basic Theory of Production Optimisation.* London: Springer. doi: 10.1007/978-3-642-14610-7

Rothgang, H. et al. (2005). The changing role of the state in healthcare systems. *European Review, 13*(1), 187-212.

Rumbold, B. et al. (2015). Improving productive efficiency in hospitals: findings from a review of the international evidence. *Cambridge University Press, 10*, 21-43. doi:10.1017/S174413311400022X

Samut, P. and Cafri, R. (2016). Analysis of the Efficiency Determinants of Health Systems in OECD Countries by DEA and Panel Tobit. *Social Indicators Research, 129*, 113-132.

Seiford, L. (1996). Data Envelopment Analysis: The Evolution of the State of the Art (1978-1995). *Journal of Productivity Analysis, 7*, 99-137.

Shepard, R. W. (1953). Cost and production functions. *Princeton University Press*.

Shepard, R. W. (1970). Theory of Cost and Production Function. *Princeton University Press*.

Spinks, J. and Hollingsworth, B. (2009). Cross-country comparisons of technical efficiency of health production: a demonstration of pitfalls. *Journal of Applied Economics, 42*, 417-427.

Stavárek, D. (2003). Banking Efficiency in Visegrad Countries Before Joining the European Union. *Prepared for the Workshop on Efficiency of Financial Management Institutions and European Integration at the Technical University of Lisbon*

Thompson, R., Langemeier, L., Lee, C.-T., Lee, E., & Thrall, R. (1990). The role of multiplier bounds in efficiency analysis with application to Kansas farming. *Journal of Economics, 46*, 93-108.

Tulkens, H. (1993). On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts, and Urban Transit. *Journal of Productivity Analysis, 4*, 183-200.

Varabyova, Y. and Schreyögg, J. (2013). International comparisons of the technical efficiency of the hospital sector: Panel data analysis of OECD countries using parametric and non-parametric approaches. *Health Policy, 112*, 70-79.

WHO (2000). The World health report 2000: health systems: improving performance. *WHO publishing*.

Yaya, S. and Danhoundo, G. (2015). Special Issue on Innovations in Health Care System Reform in. *Innovation Journal, 20*(1).

Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve, Belgique **www.uclouvain.be/lsm**

UCL
Université
catholique
de Louvain