Monitoring and forecasting hospital performance in Portugal

Sofia Ambrósio
sofiaambrosio@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

December 2020

Abstract

A great deal of resources is applied to the health sector in Portugal, with a considerable part going to public hospitals. Therefore, a lot of attention has been drawn to public hospitals' efficiency and productivity analysis. This work aims to assess and predict the performance of Portuguese public hospitals. Data Envelopment Analysis is used to calculate hospital efficiency, and the Malmquist Productivity Index to evaluate hospital productivity. A sample of 26 public hospitals and hospital centers with data from 2013 to 2017 was used for this analysis. The Malmquist Productivity Index was forecasted for 2018 and then compared, for some hospitals, with the real values. The performance of hospitals has been slowly increasing, with overall average DEA score considering CRS being 0.648 and under VRS 0.764. Hospital efficiency seems to be increasing throughout the years, as well as scale efficiency. The overall average MPI is 1.049, suggesting productivity increase but not presenting a clear trend. The terms regarding changes in technology seem to influence more the MPI than the ones considering efficiency changes. The forecasted MPI suggested a very small increase for the year 2018 but the forecast did not seem to present reliable enough results.

Keywords: Efficiency, Productivity, Data Envelopment Analysis, Malmquist Productivity Index, Portuguese public hospitals

1. Introduction

Health is one of the most powerful factors in social integration, but also in generating wealth and well-being. New demographics, socioeconomic conditions and technological progress generate the need for the health sector to adapt to new problems that are presented, different than the ones existing before, putting a great amount of pressure on the health sector. Portugal spent around 9.4% of the GDP in health expenditure in 2018 (provisory value) and 9.4% in 2019 (preliminary value).[1] However, data are provisional for 2018 and preliminary for 2019. So, the most recent final and validated data is from 2017. In this year, 9.3% of the GDP was spent on health, and 30% of the health expenditure in Portugal went to public hospitals.[1] These values draw attention to the efficiency of the management of health systems and, in particular, public hospitals. Hospitals are crucial components of a health system, offering specialized health care that cannot be provided in other settings. However, this also means that hospitals are expensive to operate with a high number of staff, equipment and other operating costs.[2]

Health care services are mostly provided by public institution, where health care is not seen as an area to obtain profits, and is seen as priceless. Doctors and nurses and other health care providing workers aim to maximize the patients' well-being, and not optimize profits or resource utilization (Prezerakos, Maniadakis, Kaitelidou, Kotsopoulos, & Yfantopoulos, 2007). This, allied with all the changing factors stated before, may lead to health care institutions being often thought of presenting inefficiency and low productivity (Prezerakos et al., 2007). Efficiency assessment can be a useful tool for health planning and evaluation of policies, being of interest to a range of people, from the general public to hospital managers and governmental policymakers (Peacock, Chris, Melvino, & Johansen, 2001). In this context, this work will assess the performance of hospitals,

[1] Conta Satélite da Saúde 2020. Instituto Nacional de Estatística - INE (Available at: www.gee.gov.pt/pt/indicadores-diarios/ultimos-indicadores/30399-ine-conta-satelite-da-saude). Accessed on: 10/12/2020

[2] OECD (2017), Tackling Wasteful Spending on Health, OECD Publishing, Paris (Availabe at: www.oecd.org/health/tackling-wasteful-spending-on-health-9789264266414-en.htm. Accessed on: 3/12/2020

making use of Data Envelopment Analysis (DEA) and Malmquist Productivity Index (MPI). Moreover, the MPI will be used to forecast hospital performance.

2. The Portuguese NHS

The Portuguese public health system's activity is mainly characterized by a National Health Service (NHS) - *"Serviço Nacional de Saúde" (SNS)* - that follows a Beveridge system (D. C. Ferreira & Marques, 2019). Additionally, the state maintains agreements with the private and social sectors to complement the health care provision (Nunes & Ferreira, 2019). Also, there are health subsystems (health insurance schemes associated with professional or occupational sectors), and private insurance schemes (Simões, Augusto, Fronteira, & Hernandez-Quevedo, 2017).

The NHS includes health promotion and surveillance, disease prevention, diagnosis and treatment of patients and medical and social rehabilitation. It holds administrative and financial autonomy and is structured in a decentralized and deconcentrated organization, comprising central, regional and local bodies.[3]

Public health care services are under the authority of the Ministry of Health, which is responsible for the development of health policies and the supervision and evaluation of their implementation (Simões et al., 2017). The management, planning and regulation of the NHS are carried out centrally by the Ministry of Health and its institutions. Thus, hospitals are not autonomous in a number of issues, such as the purchase of innovative new technologies, or the hiring of personnel (doctors, for example).

In the Portuguese NHS, four levels of care can be differentiated: (1) primary health care, in health care centers (2) secondary care, in hospital units, (3) post hospital care, involved in rehabilitation processes, and (4) palliative care for end-of-life cases (D. C. Ferreira & Nunes, 2019).

Health care management is decentralized in Regional Health Administrations (RHA). Each RHA is responsible for the regional implementation of national health policies and coordination of all levels of health care. As well as the coordination of all aspects of health care provision, supervision of the hospitals and health centres' management, and articulation of agreements with the private and social sectors, and municipal councils, in its geographical area and for its population (Simões et al., 2017).

Secondary healthcare is provided by public hospitals that are uniformly distributed across the country, according to the resident population, its health needs, and the existence of medical professionals (Nunes & Ferreira, 2019). As well as singular hospitals, secondary heath care also includes hospital centers (horizontal merging), local health units (vertical merging of a singular hospital and primary health care centres), hospitals in Public-Private Partnerships (PPPs), oncology centers (IPOs), maternity hospitals (which provide specialized Obstetrics, Gynaecology, and Paediatrics), and psychiatric hospitals. These last three represent specialized hospitals, requiring specialized physicians who may only serve in these specific specialties (D. C. Ferreira & Marques, 2020; D. C. Ferreira, Marques, & Nunes, 2018).

There are, currently, five RHAs in the country. There were, in 2017, a total of 225 health facilities, with 100,147 people as hospital staff, and 34,953 hospital beds.[4] Private and public hospitals represented 51.7% and 46.5%, respectively, of total hospitals, with PPPs representing 1.7%.[4]

In 2005, hospitals were transformed into corporate public entities (EPE, which stands for *"Entidade Pública Empresarial"*). This new management scheme incorporated management efficiency and user satisfaction.[5] Since 2005, the number of EPE hospitals increased and there are nowadays 47 health centers and 53 hospitals and hospital centers from which 41 are EPEs, six belong to the Public Administrative Sector, and from the remaining six, three are Public-Private Partneships (PPP) and the other three are managed by the *Misericórdias* (social sector).[3] The autonomy of an EPE hospital is lower, given that Ministries must approve their activity reports and budgets and deal with the most important issues (D. Ferreira & Marques, 2015).

Between 2011 and 2015, a period characterized by the economic and financial crisis, which was followed by the post-crisis recovery period (Nunes & Ferreira, 2019).

3. Literature Review

A total of 23 studies, both from Portugal and the world, were identified and analyzed in order to characterize similar already existing studies, assess the most used variables and methodologies and identify possible literature gaps. Regarding the methodologies used, 16 out of the 23 studies employed Data Envelopment Analysis (DEA) or DEA based methods, while six studies additionally used

[3] SNS - Portal SNS (Available at: www.sns.gov.pt/). Accessed on: 15/04/2020

[4] PORDATA (Available at: www.pordata.pt/). Accessed on: 30/04/2020

[5] Franca, L., Monte, A. P. (2010). Comparação entre sistemas de gestão hospitalar: SPA, SA e EPE, na perspectiva do planeamento e controlo orçamental: um estudo de caso. XIV Congreso Internacional de la Academia de Ciencias Administrativas, Monterrey. (Available at: hdl.handle.net/10198/2541). Accessed on: 30/04/2020

the Malmquist Productivity Index (MPI) to assess productivity. Other common methods used were Order-α, Free Disposal Hull (FDH) and Order-m.

Studies recurred mostly to inputs and outputs (82.6% of the studies used in/output variables), with some considering environment or exogenous variables. The average number of variables used by authors is nine, with 39 being the maximum and the minimum only four. In terms of inputs, costs are the most commonly adopted, mainly costs of goods sold and consumed, costs with supplies and external services, staff costs, and operating costs. The number of nurses, doctors and other staff (such as sanitary workers or administrative and support staff) are also regularly employed as inputs, either in total numbers of these worker types, number per patient or inhabitant, or full time equivalent (FTE). The number of inpatient discharges and the total number of medical appointments or outpatient visits are the most commonly used outputs. Other most used variables are emergency cases, ambulatory surgeries and the number of patients treated.

4. Methodology

4.1. Data Envelopment Analysis

DEA is a benchmarking technique (Ji & Lee, 2010) and linear programming method used to examine the relationship between inputs and outputs of each Decision Making Unit's (DMU) production process from observed data, comparing the result with the best practice frontier (Büchner, Hinz, & Schreyögg, 2016). It is a non-parametric, non-statistical, deterministic method that determines a frontier "enveloping" the observations (Bezat, 2009). In a DEA model, the efficiency of a DMU is defined as the ratio of the sum of its weighted outputs (for example, number of patients treated) to the sum of its weighted inputs (for example, resources used in a hospital) (Weng, Wu, Blackhurst, & Mackulak, 2009). It simultaneously identifies the optimal input/output combination, depicted as the "best practice frontier" (Ersoy, Kavuncubasi, Ozcan, & Harris, 1997). This frontier represents the production technology of the most efficient entities, with DMUs belonging to it having an efficiency score of one and being benchmarks for the other, inefficient, entities, since they can deliver the same kind of services with a more efficient use of the available resources (D. C. Ferreira & Nunes, 2019). Accordingly, DMUs operating below the frontier are assigned a score inferior to one, but greater than zero, hence being capable to improve capacity and future performance (Ersoy et al., 1997; Ji & Lee, 2010).

The modeling can be input or output oriented, depending if the objective is the reduction of resources or production increase (Lins, Lobo, Moreira Da Silva, Fiszman, & Ribeiro, 2007). In the scope of this work, input orientation is assumed, since in hospitals there is little control of production (outputs), i.e. managers can control the inputs, such as number of hired staff or hospital costs, whereas outputs, for example number of patients treated, can be considered exogenous (Büchner et al., 2016). Moreover, DEA can be carried out based on constant returns to scale (CRS), meaning that the output will change by the same proportion as inputs are changed, or variable returns to scale (VRS), which reflects that production technology may increase, decrease or maintain returns to scale (Ji & Lee, 2010; Lins et al., 2007).

Mathematically, consider a set of $j = 1, 2, ..., n$ DMUs (hospitals, in this case) that transform a vector of $i = 1, 2, ..., m$ inputs into a vector of $r = 1, 2, ..., s$ outputs. Each hospital n is characterized by the vector (x_n, y_n) of inputs and outputs, with $x \in R^m_+$ and $y \in R^s_+$. Let (x^i_j, y^r_j) be the vector defining the DMU whose efficiency is being assessed and λ_n the weights regarding the outputs and inputs.

The input oriented efficiency of each DMU j is then calculated by solving the following linear programming problem n times:

$$\min \quad \theta_j \tag{1}$$

$$\text{subject to} \quad \sum_n \lambda_n x^i_n \leq \theta_j x^i_j \tag{2}$$

$$\sum_n \lambda_n y^r_n \geq y^r_j \tag{3}$$

$$\lambda \geq 0. \tag{4}$$

in the case of assuming CRS. If it is the case of VRS then another condition is needed (Jacobs, 2001):

$$\sum_{j=1}^{n} \lambda_j = 1. \tag{5}$$

Calculating the efficiency scores with DEA under both CRS and VRS, it is possible to evaluate the scale efficiency, by dividing the score under CRS for the one considering VRS (Kirigia & Asbu, 2013). The maximum scale efficiency score is one, which implies that the DMU considered is operating at its optimal scale or size. If the score is less than one, the unit is either too small or too big relative to the optimal size (Kirigia & Asbu, 2013).

DEA can be combined with the Malmquist Productivity Index in order to assess efficiency over a period of time.

4.2. Malmquist Productivity Index

The Malmquist Productivity Index is a bilateral index that compares the production technology of two economies, evaluating the efficiency change between two time periods (Tone, 2005), and can be defined as the ratio of two input distance functions (Simar & Wilson, 1999). The MPI can be decomposed into indices describing changes in technology and efficiency (Simar & Wilson, 1999), indicating progress or regress in efficiency as well as progress or regress of the technology frontier over time (Tone, 2005).

Let's consider, as before, a DMU that produces s outputs from m inputs, with $x \in R^m_+$ and $y \in R^s_+$ being input and output vectors, respectively, and a production possibilities set at time t denoted as $P = \{(x^t, y^t) | x^t$ can produce y at time $t\}$ (Simar & Wilson, 1999). Its upper boundary can be referred as the production technology or the production frontier (Simar & Wilson, 1998). Let (x^t_i, y^t_i) be the input and output vectors of production unit i at time t.

The MPI can, then, be defined as a geometrical mean of relative productivity changes from time t to time $t+1$ (Daskovska, Simar, & van Bellegem, 2010):

$$\Pi^{t,t+1} = \left(\frac{\theta^t_{CRS}(x^{t+1}, y^{t+1})}{\theta^t_{CRS}(x^t, y^t)} \cdot \frac{\theta^{t+1}_{CRS}(x^{t+1}, y^{t+1})}{\theta^{t+1}_{CRS}(x^t, y^t)} \right)^{1/2} = \\ = \left(\frac{\theta^{t+1}_{CRS}(x^{t+1}, y^{t+1})}{\theta^t_{CRS}(x^t, y^t)} \right) \cdot \\ \cdot \left(\frac{\theta^t_{CRS}(x^{t+1}, y^{t+1})}{\theta^{t+1}_{CRS}(x^{t+1}, y^{t+1})} \cdot \frac{\theta^t_{CRS}(x^t, y^t)}{\theta^{t+1}_{CRS}(x^t, y^t)} \right)^{1/2} = \\ = \Delta Eff^{t,t+1} \cdot \Delta Tech^{t,t+1} \tag{6}$$

in which the term $\Delta Eff^{t,t+1}$ "measures the change in relative efficiency (i.e., the change in how far observed production is from maximum potential production)" between times t and $t+1$, and the term $\Delta Tech^{t,t+1}$ "captures the shift in technology between the two periods" evaluated in the hyperplanes where the inputs for production unit i are maintained constant at times t and $t+1$ (Simar & Wilson, 1998).

4.3. Forecasting the Malmquist Productivity Index

In Daskovska et al. (2010), a new method for forecasting the MPI is introduced. In this method, a required condition is the circularity property of the index. Therefore, because the MPI is not circular, Daskovska et al. (2010) also propose a new decomposition of the index into circular components based on the one proposed by Simar and Wilson (1998).

After having defined P^t, now the set V^t is defined as the convex cone with vertex at the origin spanned by P^t, meaning $P^t \subseteq V^t$. If the technology exhibits CRS everywhere, then $P^t = V^t$.

So, both terms $\Delta Eff^{t,t+1}$ and $\Delta Tech^{t,t+1}$ (from Equation 6) can be further decomposed, and the MPI can be defined in a new decomposition (Simar & Wilson, 1998), since:

$$\Delta Eff^{t,t+1} = \Delta PureEff^{t,t+1} \cdot \Delta Scale^{t,t+1} \tag{7}$$

and

$$\Delta Tech^{t,t+1} = \Delta PureTech^{t,t+1} \cdot \Delta ScaleTech^{t,t+1}. \tag{8}$$

Where:

$$\Delta PureEff^{t,t+1} = \frac{\theta^{t+1}_{CRS}(x^{t+1}, y^{t+1})}{\theta^t_{CRS}(x^t, y^t)} \tag{9}$$

measures the change in relative efficiency, meaning how far production is from the maximum potential production.

$$\Delta Scale^{t,t+1} = \frac{\theta^{t+1}_{VRS}(x^{t+1}, y^{t+1}) / \theta^{t+1}_{CRS}(x^{t+1}, y^{t+1})}{\theta^t_{VRS}(x^t, y^t) / \theta^t_{CRS}(x^t, y^t)}. \tag{10}$$

measures the changes in scale efficiency.

$$\Delta PureTech^{t,t+1} = \left(\frac{\theta^t_{CRS}(x^{t+1}, y^{t+1})}{\theta^{t+1}_{CRS}(x^{t+1}, y^{t+1})} \cdot \frac{\theta^t_{CRS}(x^t, y^t)}{\theta^{t+1}_{CRS}(x^t, y^t)} \right)^{1/2} \tag{11}$$

is the shift in technology, and

$$\Delta ScaleTech^{t,t+1} = \left(\frac{\theta^t_{VRS}(x^{t+1}, y^{t+1}) / \theta^t_{CRS}(x^{t+1}, y^{t+1})}{\theta^{t+1}_{VRS}(x^{t+1}, y^{t+1}) / \theta^{t+1}_{CRS}(x^{t+1}, y^{t+1})} \cdot \\ \cdot \frac{\theta^t_{VRS}(x^t, y^t) / \theta^t_{CRS}(x^t, y^t)}{\theta^{t+1}_{VRS}(x^t, y^t) / \theta^{t+1}_{CRS}(x^t, y^t)} \right)^{1/2} \tag{12}$$

measures the changes in scale technology, i.e. change in the shape of the technology (Simar & Wilson, 1998).

While $\Delta PureEff^{t,t+1}$ and $\Delta Scale^{t,t+1}$ have easily demonstrable circularity, the other terms, $\Delta PureTech^{t,t+1}$ and $\Delta ScaleTech^{t,t+1}$, are not circular. Starting with $\Delta PureTech^{t,t+1}$, it is a geometric mean of two factors that represent relative changes:

$$\Delta PureTech^{t,t+1} = \left(\frac{\theta^t_{CRS}(x^{t+1}, y^{t+1})}{\theta^{t+1}_{CRS}(x^{t+1}, y^{t+1})} \cdot \frac{\theta^t_{CRS}(x^t, y^t)}{\theta^{t+1}_{CRS}(x^t, y^t)} \right)^{1/2} = \\ = \left(\Delta PT^{t+1}_{t,t+1} \cdot \Delta PT^t_{t,t+1} \right)^{1/2} \tag{13}$$

And, if the production unit is fixed at times t or $t+1$, each of the following terms is circular:

$$\Delta PT^t_{t,t+2} = \Delta PT^t_{t,t+1} \cdot \Delta PT^{t+1}_{t,t+2} \tag{14}$$

$$\Delta PT^{t+1}_{t,t+2} = \Delta PT^{t+1}_{t,t+1} \cdot \Delta PT^{t+1}_{t+1,t+2} \tag{15}$$

Taking this into account, it seems possible to forecast each circular component separately. Given a production unit working at levels (x^t, y^t) for different time periods t, with $t = 1, ..., T$, we can have: $\Delta PT^t_{s,s+1}$ where $s = 1, ..., T-1; t = 1, ..., T$, which can be organized in a table such as Table 1.

The aim is to forecast, based on the data from $t = 1, ..., T$, the productivity performance of a production unit from the time period T to the time period $T+1$. So, we need to forecast:

$$\Pi^{T,T+1} = \Delta PureEff^{T,T+1} \cdot \Delta Scale^{T,T+1} \cdot \Delta PureTech^{T,T+1} \cdot \Delta ScaleTech^{T,T+1} \tag{16}$$

Firstly the circular terms ($\Delta PureEff^{t,t+1}$ and $\Delta Scale^{t,t+1}$) are forecasted using the time-series method auto-regressive moving average (ARMA), mentioned by Daskovska et al. (2010). After, $\Delta PureTech^{t,t+1}$ is forecasted. This is done treating each term, $\Delta PT^{T+1}_{T,T+1}$ and $\Delta PT^{T}_{T,T+1}$, of Equation 13 independently.

This is compiled in Table 1, in which the last two entries of the lower row correspond to the terms of interest, with their geometrical mean ($\Delta \hat{PT}^{T}_{T,T+1} \cdot \Delta \hat{PT}^{T+1}_{T,T+1}$) being the wanted forecast.

The forecast of the term $\Delta ScaleTech^{t,t+1}$ is obtained by the same procedure as the one used for $\Delta PureTech^{t,t+1}$, since it presents the same structure.

Finally, the MPI forecast is given by the product of all the previously forecasted indices:

$$\hat{\Pi}^{T,T+1} = \Delta \widehat{PureEff}^{T,T+1} \cdot \Delta \widehat{Scale}^{T,T+1} \cdot \Delta \widehat{PureTech}^{T,T+1} \cdot \Delta \widehat{ScaleTech}^{T,T+1} \tag{17}$$

4.4. Pre-processing

The data collected presented sporadic data gaps. Excluding these DMUs from the analysis would result in less reliable results. For this reason, correlation between variables and linear regression were adopted to solve the problem.

Principal Component Analysis (PCA) was used to reduce the dimensions of the used data, in order to prevent the overestimation of the efficiency ratios. This resulted in one vector for input and one for output, since the first principal component explained, in both cases, more than 90% of the total variance.

5. Case Study

Firstly, as this analysis is about public hospitals, local health units (*"Unidade Local de Saúde"*) were not included. Moreover, only entities with public management were of interest so private hospitals, public-private partnerships and hospitals run by the *Misericórdias* were not included. Specialized hospitals, with specific technology of production, such as maternities, oncology centers (*"Instituto Português de Oncologia (IPO)"*) and psychiatric hospitals are also rejected.

Taking this into account, from all the portuguese health facilities, besides the the local health units and IPOs, the following were not included:

- *Hospital de Magalhães Lemos, EPE*, which is a psychiatric hospital;
- *Hospital de Braga, EPE*, not included since it was under a public-private partnership during the studied years;
- *Centro Hospitalar do Oeste, EPE*, only created in 2018, so no data from the analyzed years was available;
- *Hospital da Senhora da Oliveira Guimarães, EPE*, which was part of the health center *Centro Hospitalar do Alto Ave, EPE* along with *Hospital São José* in Fafe until 2015 (year in which hospitals belonging to the *Misericórdias* were returned), not included since the data registered for the years prior and after 2015 is not consistent for a reliable analysis.[6]

So, a total of 26 hospital and hospital centers, shown in Table 2, were studied.

All data were collected from the Portuguese Central Health System Administration (ACSS) website.[7]. This website has data from year 2012 to the present year of 2020, however data relative to year 2012 and from 2018 until 2020 presented several gaps in several hospitals. Thus, the data analyzed in this work refers to years 2013 to 2017.

The variables were chosen according to the ones most used in the literature review performed and to what was available on the ACSS website.[7] Five variables were chosen as inputs and three as outputs. As inputs: costs of external services and suppplies, costs of staff, costs of clinical consumption material all per standard patient, standard patients per doctor FTE and standard patients per nurse FTE. And, as outputs: discharges per bed, total number of medical appointments and total number of emergency room visits.

Since the number of hospital days, was not found in the database, it was replaced with the next most used as output: total number of emergency room visits.

[6]Serviço Nacional de Saúde - Entidades de Saúde (Available at: www.sns.gov.pt/institucional/entidades-de-saude/). Accessed on: 1/6/2020

[7]ACSS - Benchmarking Hospitais (Available at: benchmarking-ACSS.min-saude.pt). Accessed on 4/5/2020

Table 1: Technological change forecast decomposition. Table adapted from Daskovska et al. (2010).

Time periods' shift	(x^1, y^1) fixed	(x^2, y^2) fixed	...	(x^T, y^T) fixed	Forecast
1,2	$\Delta\hat{PT}^1_{1,2}$	$\Delta\hat{PT}^2_{1,2}$	...	$\Delta\hat{PT}^T_{1,2}$	$\Delta\hat{PT}^{T+1}_{1,2}$
2,3	$\Delta\hat{PT}^1_{2,3}$	$\Delta\hat{PT}^2_{2,3}$	...	$\Delta\hat{PT}^T_{2,3}$	$\Delta\hat{PT}^{T+1}_{2,3}$
...	...	...	...	...	...
$T-1, T$	$\Delta\hat{PT}^1_{T-1,T}$	$\Delta\hat{PT}^2_{T-1,T}$	...	$\Delta\hat{PT}^T_{T-1,T}$	$\Delta\hat{PT}^{T+1}_{T-1,T}$
Forecast $T, T+1$	$\Delta\hat{PT}^1_{T,T+1}$	$\Delta\hat{PT}^2_{T,T+1}$	...	$\Delta\hat{PT}^T_{T,T+1}$	$\Delta\hat{PT}^{T+1}_{T,T+1}$

Table 2: Portuguese public hospitals analyzed.

DMU	Hospital
H_1	Centro Hospitalar Barreiro/Montijo, EPE
H_2	Centro Hospitalar de Leiria, EPE
H_3	Centro Hospitalar de Lisboa Ocidental, EPE
H_4	Centro Hospitalar de Setúbal, EPE
H_5	Centro Hospitalar do Baixo Vouga, EPE
H_6	Centro Hospitalar do Médio Ave, EPE
H_7	Centro Hospitalar e Universitário de Coimbra, EPE
H_8	Centro Hospitalar Entre Douro e Vouga, EPE
H_9	Centro Hospitalar Médio Tejo, EPE
H_{10}	Centro Hospitalar Póvoa de Varzim/Vila do Conde, EPE
H_{11}	Centro Hospitalar Tâmega e Sousa, EPE
H_{12}	Centro Hospitalar Tondela-Viseu, EPE
H_{13}	Centro Hospitalar Trás-os-Montes e Alto Douro, EPE
H_{14}	Centro Hospitalar Universitário Cova da Beira, EPE
H_{15}	Centro Hospitalar Universitário de Lisboa Central, EPE
H_{16}	Centro Hospitalar Universitário de São João, EPE
H_{17}	Centro Hospitalar Universitário do Algarve, EPE
H_{18}	Centro Hospitalar Universitário do Porto, EPE
H_{19}	Centro Hospitalar Universitário Lisboa Norte, EPE
H_{20}	Centro Hospitalar Vila Nova de Gaia/Espinho, EPE
H_{21}	Hospital Distrital da Figueira da Foz, EPE
H_{22}	Hospital Distrital de Santarém, EPE
H_{23}	Hospital Espírito Santo de Évora, EPE
H_{24}	Hospital Fernando Fonseca, EPE
H_{25}	Hospital Garcia de Orta, EPE
H_{26}	Hospital Santa Maria Maior, EPE

DMU: Decision Making Unit; EPE: *Entidade Pública Empresarial*

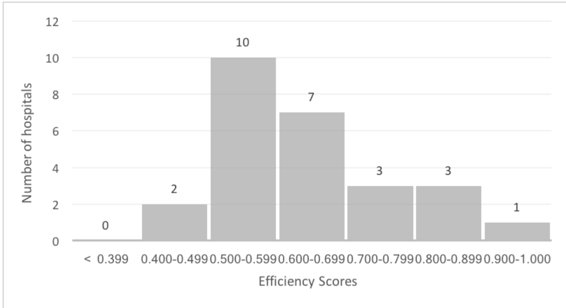

Figure 1: Distribution of the mean efficiency values, when considering Constant Returns to Scale.

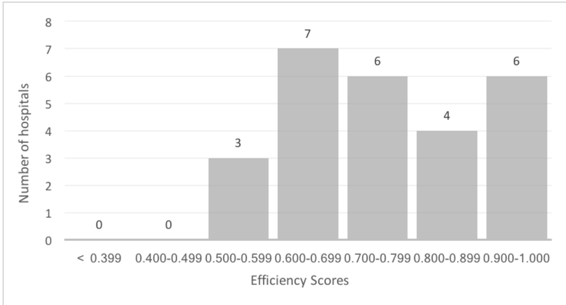

Figure 2: Distribution of the mean efficiency values, when considering Variable Returns to Scale.

6. Results & discussion

Firstly, the results obtained with DEA, are presented. These results were obtained both considering CRS and VRS. Figures 1 and 2 show the distribution of the mean efficiency scores, i.e. the number of hospitals that present an average score between the shown intervals, during the studied years.

The scores vary in a considerate range, from the minimum observed 0.116 to 1.000. A score of 1.000 corresponds to an efficient unit, managing correctly their resources. In all analyzed years, both under CRS and VRS, at least one unit is considered efficient in every month, being this a benchmark for the other less efficient units. The other hospitals present inefficiency, that can come from different sources, and they could decrease their inputs to produce the same quantity of outputs. Considering CRS, the average efficiency score is 0.648 with an average standard deviation of 0.143. And when assuming VRS, the average efficiency score is 0.764 with an average standard deviation of 0.097. Hospitals have better efficiency scores when considering VRS, and the results are more homogeneous. Moreover, the number of efficient units increases when considering VRS. Observing these Figures, it can be verified that, if assuming CRS, most hospitals do not show great efficiency scores, with most having values between 0.500 and 0.599, and with 17 between 0.500 and 0.699, which represents 65% of the analyzed hospitals. Only one has an average between 0.900 and 1.000. When assuming VRS, hospitals seem to perform better, with most - 88%, compared to only 54% with CRS - presenting an average score of more than 0.600, and six between 0.900 and 1.000.

Considering VRS, H_{26} has the best average efficiency score, being efficient every month from 2014 on, having a mean of 1.000 every year except 2013. Other two hospitals are also efficient during entire years: H_7 from 2015 to 2017 and H_{11} in 2015 and 2017, also presenting mean of 1.000 for these years. This does not happen with CRS, since no hospital is efficient every month of one year, so no hospital presents a mean of 1.000. Nevertheless, H_8 and H_{11} are the ones with the best average efficiency scores, considering CRS. Together with H_6, H_{22} and H_{26} they are the only that are benchmarks in any of the periods analyzed, under CRS, as they present maximum values of 1.000 in at least one year. Considering VRS, 13 hospitals are benchmarks in at least one of the periods considered. H_7, H_{16} and H_{18} are the ones most consistent, presenting an average score greater than 0.9 for all years.

H_{17} is always the one with the lowest average efficiency scores for all years under CRS and VRS, with the exception of 2013 under VRS. Despite presenting better results for some months (mostly June and August, under VRS), its average efficiency scores are not bigger than 0.600. These results go accordingly to what is commonly known about this hospital - *Centro Hospitalar Universitário do Algarve, EPE* - which is that it has a greater influx of people during the summer months since there is a large movement of people to this region.[8]

June is the month which presents the lowest average efficiency scores more often, both under CRS and VRS especially from 2014 on.

As can be seen in Table 3, the average efficiency scores both under CRS and VRS, has been overall increasing over time. However, its behaviour per month is not always increasing, presenting peaks of minimum values every June. This behaviour reflects the existence of seasonality in the results. The lower peaks correspond to a month where minimum values of efficiency occur in several hospitals.

There was a bigger increase in efficiency during the analyzed years under CRS than under VRS. The values of average efficiency score are the most heterogeneous in 2013 under CRS, as can be verified by the standard deviation values. The most homogeneous values occur in 2016 under VRS.

Regarding scale efficiency, the results are very heterogeneous, with hospitals having efficiency scores ranging from as low as 0.174 to the maximum possible of one. Moreover they show that, in every month of every analyzed year, there is at least one hospital with a scale efficiency of one, which is the maximum value. These correspond

[8]Centro Hospitalar Universitário do Algarve (Available at: www.chualgarve.min-saude.pt/chalgarve-em-numeros/area-de-influencia/) Accessed on: 25/11/2020

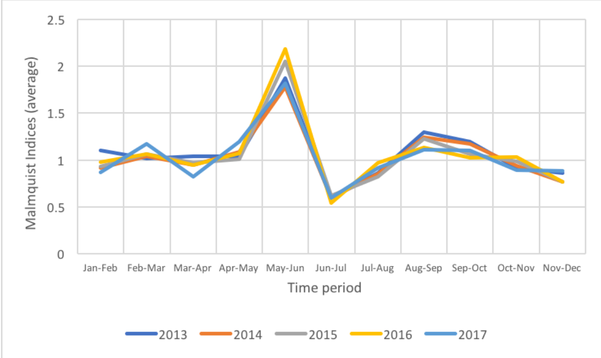

Figure 3: Evolution of the average Malmquist Productivity Index for all hospitals through all time periods in each year.

to the hospitals that present the same efficiency score considering both CRS and VRS, showing scale efficiency, and meaning that the DMU is operating at the optimal scale. H_7 is the one with the lowest scale efficiency scores for all analyzed years, and H_{11} presents the highest average scores in most years (2014, 2015 and 2017). Only five hospitals present scale efficiency at some month in these five years. H_{11} is the one that presents a score of 1.000 more times, being the one that is scale efficient in most months, in particular from 2015 on and being scale efficient for half of the periods studied. H_8 and H_{26} follow this one being also scale efficient in several months. In addition to these, only H_{22} and H_6 are also scale efficient in some period. The rest always present scale inefficiency. These hospitals are not operating at their optimum size and could benefit from an adjustment of their production capacity. The average scale efficiency scores have been slowly increasing over the years, as well as the minimum values. This means that, hospitals may be slowly approaching their appropriate and optimal size, in particular the ones that present the lowest scores.

The MPI presents a total average of 1.049 and standard deviation of 0.475. Its values range from 0.054 to 6.137. However, the average value of MPI for each year does not vary very much, being around 1.000 for all five analyzed years. This suggests that the productivity of the Portuguese hospitals has not changed significantly through these years.

Even though none of the hospitals have values bigger than 1.000 in every analyzed period, 12 present an average per year that is greater than 1.000 for all years. From these, H_{26} is the one that presents the biggest average MPI and is the highest in 2017. H_6, H_{17} and H_{22} present the highest indices in 2014, 2015 and 2016 and 2013, respectively. H_{17} and H_{22}, however, are not consistently productive. Firstly, H_{22} presents very good efficient scores at the end of 2013 due to a de-

Table 3: Evolution of the statistical values of the efficiency scores obtained with Data Envelopment Analysis over the analyzed years.

	CRS[1]				VRS[2]			
	Mean	Std. Deviation	Minimum	Maximum	Mean	Std. Deviation	Minimum	Maximum
2013	0.494	0.220	0.200	0.741	0.710	0.087	0.591	0.819
2014	0.669	0.087	0.389	0.736	0.764	0.056	0.585	0.805
2015	0.678	0.092	0.389	0.735	0.763	0.052	0.594	0.800
2016	0.694	0.089	0.406	0.738	0.781	0.039	0.654	0.808
2017	0.706	0.079	0.466	0.773	0.784	0.054	0.611	0.821

[1]CRS: Constant Returns to Scale; [2]VRS: Variable Returns to Scale

crease in the costs and increase in the number of medical appointments, which leads to an average high MPI score. H_{17} was already discussed and presents high efficiency scores when the rest of of the hospitals don't (vacation months), leading to high indexes which average to a good overall result. There are also hospitals that never present average values in any year bigger than 1.000, suggesting they are not progressing in terms of productivity. These are H_{18} and H_{19}. Despite having an MPI bigger than 1.000 in some time periods, H_{18} and H_{19} do not present averages bigger than 1.000. H_{18}, however, is one of the most efficient, so this low average productivity indexes are justifiable since, because it performs very well in all months except the summer months, the index for these is low, leading to a low average.

Overall, the results of the MPI are very homogeneous, and there is not one hospital that clearly stands out, for example, in terms of presenting indexes bigger than 1.000 for all periods. It can be seen that, regardless of the year, the period May-June presents the higher values of productivity. On the other hand, the period June-July exhibits the lowest values, being very often and for most hospitals less than 1.000. This suggests a seasonal effect, which means, in the case of May-June, that during these periods, there is an increase in production, a decrease in spending and in the resources used, progress in the production technology or any combination of these, and the contrary in June-July. Thus, from May to June there is an increase in productivity and from June to July there is a decrease. The August-September increase also observable may be in line with the fact that during August, which is the principal vacation month, the majority of hospitals will present less amount of outputs but no change in the inputs, being less productive.

As previously mentioned, the MPI can be divided into change in efficiency (ΔEff) and technology change (ΔTech), consisting in the geometric mean of these two terms. ΔTech is generally higher than ΔEff, which may lead to ΔTech influencing more the total MPI. With the exception of 2013, ΔEff presents its higher values in the period June-July and ΔTech in the period May-June. It makes sense that June-July presents the highest values for ΔEff since, as was seen in the DEA results, efficiency presents its lower values in June, which leads to the biggest increase in June-July. The lowest and highest values of these two terms belong to the same periods, but are switched. Hence, when ΔEff presents its highest values, ΔTech presents its lowest and vice versa, with the exception of 2013. In terms of ΔEff, H_{20} presents the best average from 2013 to 2015. And H_{23} and H_{19} for 2016 and 2017, respectively. The lowest averages belong to H_{22}, H_6, H_{17} and H_9.

The term contributing more to the MPI change seems to be ΔTech, the term regarding the technological change, so when the total MPI presents a peak, either low or high, it is the term ΔTech that presents a peak in the same period, even if the other term, ΔEff, presents values with the opposite trend. Thus, a hospital with a decrease in efficiency can still present an index that suggests productivity increase. ΔTech has been very slowly decreasing over the years, but never reaching an average lower than one. And the same can be said about ΔEff. However, as was seen before, DEA is increasing. So, even though the hospital's efficiency has been increasing over the studied years, the rate at which they have been becoming more efficient has been declining.

The MPI can be further decomposed into the four different terms: ΔPureEff, ΔScale, ΔPureTech and ΔScaleTech. The MPI calculated with its decomposition in four terms presents lower values for the averages of each year, when compared to the MPI calculated with the decomposition into only two terms. And when decomposed in four terms, the MPI shows a clear increase in its average values throughout the years, as seen in Table 4. This can, however, be because the two hospitals that did not present the lowest results - H_{18} and H_{19} - are not being considered here due to what was already mentioned about the computation of this decomposition terms.

Note that the decomposition into the four terms involves the computation of efficiency assuming VRS, which may lead to no feasible point being found in the linear programming method when computing the efficiency for one time period projected into another, which then sets some hospitals' MPI results to NaN. So, these hospitals can not be considered in this MPI decomposition's results.

In this case, it is again a term relative to technology, △PureTech, that affects more the total MPI. The upper and lower peaks more evident of the MPI overlap with upper and lower peaks of the △PureTech term.

Lastly, the MPI for the year 2018 was forecasted. The results show that, despite the values of the MPI predicted being a bit higher, they continue to be in line with the ones from previous years, which present indices around 1.000. The average MPI forecasted for 2018 is 1.229, meaning a productivity increase should be expected in 2018. All hospitals present an average index bigger than 1.000, and the same is observed for the 11 time periods forecasted. The two hospitals with the best average forecasted MPI are H_{17} and H_{26}, and the ones with the lowest values are H_{23} and H_{25}. This does not differ significantly from what could be expected, since these units are also the ones presenting some of the highest and lowest values, respectively, of MPI in some of the previous years. In another way, however, the period with the lowest forecast is May-June, which does not meet the MPI pattern from previous years. February-March presents the highest forecasted MPI.

△PureEff, which represents the pure efficiency change, has an average forecast of 1.054, which suggests an increase, even though small, in hospital efficiency in 2018. This somewhat makes sense given the DEA results obtained and the efficiency scores tendency observed. △Scale, which measures the changes in scale efficiency of the production unit, presents a mean of 1.046, suggesting changes in the returns to scale faced by the production unit, especially an increase in scale efficiency in 2018. △PureTech and △ScaleTech present forecasted average values of 1.119 and 1.001, respectively. Both these terms indicate changes related to the technology frontier, hence meaning an increase in technology in 2018. △ScaleTech is close to 1.000, which means the shape of technology does not change significantly.

The actual MPI for the year 2018 can be calculated for some (21) hospitals and some time periods. The hospitals not included did not have several variables data for the year 2018, and the time period for which there are no results presented (November-December) did not have reliable data for the year under analysis. Regardless, with the MPI calculated for these hospitals and time periods of 2018, it is possible to assess the reliability and accuracy of the forecast performed. In order to do so, the mean absolute error (MAE) and root mean square error (RMSE) were calculated. The average MAE is 0.392 and the total RMSE is 0.294.

In particular, the best forecast is done for hospital H_2. The worst is H_{21}. In terms of periods, the one with the worst average MAE is May-June. Taking into account the forecasted values for this period, it was already pretty obvious that this would be the period with the biggest error, as was discussed before, since it presented in the past years a peak in this period and so it would be expected that 2018 would follow the example, which was not seen in the forecast. This suggests that the forecast method is probably not the most adequate to account for seasonality in the data.

Some terms are forecasted with better precision than others (Table 5). While △PureTech and △PureEff have a bigger average of MAE and RMSE values, the other two terms, △ScaleTech and △Scale, have lower values, with △ScaleTech having the lowest. The △ScaleTech term presents very close values between hospitals, but without a trend as clear as other terms. For example △PureTech, the term with the biggest error, presents a more clear trend of the data, presenting more clear ups and downs, but with values varying over a larger range.

7. Conclusions

Making use of the DEA method, the efficiency of all public hospitals and hospital centers was calculated and interpreted. 26 hospitals and hospital centers were analyzed for the years 2013 to 2017. The Malmquist productivity index was then used to assess the productivity of these same hospitals, using two different decompositions based on the works of Simar and Wilson (1998). The MPI was also forecasted following the theory developed by Daskovska et al. (2010). The forecast was calculated for the year 2018 since the MPI results were obtained until 2017. This allowed for the evaluation of the forecasting technique, comparing results of some hospitals, after calculating their MPI for 2018.

Overall, the hospitals that presented the best results in terms of efficiency are either because they present the best average efficiency score of a year or because they are the ones presenting good values for the majority of the analyzed time periods. In general, the performance of hospitals has been slowly increasing. The overall average DEA score considering CRS was 0.648 and under VRS 0.764 and seems to be increasing throughout the years. Scale efficiency is also globally increasing. In terms of productivity, the MPI shows seasonality, presenting high peaks in May-June for every year

Table 4: Malmquist Productivity Index results for the total Malmquist Productivity Index and its terms, considering the decompositions into two and four terms.

	MPI	∆Eff	∆Tech	MPI	∆PureEff	∆Scale	∆PureTech	∆ScaleTech
2013	1.073	1.371	1.425	0.812	1.133	1.321	1.248	0.659
2014	1.031	1.075	1.139	0.964	1.053	1.023	1.102	0.897
2015	1.047	1.088	1.168	0.973	1.059	1.048	1.150	0.909
2016	1.063	1.083	1.187	0.985	1.059	1.048	1.150	0.909
2017	1.033	1.055	1.111	1.009	1.052	1.011	1.107	0.946

Table 5: Errors (Mean Absolute Error and Root Mean Square Error) of the Malmquist Productivity Index forecast, for the total Malmquist Productivity Index and its four different terms.

	Total MPI	∆PureEff	∆Scale	∆PureTech	∆ScaleTech
MAE	0.392	0.214	0.104	0.391	0.081
RMSE	0.294	0.171	0.020	0.444	0.013

between 2013 and 2017. The overall average MPI is 1.049, showing productivity increase. The terms regarding changes in technology seem to influence more the MPI than the ones considering efficiency changes, maybe because their values are bigger. The results obtained in this work are consistent with other studies found in the literature. For example, the hospitals that perform better all belong to more coastal areas, in line with (D. C. Ferreira, Nunes, & Marques, 2018) and not the interior of the country. Moreover, the RHA to which most belong to is the North RHA (*ARS do Norte*), in line with (D. C. Ferreira & Nunes, 2019).

Considering the second part of this work, the forecasted MPI did not present good enough results, forecasting values that are not close enough to the real ones for it to be considered a reliable forecast, which may be due to the complexity of the healthcare data as well as the method considered for the forecast. However, to the extent of the research performed, this forecast had not been applied before.

7.1. Limitations

Firstly, the lack of data available and the data gaps existing present one of the limitations of this work. This was the case of years 2018 and 2019, for which, at the date of the work, there was not enough data available to perform a reliable analysis. Another limitation is the fact that the forecast was done for a year that has already passed and for which there was already some information, even though not enough for a complete analysis. The heterogeneity of the sample of hospitals, both as group and individually, is also another limitation. The sample is composed of hospitals and hospital centers, which are inevitably different in dimensions and activity. Also, hospitals are very heterogeneous in the data, meaning that values are very disperse, and vary a lot, being difficult to model. Moreover, no quality variables were considered. As well as the complexity of the environment and patients treated. The healthcare area is complex and very particular, making the services provided quite complicated to evaluate. Despite being easy to identify and quantify inputs, such as spendings or number of staff, it is much harder to quantify outputs.

7.2. Future work

The main future work suggestions are to resolve the limitations, already stated, of this work. More specifically, include quality and access factors in the analysis, to provide a more complete study. As well as the use of exogenous variables, adjustment to environmental factors or case mix index (CMI), thus taking into account the environment in which a hospitals operates. Moreover, an analysis comprising also private hospitals and PPPs could be of interest. Future work suggestions include also the forecasting of the MPI for more recent years, perhaps for the present year of 2020 since it could be interesting to compare with actual values to assess how the COVID-19 pandemic impacted the efficiency of hospitals and in what ways. Other relevant future work is the exploitation of other forecasting techniques, to assess if it is possible to obtain better results. The smooth bootstrap adaptation mentioned in Daskovska et al. (2010) can also be explored and developed in practice, to make inferences about the forecasted MPI.

Acknowledgments

I would like to thank my supervisors Prof. Diogo and Prof. Alexandre, without whom this work would not have been possible. Thank you also to my friends and family who have accompanied me throughout this journey.

References

Bezat, A. (2009). Comparison of the Deterministic and Stochastic Approaches for Estimating Technical Efficiency on the Example of Non-Parametric Dea and Parametric Sfa Methods. *Metody Ilościowe w Badaniach Ekonomicznych*, *10*(1), 20–29.

Büchner, V. A., Hinz, V., & Schreyögg, J. (2016). Health systems: changes in hospital efficiency and profitability. *Health Care Management Science*, *19*(2), 130–143. doi: 10.1007/s10729-014-9303-1

Daskovska, A., Simar, L., & van Bellegem, S. (2010). Forecasting the malmquist productivity index. *Journal of Productivity Analysis*, *33*(2), 97–107. doi: 10.1007/s11123-009-0147-5

Ersoy, K., Kavuncubasi, S., Ozcan, Y. A., & Harris, J. M. (1997). Technical efficiencies of Turkish hospitals: DEA approach. *Journal of Medical Systems*, *21*(2), 67-74. doi: 10.1023/A:1022801222540

Ferreira, D., & Marques, R. C. (2015). Did the corporatization of Portuguese hospitals significantly change their productivity? *European Journal of Health Economics*. doi: 10.1007/s10198-014-0574-8

Ferreira, D. C., & Marques, R. C. (2019). Do quality and access to hospital services impact on their technical efficiency? *Omega (United Kingdom)*. doi: 10.1016/j.omega.2018.07.010

Ferreira, D. C., & Marques, R. C. (2020). Public-private partnerships in health care services: Do they outperform public hospitals regarding quality and access? Evidence from Portugal. *Socio-Economic Planning Sciences*. doi: 10.1016/j.seps.2020.100798

Ferreira, D. C., Marques, R. C., & Nunes, A. M. (2018). Economies of scope in the health sector: The case of Portuguese hospitals. *European Journal of Operational Research*. doi: 10.1016/j.ejor.2017.09.044

Ferreira, D. C., & Nunes, A. M. (2019). Technical efficiency of Portuguese public hospitals: A comparative analysis across the five regions of Portugal. *International Journal of Health Planning and Management*. doi: 10.1002/hpm.2658

Ferreira, D. C., Nunes, A. M., & Marques, R. C. (2018). Doctors, nurses, and the optimal scale size in the Portuguese public hospitals. *Health Policy*. doi: 10.1016/j.healthpol.2018.06.009

Jacobs, R. (2001). Alternative methods to examine hospital efficiency: Data envelopment analysis and stochastic frontier analysis. *Health Care Management Science*, 103–115. doi: 10.1023/A:1011453526849

Ji, Y. B., & Lee, C. (2010). Data envelopment analysis. *Stata Journal*, *10*(2), 267-280. doi: 10.26537/rebules.v0i6.811

Kirigia, J., & Asbu, E. (2013). Technical and scale efficiency of public community hospitals in Eritrea: An exploratory study. *Health Economics Review*, *3*(1), 1–16.

Lins, M. E., Lobo, M. S. D. C., Moreira Da Silva, A. C., Fiszman, R., & Ribeiro, V. J. D. P. (2007). The use of Data Envelopment Analysis (DEA) for Brazilian teaching hospitals' evaluation. *Ciencia e Saude Coletiva*, *12*(4), 985–998. doi: 10.1590/s1413-81232007000400020

Nunes, A. M., & Ferreira, D. C. (2019). The health care reform in Portugal: Outcomes from both the New Public Management and the economic crisis. *International Journal of Health Planning and Management*. doi: 10.1002/hpm.2613

Peacock, S., Chris, C., Melvino, M., & Johansen, D. (2001). Techniques for measuring efficiency in health services. *Productivity Commision Staff Working Paper*(July), 1–97.

Prezerakos, P., Maniadakis, N., Kaitelidou, D., Kotsopoulos, N., & Yfantopoulos, J. (2007). Measuring Across Hospital Efficiency and Productivity: The Case of Second Regional Health Authority of Attica. *European Research Studies Journal*, *X*(1-2), 83–96.

Simar, L., & Wilson, P. (1998). Productivity Growth in Industrialized Countries. *Papers*.

Simar, L., & Wilson, P. W. (1999). Estimating and bootstrapping Malmquist indices. *European Journal of Operational Research*, *115*(3), 459–471. doi: 10.1016/S0377-2217(97)00450-5

Simões, J. A., Augusto, G. F., Fronteira, I., & Hernandez-Quevedo, C. (2017). Portugal: Health System Review. *Health systems in transition*, *19*(2), 1–184.

Tone, K. (2005). Malmquist Productivity Index. In *Handbook on data envelopment analysis* (p. 203-227). Springer, Boston, MA. doi: 10.1007/1-4020-7798-x_8

Weng, S. J., Wu, T., Blackhurst, J., & Mackulak, G. (2009). An extended DEA model for hospital performance evaluation and improvement. *Health Services and Outcomes Research Methodology*, *9*(1), 39-53. doi: 10.1007/s10742-008-0042-y