



Pay for performance in health care: a new best practice tariff-based tool using a log-linear piecewise frontier function and a dual–primal approach for unique solutions

Diogo Cunha Ferreira^{1,2}  · Rui Cunha Marques¹ · Alexandre Morais Nunes^{1,2}

Received: 3 August 2018 / Revised: 7 April 2019 / Accepted: 7 June 2019 /

Published online: 19 June 2019

© The Author(s) 2020, corrected publication 2020

Abstract

Health care systems worldwide have faced a problem of resources scarcity that, in turn, should be allocated to the health care providers according to the corresponding population needs. Such an allocation should be as much as effective and efficient as possible to guarantee the sustainability of those systems. One alternative to reach that goal is through (prospective) payments due to the providers for their clinical procedures. The way that such payments are computed is frequently unknown and arguably far from being optimal. For instance, in Portugal, public hospitals are clustered based on criteria related to size, consumed resources, and volume of medical acts, and payments associated with the inpatient services are equal to the smallest unitary cost within each cluster. First, there is no reason to impose a single benchmark for each inefficient hospital. Second, this approach disregards dimensions like quality (and access) and the environment, which are paramount for fair comparisons and benchmarking exercises. This paper proposes an innovative tool to achieve best-practices tariff. This tool merges both quality and financial sustainability concepts, attributing a hospital-specific tariff that can be different from hospital to hospital. That payment results from the combination of costs related to a set of potential benchmarks, keeping quality as high as possible and higher than a user-predefined threshold, and being able to generate considerable cost savings. To obtain those coefficients we propose and detail a log-linear piecewise frontier function as well as a dual–primal approach for unique solutions.

Keywords Pay for performance · Best practice tariff · Hospital · Services · Log-linear piecewise frontier functions · Unique solutions-based log-data envelopment analysis

✉ Diogo Cunha Ferreira
diogo.cunha.ferreira@tecnico.ulisboa.pt

Extended author information available on the last page of the article

1 Introduction

In several countries worldwide, health care providers are paid by the services that they provide to the citizens. These providers include physicians, nurses, and mostly primary care centres and hospitals. Usually these payments depend on dimensions included in contracts celebrated between the payer (e.g. the State) and the payee (the provider). Several schemes are available: (a) block budget, featured by a periodic prospective payment associated with the activity, (b) capitation, based on the number of enrolled patients, multiplied by a unitary price, (c) case-based payments, such that health care providers are paid a prospective/retrospective lump sum per episode of care, and (d) fee for service, characterized by the payment of a price per medical act in a retrospective way. Advantages and disadvantages of these schemes can be found in the relevant literature (e.g. Marshall et al. 2014; Friesner and Rosenman 2009; Street et al. 2011).

It is important to stress that resources, namely the financial ones, are scarcer each passing day, motivating an optimization exercise to better allocate them. To the best of our knowledge, this is the first attempt to optimize payments to the health care providers. A common feature of all payment schemes is the paid price or tariff, which is usually set as the average cost per patient/medical act/episode of care. We remark that health care is a public interest service and, as such, resources allocation and payments should account not only for the efficiency of providers but also for their quality and access (Ferreira and Marques 2018b). Despite some contracts contain a number of quality and access parameters to be fulfilled, in most of the cases penalties are not sufficient to induce good practices. Instead of fixing these parameters (quality, access) and hoping that providers do not adopt misconducts, we should find out a reference set of providers for each health care provider. Such a reference set should contain only those entities whose quality and access observations are above a pre-defined threshold that indicates a minimum acceptable level of social performance. This intends to avoid that poor quality providers can be potential benchmarks. Additionally, optimal allocation of resources requires fair comparisons in terms of internal and external operational environment (Karagiannis and Velentzas 2012; Cordero et al. 2018). Efficiency of health care providers is dependent on epidemiology and demographics as well as on their specialization degree (Ferreira et al. 2017), hence disregarding these dimensions from the payment optimization is likely to produce biased results. Hence, we restrict even more the aforementioned reference set using operational environment variables such that the reference set contains the entities with good quality and access levels and, simultaneously, operate under similar conditions as the provider whose payment we want to optimize. The construction of this reference set constitutes a bridge between operational research and health economics/management, which appears to be innovative in the field.

Once the reference set has been constructed, we should use a benchmarking tool to derive an efficiency frontier, where benchmarks are placed. It is usual to construct a common frontier by using the entire set of health care providers. But since we want neither unfair comparisons nor potential benchmarks with poor quality, we use

the reference set to construct such a frontier. Because each provider has its own comparability set, it is difficult to parametrically define a frontier shape equal for all providers and expect good outcomes, including fitting parameters. Therefore, the frontier should be empirically constructed, i.e. based only on inputs, outputs, and the reference set that is fixed as the overall comparability set for each provider. Data Envelopment Analysis (DEA) (Charnes et al. 1979; Banker et al. 1984) is probably the most common method to reach this goal. Because of the frontier convexity, DEA-based efficiency levels can be sometimes underestimated, meaning that inputs (outputs) should be decreased (increased) beyond what they would if there was not that bias. Indeed, nonconvexity is usually a more natural assumption for the frontier. The extreme case corresponds to (FDH) (Deprins et al. 1984; Daraio and Simar 2007), although it assumes that each provider can have one and only one benchmark. Nevertheless, we believe that such an assumption is too restrictive but still we should keep nonconvexity. In view of that, we extend the work of Banker and Maindiratta (1986) and construct a log-linear piecewise frontier with a directional nature. Properties of this model are studied, including the existence of multiple solutions. By consequence, we propose an extension of that model that results from the work of Sueyoshi and Sekitani (2009) and mitigates the aforementioned problem. Moreover, in some circumstances, the log-linear program may become infeasible; thus, we propose a slightly change on the model to cope with this problem. Based on the optimized parameters resulting from the log-linear programming tool, we can derive optimal payments as well as potential cost savings related to the achieved tariffs. These are innovations in operational research.

The structure of the manuscript is as follows. Section 2 presents some useful concepts and definitions. Section 3 details and explores the new best practice tariff-based tool (the core of this study). Section 4 related efficiency, optimal payments, and cost savings for the commissioner. Section 5 applies the new tool to the Portuguese public hospitals. Section 6 provides some concluding remarks and explains the economic impact of this tool to the Portuguese National Health Service.

2 Useful concepts and definitions

2.1 Overview

Defining optimal payments due to the health care providers depends on a number of distinct features.¹ The first one is the level of care, service, or Diagnostic Related Groups (DRG) under analysis. Each country or health care service pays for clinical acts on different levels, e.g. directly to the physician or to the nurse, per service (inpatients, outpatients, surgeries, . . .), per speciality, per diagnosis group, and per severity level of each diagnosis group. The concept of DRG, for instance, has long been accepted by health care management researchers as a way of accounting for the patient-mix, which is to say a measure of heterogeneity among the patients

¹ Henceforward, we will use the term 'hospital' to denote the health care provider. Though, this does not mean that our approach cannot be used to optimize payments to other health care providers, such as physicians. This is only a matter of simplification.

admitted to the hospital services. Researchers have analysed the efficiency and productivity of hospitals based on such a concept. For instance, recently Johannessen et al. (2017) considered some DRG scores resulting from hospitalization, day-care, and outpatient consultations, so as to investigate the productivity improvement following a Norwegian hospital reform. Some references cited therein also considered DRG to homogenize the hospital activity.

The second feature is the set of variables used to establish such a payment (or tariff) and to avoid that it results from unfair comparisons and/or from benchmarks that disregarded important dimensions for citizens, such as quality and access. Hence, we need inputs (traditionally defined as the operational expenses required to treat patients), outputs (the quantity of treated patients), operational environment (which can be internal and/or external to the hospital), and quality (which also includes access). It is important to point out that, in our framework, quality is assumed to positively contribute to the hospital overall performance. In other words, from two quality observations, the largest one presents a higher utility to the hospital. However, in some cases, quality is measured through undesirable dimensions (e.g. avoidable mortality in low severity levels), demanding for an appropriate rescaling. It is not the goal of our paper, although an example is given in the next section.

Although DEA and models alike have been extensively utilized with DRG altogether, to the best of our knowledge, no other study has previously employed them to optimize payments in the health sector, particularly in DRG-based financing systems. This study can, then, be seen as a *first step* in order to optimize those payments, making them fairer and more sustainable, with adjustments for quality and operational environment.

The idea underlying the use of quality and operational environment-based dimensions is to construct comparability sets for each health care provider. Such sets should contain only those observations that are close to the provider whose tariff (payment) we want to optimize. Closeness (or proximity) is fixed by the so-called bandwidth. At this point, the comparability set associated with quality dimensions are formulated in a distinct way because, differently from size and environment, the higher the quality of potential benchmarks (best practices), the better. In fact, given a certain health care provider, if there is, at least, another entity delivering better health care with fewer resources per patient, then it should be a potential benchmark for the former one. Health care is a public interest service with several stakeholders, including citizens, staff, and the Government (either central or not). They usually have a minimum level of quality that is acceptable to a provider be considered a good performer. Because in our framework poor performers cannot be benchmarks, we impose a threshold that is defined as that minimum acceptable level per quality dimension.

2.2 Notation

In this paper, we have adopted the following notation:

- $\Omega = \{1, \dots, j, \dots, J\}$ is the list of indexes denoting each hospital from the sample, with $j, J \in \mathbb{N}$;
- $\Psi = \{1, \dots, \ell, \dots, L\}$ is the list of indexes denoting each service or DRG under analysis, for which the optimal tariff is being assessed, with $\ell, L \in \mathbb{N}$;
- $\Gamma^{(\ell)} = \{1, \dots, r^{(\ell)}, \dots, R^{(\ell)}\}$, $\ell \in \Psi$, is the list of indexes denoting each quality dimension associated with the ℓ th service/DRG, with $r^{(\ell)}, R^{(\ell)} \in \mathbb{N}$;
- $\Phi^{(\ell)} = \{1, \dots, c^{(\ell)}, \dots, C^{(\ell)}\}$, $\ell \in \Psi$, is the list of indexes of environment-related dimensions, with $c^{(\ell)}, C^{(\ell)} \in \mathbb{N}$;
- $x^{j(\ell)}$ denotes the total operational expenses related to the hospital $j \in \Omega$ and the service/DRG $\ell \in \Psi$, being $x^{j(\ell)} > 0$;
- $y^{j(\ell)}$ denotes the total number of in/outpatients associated with the hospital $j \in \Omega$ and the service/DRG $\ell \in \Psi$, being $y^{j(\ell)} > 0$;
- $q_{r^{(\ell)}}^{j(\ell)}$ is the $r^{(\ell)}$ th quality observation related to the hospital $j \in \Omega$ and the service/DRG $\ell \in \Psi$;
- $z_{c^{(\ell)}}^{j(\ell)}$, $\ell \in \Psi$, is the observation associated with the $c^{(\ell)}$ th environment dimension, the hospital $j \in \Omega$, and the service/DRG $\ell \in \Psi$;
- $\mathbf{X}^{(\ell)} = (x^{1(\ell)}, \dots, x^{j(\ell)}, \dots, x^{J(\ell)})^\top$, $\ell \in \Psi$, is the column vector of total operational expenses for the J hospitals in dataset Ω ;
- $\mathbf{Y}^{(\ell)} = (y^{1(\ell)}, \dots, y^{j(\ell)}, \dots, y^{J(\ell)})^\top$, $\ell \in \Psi$, is the column vector of total in/outpatients handled by the J hospitals in dataset Ω ;
- $\mathbf{Q}^{(\ell)}$ is the matrix of J observations related to the $R^{(\ell)}$ quality dimensions, where

$$\mathbf{Q}^{(\ell)} = \begin{bmatrix} q_1^{1(\ell)} & \cdots & q_1^{j(\ell)} & \cdots & q_1^{J(\ell)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ q_{r^{(\ell)}}^{1(\ell)} & \cdots & q_{r^{(\ell)}}^{j(\ell)} & \cdots & q_{r^{(\ell)}}^{J(\ell)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ q_{R^{(\ell)}}^{1(\ell)} & \cdots & q_{R^{(\ell)}}^{j(\ell)} & \cdots & q_{R^{(\ell)}}^{J(\ell)} \end{bmatrix}, \ell \in \Psi;$$

- $\mathbf{Z}^{(\ell)}$ is the matrix of J observations associated with the $C^{(\ell)}$ environment dimensions, where

$$\mathbf{Z}^{(\ell)} = \begin{bmatrix} z_1^{1(\ell)} & \cdots & z_1^{j(\ell)} & \cdots & z_1^{J(\ell)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{c^{(\ell)}}^{1(\ell)} & \cdots & z_{c^{(\ell)}}^{j(\ell)} & \cdots & z_{c^{(\ell)}}^{J(\ell)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{C^{(\ell)}}^{1(\ell)} & \cdots & z_{C^{(\ell)}}^{j(\ell)} & \cdots & z_{C^{(\ell)}}^{J(\ell)} \end{bmatrix}, \ell \in \Psi;$$

- $b_y^{j(\ell)}$ is a strictly positive bandwidth related to the j th hospital, the ℓ th service/DRG, and the total number of in/outpatients treated in that hospital, $y^{j(\ell)}$;

- $b_{q_r}^{j(\ell)}$ is a strictly positive bandwidth associated with the j th hospital, the ℓ th service/DRG, and the $r^{(\ell)}$ th quality dimension;
- $b_{z_c}^{j(\ell)}$ is a strictly positive bandwidth related to the j th hospital, the ℓ th service/DRG, and the $c^{(\ell)}$ th environment dimension;
- $t_{q_r}^{(\ell)}$ is a strictly positive user-defined threshold associated with the $r^{(\ell)}$ th quality dimension and the ℓ th service/DRG—thresholds do not depend on $j \in \Omega$.

2.3 Technology

A production technology that transforms the total operational expenses, $\mathbf{X}^{(\ell)}$, into inpatients, $\mathbf{Y}^{(\ell)}$, can be featured by the technology set $\mathcal{T}^{(\ell)} \subset \mathbb{R}_+ \times \mathbb{R}_+$:

$$\mathcal{T}^{(\ell)} = \left\{ (\mathbf{X}^{(\ell)}, \mathbf{Y}^{(\ell)}) \in \mathbb{R}_+^J \times \mathbb{R}_+^J \mid \mathbf{X}^{(\ell)} \text{ can produce } \mathbf{Y}^{(\ell)} \right\}.$$

The technology set $\mathcal{T}^{(\ell)}$ follows some axioms.

Axiom 1 (Closeness) $\mathcal{T}^{(\ell)}$ is a closed set. Accordingly, the input requirement set, $\mathcal{C}(\mathbf{y}^{(\ell)})$, and the output correspondence set, $\mathcal{P}(\mathbf{x}^{(\ell)})$, are also closed, being $\mathcal{C}(\mathbf{y}^{(\ell)}) = \{ \mathbf{x}^{(\ell)} \in \mathbb{R}_+^J \mid (\mathbf{x}^{(\ell)}, \mathbf{y}^{(\ell)}) \in \mathcal{T}^{(\ell)} \}$ and $\mathcal{P}(\mathbf{x}^{(\ell)}) = \{ \mathbf{y}^{(\ell)} \in \mathbb{R}_+^J \mid (\mathbf{x}^{(\ell)}, \mathbf{y}^{(\ell)}) \in \mathcal{T}^{(\ell)} \}$.

Axiom 2 (No free lunch) Should $x^{j(\ell)} = 0$ and $y^{j(\ell)} > 0$ for $j \in \Omega$, and $(x^{j(\ell)}, y^{j(\ell)}) \notin \mathcal{T}^{(\ell)}$.

Axiom 3 (Free disposability) If $(x^{j(\ell)}, y^{j(\ell)}) \in \mathcal{T}^{(\ell)}$, $\tilde{x}^{j(\ell)} \geq x^{j(\ell)}$, and $\tilde{y}^{j(\ell)} \leq y^{j(\ell)}$, then $(\tilde{x}^{j(\ell)}, y^{j(\ell)}) \in \mathcal{T}^{(\ell)}$, $(x^{j(\ell)}, \tilde{y}^{j(\ell)}) \in \mathcal{T}^{(\ell)}$, and $(\tilde{x}^{j(\ell)}, \tilde{y}^{j(\ell)}) \in \mathcal{T}^{(\ell)}$, (p. 21, Daraio and Simar 2007).

Axiom 4 (Boundedness) The set $\mathcal{A}^{(\ell)}(x^{(\ell)}) = \{ (\tilde{x}^{(\ell)}, y^{(\ell)}) \in \mathcal{T}^{(\ell)} \mid \tilde{x}^{(\ell)} \leq x^{(\ell)} \}$ is bounded for each $x^{(\ell)} \in \mathbb{R}_+^J$ (Mehdiloozad et al. 2014).

Axiom 5 (Normal convexity of $\log \mathcal{T}^{(\ell)}$) Let $\log \mathcal{T}^{(\ell)} \leftarrow \{ (\log \mathbf{X}^{(\ell)}, \log \mathbf{Y}^{(\ell)}) \in \mathbb{R}^J \times \mathbb{R}^J \mid (\mathbf{X}^{(\ell)}, \mathbf{Y}^{(\ell)}) \in \mathcal{T}^{(\ell)} \}$. Let $(\log \tilde{x}^{j(\ell)}, \log \tilde{y}^{j(\ell)}) \in \log \mathcal{T}^{(\ell)}$ and $(\log \tilde{x}^{k(\ell)}, \log \tilde{y}^{k(\ell)}) \in \log \mathcal{T}^{(\ell)}$ for $j, k \in \Omega$. The set $\log \mathcal{T}^{(\ell)}$ is (normal) convex if and only if $(\zeta \log \tilde{x}^{j(\ell)} + (1 - \zeta) \log \tilde{x}^{k(\ell)}, \zeta \log \tilde{y}^{j(\ell)} + (1 - \zeta) \log \tilde{y}^{k(\ell)}) \in \log \mathcal{T}^{(\ell)}$, for any $\zeta \in [0, 1]$.

Axiom 6 (Geometric convexity) $\mathcal{T}^{(\ell)}$ is (geometric or log-) convex if and only if $\log \mathcal{T}^{(\ell)}$ is (normal convex). It is to say that, for any $(\tilde{x}^{j(\ell)}, \tilde{y}^{j(\ell)}) \in \mathcal{T}^{(\ell)}$ and $(\tilde{x}^{k(\ell)}, \tilde{y}^{k(\ell)}) \in \mathcal{T}^{(\ell)}$ with $j, k \in \Omega$, then

$$\forall \zeta \in [0, 1] : \left((\tilde{x}^{j(\ell)})^\zeta (\tilde{x}^{k(\ell)})^{1-\zeta}, (\tilde{y}^{j(\ell)})^\zeta (\tilde{y}^{k(\ell)})^{1-\zeta} \right) \in \mathcal{T}^{(\ell)}.$$

2.4 Concepts and definitions

This subsection provides the mathematical formulations of (1) bandwidths; (2) comparability sets used to constraint the original sample to those hospitals similar to the one whose optimal payment we want to assess and also exhibiting quality levels above a user-defined threshold; and (3) best practice tariff, also called optimal payment (or optimal price), based on a set of coefficients to be optimized and the values observed for the entities belonging to the overall comparability set.

2.4.1 Bandwidths

In general and because hospitals should be comparable, some measures of proximity (or closeness) among them are desirable. Such measures are the so-called bandwidths. The larger the bandwidth, the broader the set of hospitals accepted as comparable with the hospital k , whose optimal payment we want to assess. In view of that, optimal bandwidths are paramount.

The choice of optimal bandwidths, $b_y^{k(\ell)}$, $b_{q_r}^{k(\ell)}$, and $b_{z_c}^{k(\ell)}$, can be debatable since there is a number of ways to compute a bandwidth. All of those ways present advantages but also shortcomings. Particularly, a bandwidth can be either *global* or *local*. A bandwidth is global if it is the same across the whole sample of hospitals; otherwise, it is local.

On the one hand, global bandwidths can easily be computed by using, for instance, the Silverman’s rule of thumb. Let f denote a probability density function. We usually plug-in f by a kernel with order $\gamma > 0$. Hence, the global bandwidth related to a variable V , whose observations in Ω have standard deviation $\hat{\sigma}_V$, is (Silverman 1986):

$$b_V = 2\hat{\sigma}_V J^{-1/(2\gamma+1)} \left(\frac{\sqrt{\pi}(\gamma!)^3 \int_{-\infty}^{+\infty} f^2(u) du}{2\gamma(2\gamma)! \left(\int_{-\infty}^{+\infty} u^2 f(u) du \right)^2} \right)^{1/(2\gamma+1)} \tag{1}$$

According to Ferreira et al. (2017), we should use kernel functions with compact support, i.e., $f(u) > 0$ if $|u| \leq 1$ and $f(u) = 0$ otherwise, and symmetric around k , i.e. γ should be equal to 2 to avoid negative parts on f . In that case, the bandwidth

simplifies to $b_V = 2\hat{\sigma}_V \left(\frac{\sqrt{\pi} \int_{-\infty}^{+\infty} f^2(u) du}{6J \left(\int_{-\infty}^{+\infty} u^2 f(u) du \right)^2} \right)^{1/5}$. For example, if the triweight kernel

is used, we have $b_V \approx 3.62\hat{\sigma}_V J^{-1/5}$. The smaller and the more heterogeneous the sample, the larger the bandwidth.

On the other hand, local bandwidths can be obtained e.g. through the so-called k -Nearest Neighbor method (Daraio and Simar 2007). First, one defines a grid of N units, say $N \in [\epsilon_1 J, \epsilon_2 J]$, with $0 < \epsilon_1 \leq \epsilon_2 \leq 1$. Then, one finds the value of N within such range that minimises the score function $CV(N)$. For the case of bandwidth $b_{q_r}^{k(\ell)}$, the CV function is as follows:

$$CV(N) = \frac{1}{J} \sum_{i=1}^J \log \left(\frac{1}{(J-1)b_{q_r}^{k(\ell)}} \sum_{j=1, j \neq i}^J f \left(\frac{q_{r(\ell)}^{j(\ell)} - q_{r(\ell)}^{i(\ell)}}{b_{q_r}^{k(\ell)}} \right) \right), \ell \in \Psi, r^{(\ell)} \in \Gamma^{(\ell)}, \quad (2)$$

where f denotes a univariate kernel function. Therefore, $b_{q_r}^{k(\ell)}$ is the local bandwidth associated with hospital k chosen such that there are N points verifying $|q_{r(\ell)}^{j(\ell)} - q_{r(\ell)}^{i(\ell)}| \leq b_{q_r}^{k(\ell)}, r^{(\ell)} \in \Gamma^{(\ell)}$. Naturally, we can specify similar local bandwidths for the case of the output and the operational environment data. There are other alternatives for the local bandwidths' formulation, including the (data-driven) least squares cross validation procedure to minimise the integrated squared error (Bădin et al. 2010; Hall et al. 2004; Li and Racine 2007).

2.4.2 Comparability sets

The idea underlying our approach is to restrict the original set of hospitals Ω based on constraints related to size (because of potential economies of scale), quality, access, and environment. To simplify the exposition, we will assume that access dimensions can be included into the group of quality variables and handled like them. Therefore, we may introduce three comparability sets that are subsets of Ω .

The first comparability set presented regards the size of the health care provider. It is important to guarantee that best practices related to hospital k have similar operations' scale as this one. For that reason, we constrain the set of admissible best practices for k by using the concept of bandwidth and centring the observations associated with that comparability set on $y^{k(\ell)}$, which is assumed to be the proxy for the size of k .

Definition 1 (*Size-related comparability set, $\Omega_y^{k(\ell)}$*) Given a service or DRG $\ell \in \Psi$, the size-related comparability set for hospital $k \in \Omega$ is defined as the set of hospitals whose sizes, as measured by the output level $\mathbf{Y}^{(\ell)}$, are close to the dimension of hospital k , i.e., $y^{k(\ell)}$. Let the bandwidth $b_y^{k(\ell)}$ be the closeness measure. Hence, $\Omega_y^{k(\ell)}$ is as follows:

$$\Omega_y^{k(\ell)} = \left\{ \bigcup_{j \in \Omega} \text{s.t. } y^{k(\ell)} - b_y^{k(\ell)} \leq y^{j(\ell)} \leq y^{k(\ell)} + b_y^{k(\ell)} \right\}. \quad (3)$$

In the following definition, we introduce the so-called quality-related comparability set, which, as before, requires a bandwidth. However, it also needs a user predefined threshold per quality measure so as to ensure that no unit with low levels of quality (including access to health care services) can be addressed as a potential best practice for the hospital k whose optimal payment is being computed. In fact, we could achieve lower levels of resources consumption at the cost of reducing the quality of supplied services, endangering the societal mission of health care providers. In practice, we would be interested on best practices that would

outperform k in terms of both quality and access, i.e. verifying $q_{r^{(\ell)}}^{j^{(\ell)}} \geq q_{r^{(\ell)}}^{k^{(\ell)}}$ for any $r^{(\ell)} = 1, \dots, R^{(\ell)}$. Nonetheless, we note that hospitals with very high quality levels may sometimes be technically inefficient on resources consumption for the quantity of services delivered. It means that, if hospital k reaches the maximum quality within the dataset, then, according to our previous two conditions, no other hospital could be a benchmark for k even if it would be more technically efficient than the latter. This constitutes a problem because the optimal payment for k would not be Pareto-efficient, jeopardising the health system sustainability. In practice, usually one can reduce a little the quality levels of k by using bandwidths, if it results on meaningful cost savings and as long as the best practices found for k would present quality levels at least above the user-defined threshold.

Definition 2 (*Quality-related comparability set, $\Omega_q^{k^{(\ell)}}$*) Given a service or DRG $\ell \in \Psi$, the quality-related comparability set for hospital $k \in \Omega$ is defined as the set of hospitals whose quality dimensions $\mathbf{Q}^{(\ell)}$ are close to the quality observed for hospital k (or even above it), $q_{r^{(\ell)}}^{k^{(\ell)}}$, and simultaneously larger than the user-defined threshold $t_{q_r}^{(\ell)}$ for all $r^{(\ell)} \in \Gamma^{(\ell)}, \ell \in \Psi$. Let the bandwidth $b_{q_r}^{k^{(\ell)}}$ be the closeness measure. Hence, $\Omega_q^{k^{(\ell)}}$ is as follows:

$$\Omega_q^{k^{(\ell)}} = \left\{ \bigcup_{j \in \Omega} \text{s.t.} \begin{array}{l} (q_{r^{(\ell)}}^{k^{(\ell)}} - b_{q_r}^{k^{(\ell)}} \leq q_{r^{(\ell)}}^{j^{(\ell)}} \leq q_{r^{(\ell)}}^{k^{(\ell)}} + b_{q_r}^{k^{(\ell)}}) \vee (q_{r^{(\ell)}}^{j^{(\ell)}} \geq q_{r^{(\ell)}}^{k^{(\ell)}}), r^{(\ell)} = 1, \dots, R^{(\ell)} \\ \wedge \\ (q_{r^{(\ell)}}^{j^{(\ell)}} \geq t_{q_r}^{(\ell)}), r^{(\ell)} = 1, \dots, R^{(\ell)} \end{array} \right\}. \tag{4}$$

The definition of an environment-related comparability set is similar to the one associated with size (vide supra).

Definition 3 (*Environment-related comparability set, $\Omega_z^{k^{(\ell)}}$*) Given a service or DRG $\ell \in \Psi$, the environment-related comparability set for hospital $k \in \Omega$ is defined as the set of hospitals whose environment $\mathbf{Z}^{(\ell)}$ are close to the conditions observed for hospital k , $z_{c^{(\ell)}}^{k^{(\ell)}}$, for all $c^{(\ell)} \in \Phi^{(\ell)}, \ell \in \Psi$. Let the bandwidth $b_{z_c}^{k^{(\ell)}}$ be the closeness measure. Hence, $\Omega_z^{k^{(\ell)}}$ is as follows:

$$\Omega_z^{k^{(\ell)}} = \left\{ \bigcup_{j \in \Omega} \text{s.t.} \begin{array}{l} z_{c^{(\ell)}}^{k^{(\ell)}} - b_{z_c}^{k^{(\ell)}} \leq z_{c^{(\ell)}}^{j^{(\ell)}} \leq z_{c^{(\ell)}}^{k^{(\ell)}} + b_{z_c}^{k^{(\ell)}}, c^{(\ell)} = 1, \dots, C^{(\ell)} \end{array} \right\}. \tag{5}$$

Using these three definitions, we can construct the overall comparability set, $\Omega^{k^{(\ell)}}$, which is a subset of Ω .

Definition 4 (Overall comparability set, $\Omega^{k(\ell)}$) The overall comparability set related to service/DRG $\ell \in \Psi$ and the hospital $k \in \Omega$, results from the intersection of the three previously defined comparability sets:

$$\Omega^{k(\ell)} = \Omega_y^{k(\ell)} \cap \Omega_q^{k(\ell)} \cap \Omega_z^{k(\ell)}. \tag{6}$$

Figure 1 exemplifies the achievement of the overall comparability set associated with the hospital $k = 4$ as well as the corresponding comparability sets. In this example, there are fifteen hospital in the dataset; they could be potential best practices for k if features like size, quality, and environment would be disregarded. However, we have considered one quality and another environment dimensions. Let the blue and the green areas (left) define the the size- and the environment-related comparability sets for $k = 4$, respectively. The intersection of them will result on a subset composed of the following hospitals: $\Omega_y^{k(\ell)} \cap \Omega_z^{k(\ell)} = \{3, 4, 7, 14, 15\}$. The red line (right) identified the user-defined threshold. Arrows explain that the quality of potential best practices should be larger than or equal to 30. However, the quality observed for hospital k is equal to 70 and the bandwidth is 22, meaning that the quality of those best practices for hospital k should, in fact, be larger than 48. Accordingly, we have $\Omega_y^{k(\ell)} \cap \Omega_q^{k(\ell)} = \{1, 4, 7, 15\}$ and $\Omega^{k(\ell)} = \{4, 7, 15\}$. Therefore, the set of potential best practices related to k is composed of itself and of two more hospitals: 7 and 15. In the next subsection we will explain how the best practice tariff of k can be obtained from $\Omega^{k(\ell)}$. To do so, we have to introduce the following notation:

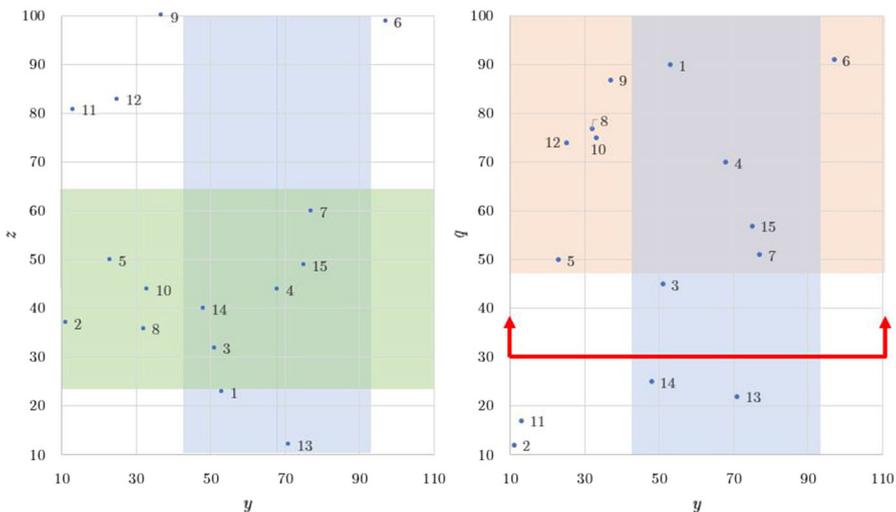


Fig. 1 On defining the overall comparability set and its subsets (color figure online)

- J^k is the length of the list $\Omega^{k(\ell)}$;
- $\mathbf{X}^{k(\ell)}, \ell \in \Psi$, is the vector of total operational expenses associated with the J^k hospitals in $\Omega^{k(\ell)}$;
- $\mathbf{Y}^{k(\ell)}, \ell \in \Psi$, is the vector of total in/outpatients handled by the J^k hospitals in $\Omega^{k(\ell)}$.

In the previous example, $J^k = 3$, $\mathbf{X}^{k(\ell)} = (x^{4(\ell)}, x^{7(\ell)}, x^{15(\ell)})^\top$, and $\mathbf{Y}^{k(\ell)} = (y^{4(\ell)}, y^{7(\ell)}, y^{15(\ell)})^\top$.

2.4.3 Best practice tariffs

The *best practice tariff* of hospital $k \in \Omega$ and service/DRG $\ell \in \Psi$ is, by definition, the paid price per medical act which is assessed using the information of benchmarks (best practices), which, in turn, are at least as good performer as hospital k in the provision of service/DRG ℓ . Because hospitals present distinct technologies and face heterogeneous environments, they must be comparable using data related to the inn- (size, complexity of inpatients) and the out-operational environment (demographics, epidemiology). Likewise, poor quality hospitals should not be considered potential best practices for k . For that reason, we have created the concept of comparability sets associated with this hospital. Using it we can formulate the optimal (best practice based) tariff paid to hospital k .

Definition 5 (*O-order best practice tariff paid to hospital k and service/DRG ℓ , $\mathcal{P}^{k(\ell, \mathcal{O})}$*) Let $k \in \Omega$ be a hospital, $\ell \in \Psi$ a service or a DRG, and $\Omega^{k(\ell)}$ the overall comparability related to the former two. The weighted Hölder (or power/generalized) mean with order \mathcal{O} of a vector $\mathbf{V} = (V_1, \dots, V_i, \dots, V_n)^\top$ and weights $\boldsymbol{\mu} = (\mu_1, \dots, \mu_i, \dots, \mu_n)^\top$, verifying $\sum_{i=1}^n \mu_i = 1$, is

$$\mathcal{H}(\mathbf{V}, \boldsymbol{\mu}, \mathcal{O}) = \left\langle \boldsymbol{\mu}, (\mathbf{V})^\mathcal{O} \right\rangle^{1/\mathcal{O}}, \tag{7}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner-product of two vectors. The \mathcal{O} -order best practice tariff that should be paid to hospital k is the relationship between the optimal costs and the number of in/outpatients for that hospital. We assume that optimal expenses can be modelled as weighted Hölder averages with order \mathcal{O} and weights $\boldsymbol{\mu}^k$:

$$\mathcal{P}^{k(\ell, \mathcal{O})} = \frac{\mathcal{H}(\mathbf{X}^{k(\ell)}, \boldsymbol{\mu}^k, \mathcal{O})}{y^{k(\ell)}}. \tag{8}$$

In this case, we assume that the number of in/outpatients is out of control by the hospital managers. If this is not the case, we simply have to replace the denominator of Eq. (8) by $\mathcal{H}(\mathbf{Y}^{k(\ell)}, \boldsymbol{\mu}^k, \mathcal{O})$.

The \mathcal{O} -order best practice tariff paid to hospital k for the service/DRG ℓ in the previous example is, according to Eq. (8):

$$\mathcal{P}^{k(\ell, \mathcal{O})} = \frac{\left(\mu_1^k \cdot (x^{4(\ell)})^{\mathcal{O}} + \mu_2^k \cdot (x^{7(\ell)})^{\mathcal{O}} + \mu_3^k \cdot (x^{15(\ell)})^{\mathcal{O}} \right)^{1/\mathcal{O}}}{y^{4(\ell)}}.$$

The problem associated with Eq. (8) is that, in general, we do not know the weights μ^k to use. Therefore, instead of imposing quantities with little or even no empirical support, these weights should be optimized, for instance, using linear programming tools. In the next subsections, we discuss how cost savings can arise.

3 A new best practice tariff-based tool

In the last section, we have defined the overall comparability set and explained how the best practice tariffs can be expressed in terms of some coefficients that should be optimized using benchmarking tools. In our case, we develop and explore a model similar to DEA that constructs a log-linear piecewise frontier and which is more flexible than DEA itself. This flexible model is named multiplicative (or log-) DEA. Those coefficients μ^k can be optimized using log-linear programming model, as detailed in Sect. 3.3. This one extends the work of Banker and Maindiratta (1986). Some properties of the new model are studied in Sect. 3.4. The reader should be aware of the problem of multiple/degenerate solutions associated with those coefficients, which can be critical to our optimal payments. Therefore, we extend our DEA model to solve the problem of multiple solutions (Sect. 3.5). In some circumstances, this model may become infeasible; hence, we propose a strategy to overcome this problem, vide Sect. 3.6. Finally, Sect. 3.7 presents a simple step-by-step procedure to simplify the exposition and to sum up our approach to optimize payments.

3.1 Past research on multiplicative DEA

The proposed way of optimizing tariffs imposes the achievement of Pareto-optimal weights μ^k . DEA can be very useful in such a situation. Since its foundation, it seems to be the widest employed model for efficiency assessment, especially in health care (Hollingsworth 2008). Because of the underlying convexity, the DEA formulation proposed by Banker et al. (1984) requires non-increasing marginal products. However, according to the classical production theory, there are three main stages characterizing the consumption of resources and the associated quantity of produced outputs (vide Fig. 2). For input quantities smaller than a , the fixed input is being utilized more effectively as the variable input increases (Kao 2017), which is to say that the marginal product increases accordingly (Henderson and Quandt 1980). Before b , the production function (red line) is non-concave and the production possibility set cannot be convex (Banker and Maindiratta 1986). If the standard DEA model is used to estimate this frontier (dashed line [Ob]), then some efficient observations exploiting gains from increasing specialization with larger scale sizes (i.e., close to a) would be rated as inefficient.

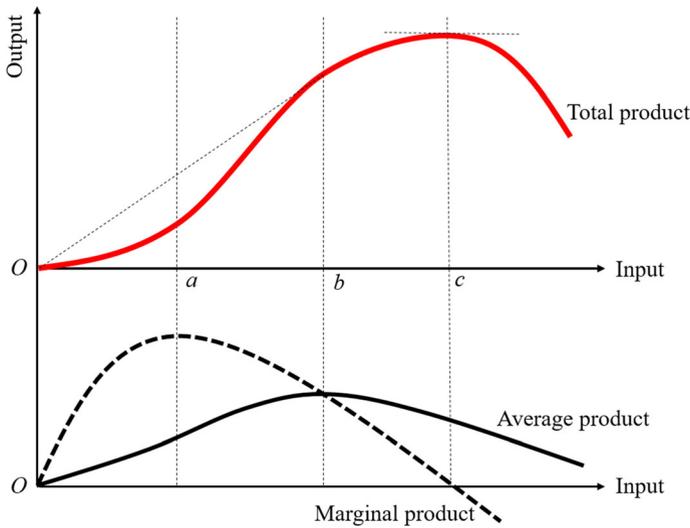


Fig. 2 The three main stages characterizing the consumption of resources and the associated quantity of produced outputs, according to the classical production theory (color figure online)

An immediate conclusion is that the convexity assumption must be smoothed to account for the case of increasing marginal products. An alternative is the so-called FDH (Deprins et al. 1984), which constructs a frontier with a *staircase* nature that, in turn, is discontinuous in some points and not differentiable (note that we do not require that the production function must be differentiable). Situations including non-variable returns to scale and imposition of restrictions over multipliers through FDH (Ferreira and Marques 2017) require solving a mixed-integer linear program (Agrell and Tind 2001), which can be difficult to implement in some solvers. Another alternative is the multiplicative DEA, also named log-DEA or DEA-Cobb–Douglas (whose mathematical details will be detailed soon).

Back nearly 40 years ago, Banker et al. (1984) constructed a log-linear model in which outputs are modelled following a Cobb–Douglas function. Under such a framework, outputs do not compete for the inputs, although they do in most of the empirical scenarios, including hospitals. Therefore, this model is not applicable in these situations. Charnes et al. (1982) suggested a multiplicative efficiency measure, but which is not consistent with the postulated underlying production of Banker and Maindiratta (1986). One year later, the same authors modified their model to account for non-radial inefficiency sources. Nonetheless, they did not examine the characteristics of the production technology, ignoring concepts like returns to scale, optimal scale, and non-competing outputs. Later on, Banker and Maindiratta (1986) proposed a radial log-DEA that account for competing outputs. These authors have verified that their model outperforms the piecewise linear DEA model of Banker et al. (1984) in terms of rates of substitution of inputs, and especially whenever the production technology frontier exhibits non-concavity in some regions. In the present work, we follow the model of Banker and Maindiratta, which has not seen

further meaningful developments and enhancements, despite its advantages (vide infra).

Chang and Guh (1995) criticize the remedy of Suayoshi and Chang (1989) to transform the efficiency score associated with the multiplicative model of Charnes et al. (1982) into unit invariant. They propose another model, which according to Seiford and Zhu (1998) reduces itself to the proposal of Charnes et al. This is because the model of Chang and Guh is for the case of constant output and linearly homogeneous production technology. The authors fail to recognize that the constraint associated with the convexity of the DEA model of Banker et al. is the basis for units invariance. Therefore, according to Seiford and Zhu, the distance efficiency measure of Chang and Guh is of no value.

Marginal products are usually related to returns to scale (Ferreira and Marques 2018a). Since the standard DEA fails to identify the production regions of increasing marginal products, multiplicative DEA models have been used to analyse returns to scale and most productive scale size; vide the works of Zarepisheh et al. (2010) as well as Davoodi et al. (2015).

Only a few applications of the multiplicative DEA model have been published. We point out three: Emrouznejad and Cabanda (2010), Tofallis (2014) and Valadkhani et al. (2016). The former authors integrated the so-called Benefit of Doubt (Cherchye et al. 2011) and the log-DEA to construct a composite indicator associated with six financial ratios and twenty-seven UK industries. They confirm that the multiplicative model is more robust than the standard DEA model. In the same vein of Emrouznejad and Cabanda, Tofallis also constructed composite indicators with multiplicative aggregation and remarked that the log-DEA avoids the *zero weight problem* of the standard DEA. Meanwhile, Valadkhani et al. proposed a multiplicative extension of environmental DEA models, accounting for weakly disposable undesirable outputs, and used it to measure efficiency changes in the world's major polluters. It is interesting to note that the adopted model is very close to the one proposed by Mehdiloozad et al. (2014). These authors developed a generalized multiplicative directional distance function as a comprehensive measure of technical efficiency, accounting for all types of slacks and satisfying several desirable properties.

Previous studies have considered that every observation is enveloped by the log-linear frontier and, for that reason, no infeasibility may arise. If this is not the case, projecting super-efficient observations into the frontier can be infeasible, mostly because of the undertaken path (model direction) (Chen et al. 2013). Infeasibility may, then, occur if the hospital whose tariff we are optimizing is not enveloped by the frontier. However, should it belong to the reference set used to construct such a frontier, and multiple solutions may be achieved, meaning that the optimal tariff could be non-unique. Neither of these two problems has been properly addressed in the literature with respect to either the multiplicative or log-DEA or the payments optimization.

3.2 Advantages of using log-DEA instead of the standard DEA

The advantages of log-DEA can be summarized as follows:

- log-DEA is more flexible than DEA, allowing for increasing marginal products (Kao 2017);
- log-DEA allows for outputs competing for inputs, as happens in most of empirical situations, including health care, (Banker and Maindiratta 1986);
- log-DEA outperforms DEA in terms of rates of substitution of inputs (Banker and Maindiratta 1986);
- log-DEA outperforms DEA when the (unknown) production technology frontier exhibits non-concavity regions, (Emrouznejad and Cabanda 2010);
- log-DEA avoids the *zero weight problem* of the standard DEA (Tofallis 2014);
- log-DEA keeps the production technology globally and geometrically convex (or non-convex under the arithmetic definition), which is a more natural solution. Indeed, there is no reason to believe that the production technology must be (piecewise) linear and the associated set must be convex. Frontiers related to non-convex sets are always consistent, even if the technology is convex (Daraio and Simar 2007). Because of this property, log-DEA can be used to estimate economies of scope (vide a comprehensive discussion in Ferreira et al. 2017);
- log-DEA mitigates the inefficiency overestimation resulting from infeasible regions constructed by two or more very distant efficient observations (Tiedemann et al. 2011);
- log-DEA classifies all non-dominated observations as efficient, unlikely DEA.

3.3 Performance assessment through nonparametric log-linear benchmarking tools

A very important feature on defining the optimal (best practice based) tariff is the vector $\mu^k = (\mu_1^k, \dots, \mu_i^k, \dots, \mu_{j^k}^k)^\top$. As previously mentioned, the components of this vector should be optimized in the Pareto sense. This subsection explains how this can be done for those components. We will describe the construction of a log-linear piecewise frontier function from the J^k hospitals composing $\Omega^{k(\ell)}$. This frontier contains the potential benchmarks related to hospital k and the service/DRG ℓ .

First of all, we have to define the concepts of *target* and *target setting*.

Definition 6 (*Targets of hospital k*) Targets are the optimal values associated with hospital k for inputs and outputs, $\tilde{x}^{k(\ell)}$ and $\tilde{y}^{k(\ell)}$ respectively. Targets must verify the conditions:

$$\begin{cases} \tilde{x}^{k(\ell)} \leq x^{k(\ell)}, \\ \tilde{y}^{k(\ell)} \geq y^{k(\ell)}. \end{cases} \tag{9}$$

Definition 7 (*Target setting for hospital k*) Targets are mathematically defined by the \mathcal{O} -order Hölder average:

$$\begin{cases} x^{k(\ell)} = \langle \boldsymbol{\mu}^k, (\mathbf{X}^{k(\ell)})^{\mathcal{O}} \rangle^{1/\mathcal{O}}, \\ y^{k(\ell)} = \langle \boldsymbol{\mu}^k, (\mathbf{Y}^{k(\ell)})^{\mathcal{O}} \rangle^{1/\mathcal{O}}. \end{cases} \tag{10}$$

In the case of $\mathcal{O} = 0$ (Banker and Maindiratta 1986), Eq. (10) reduces to:

$$\begin{cases} x^{k(\ell)} = \prod_{j \in \Omega^{k(\ell)}} (x^{j(\ell)})^{\mu_j^k}, \\ y^{k(\ell)} = \prod_{j \in \Omega^{k(\ell)}} (y^{j(\ell)})^{\mu_j^k} \end{cases} \iff \begin{cases} \log x^{k(\ell)} = \sum_{j \in \Omega^{k(\ell)}} \mu_j^k \log x^{j(\ell)}, \\ \log y^{k(\ell)} = \sum_{j \in \Omega^{k(\ell)}} \mu_j^k \log y^{j(\ell)}. \end{cases} \tag{11}$$

Because the logarithmic function is monotonically increasing, Eq. (9) can be rewritten as:

$$\log x^{k(\ell)} \leq \log x^{k(\ell)} \wedge \log y^{k(\ell)} \geq \log y^{k(\ell)},$$

and, by Eq. (11), we have:

$$\sum_{j \in \Omega^{k(\ell)}} \mu_j^k \log x^{j(\ell)} \leq \log x^{k(\ell)} \wedge \sum_{j \in \Omega^{k(\ell)}} \mu_j^k \log \frac{1}{y^{j(\ell)}} \leq \log \frac{1}{y^{k(\ell)}}. \tag{12}$$

Inequalities in (12) can be transformed into equations by using (nonnegative) slacks. These slacks can be decomposed into controllable (vector $\mathbf{D}^k = (d_1^k, d_2^k)$ with $d_1^k, d_2^k \geq 0$) and uncontrollable quantities (coefficients β^k, s_1^k , and s_2^k). Hence,

$$\begin{aligned} \sum_{j \in \Omega^{k(\ell)}} \mu_j^k \log x^{j(\ell)} + d_1^k \beta^k + s_1^k &= \log x^{k(\ell)} \wedge \sum_{j \in \Omega^{k(\ell)}} \mu_j^k \log \frac{1}{y^{j(\ell)}} + d_2^k \beta^k + s_2^k = \log \frac{1}{y^{k(\ell)}} \iff \\ \iff \langle \boldsymbol{\mu}^k, \log \mathbf{X}^{k(\ell)} \rangle + d_1^k \beta^k + s_1^k &= \log x^{k(\ell)} \wedge \langle \boldsymbol{\mu}^k, \log \frac{\mathbf{1}}{\mathbf{Y}^{k(\ell)}} \rangle + d_2^k \beta^k + s_2^k = \log \frac{1}{y^{k(\ell)}}, \end{aligned} \tag{13}$$

where $\mathbf{1}/\mathbf{Y}^{k(\ell)}$ is the Hadamard’s componentwise division between two vectors, $\mathbf{1} = (1, \dots, 1)$ and $\mathbf{Y}^{k(\ell)}$, and the logarithmic function applies to each component of vectors, i.e., $\log \mathbf{V} = (\log V_1, \dots, \log V_{J^k})^\top$.

These constraints create a log-linear piecewise frontier function assuming that $\langle \sqrt{\boldsymbol{\mu}^k}, \sqrt{\boldsymbol{\mu}^k} \rangle = 1$ and all components of $\boldsymbol{\mu}^k$ are nonnegative. Let us find the maximum value of the uncontrollable components using a linear program. In this case, we first optimize β^k and, once its value has been achieved, we maximize both s_1^k and s_2^k , which justifies the adoption of a non-Archimedean, ε . The linear program is as follows:

$$\begin{aligned}
 & \text{[log-DEA primal] maximize } \beta^k + \varepsilon(s_1^k + s_2^k) \\
 & \text{subject to:}
 \end{aligned}
 \tag{14a}$$

$$\left\langle \boldsymbol{\mu}^k, \log \mathbf{X}^{k(\ell)} \right\rangle + d_1^k \beta^k + s_1^k = \log x^{k(\ell)},
 \tag{14b}$$

$$\left\langle \boldsymbol{\mu}^k, \log \frac{\mathbf{1}}{\mathbf{Y}^{k(\ell)}} \right\rangle + d_2^k \beta^k + s_2^k = \log \frac{1}{y^{k(\ell)}},
 \tag{14c}$$

$$\langle \sqrt{\boldsymbol{\mu}^k}, \sqrt{\boldsymbol{\mu}^k} \rangle = 1,
 \tag{14d}$$

$$\boldsymbol{\mu}^k \geq 0,
 \tag{14e}$$

$$s_1^k, s_2^k \geq 0.
 \tag{14f}$$

Note that β^k is not an efficiency score per se. In fact, the efficiency of hospital k in service/DRG ℓ can be defined as the relationship between optimal targets and observations for both inputs and outputs.

Definition 8 (*Efficiency score, $\theta^{k(\ell)}$*) Using the concept of targets, we define the efficiency score of hospital k in service/DRG ℓ as Portela and Thanassoulis (2006):

$$\theta^{k(\ell)} = \frac{\bar{x}^{k(\ell)} / x^{k(\ell)}}{\bar{y}^{k(\ell)} / y^{k(\ell)}}.
 \tag{15}$$

Because of Definition 7 and the assumption $\mathcal{O} = 0$, which returns a log-linear piecewise frontier, Eq. (15) becomes:

$$\theta^{k(\ell)} = \frac{\frac{1}{x^{k(\ell)}} \prod_{j \in \Omega^{k(\ell)}} (x^{j(\ell)})^{\mu_j^k}}{\frac{1}{y^{k(\ell)}} \prod_{j \in \Omega^{k(\ell)}} (y^{j(\ell)})^{\mu_j^k}} = \frac{y^{k(\ell)}}{x^{k(\ell)}} \prod_{j \in \Omega^{k(\ell)}} \left(\frac{x^{j(\ell)}}{y^{j(\ell)}} \right)^{\mu_j^k},
 \tag{16}$$

where $\boldsymbol{\mu}^k$ is obtained from Model (14). Using Eq. (16) and after some easy algebraic manipulations, we can associate $\theta^{k(\ell)}$ with β^k :

$$\log \frac{1}{\theta^{k(\ell)}} = \beta^k (d_1^k + d_2^k) + s_1^k + s_2^k.$$

This relationship allows us to conclude: (a) if $\beta^k = s_1^k = s_2^k = 0$, then $1/\theta^{k(\ell)} = 1 \Leftrightarrow \theta^{k(\ell)} = 1$, i.e., hospital k is efficient; (b) if $\beta^k \neq 0 \vee s_1^k + s_2^k > 0$, then k is not efficient because $\theta^{k(\ell)} < 1$. Of course, k can outperform the frontier constructed by $\Omega^{k(\ell)}$, which means that $\beta^k < 0$ and $\theta^{k(\ell)} > 1$, indicating super-efficiency.

3.4 Some properties of the developed model

Model (14) exhibits several important properties: efficiency requirement, scale invariance, nonconvexity, strict monotonicity, and directional nature. Notwithstanding, it is not translation invariant and it may exhibit multiple optima solutions. Because of the latter, the optima of log-DEA primal are not sufficient to compute the optimal paid price, $\mathcal{P}^{k(\ell, \mathcal{O})}$, $\mathcal{O} = 0$, through the optimized weights μ^k . We will explore an alternative to provide the unique solution for those weights—the so-called *primal–dual* approach for multiple optima.

Proposition 1 (Efficiency requirement) *If either $k \in \Omega^{k(\ell)}$ or $\exists j \in \Omega^{k(\ell)} : (x^{j(\ell)}, y^{j(\ell)}) \succ (x^{k(\ell)}, y^{k(\ell)})$,² then $\theta^{k(\ell)}$ is bounded by a non-Archimedean, $\varepsilon \sim 0$, and by 1, i.e., the log-DEA primal model verifies the efficiency requirement property.*

Proof Under the condition $k \in \Omega^{k(\ell)}$ or the condition $\exists j \in \Omega^{k(\ell)} : (x^{j(\ell)}, y^{j(\ell)}) \succ (x^{k(\ell)}, y^{k(\ell)})$, all slacks $(d^k \beta^k + s^k)$ are nonnegative, and the optimal objective is nonnegative as well. Hence, $\log \frac{1}{\theta^{k(\ell)}} \geq 0 \Leftrightarrow \frac{1}{\theta^{k(\ell)}} \geq \exp 0 = 1 \Leftrightarrow \theta^{k(\ell)} \leq 1$. □

Proposition 2 (Maximum value of β^k) *Let $\mathcal{X}^{k(\ell)} = \frac{1}{d_1^k} \log \frac{x^{k(\ell)}}{\min_{j \in \Omega^{k(\ell)}} x^{j(\ell)}}$ and $\mathcal{Y}^{k(\ell)} = \frac{1}{d_2^k} \log \frac{1}{y^{k(\ell)} \min_{j \in \Omega^{k(\ell)}} \frac{1}{y^{j(\ell)}}}$. Therefore, $\beta^k \leq \min\{\mathcal{X}^{k(\ell)}, \mathcal{Y}^{k(\ell)}\}$.*

Proof From the definition of the log-DEA primal model, we know that $\langle \mu^k, \log \mathbf{X}^{k(\ell)} \rangle \leq \log x^{k(\ell)} - d_1^k \beta^k$ and $\langle \mu^k, \log 1/\mathbf{Y}^{k(\ell)} \rangle \leq \log 1/y^{k(\ell)} - d_2^k \beta^k$. Moreover, from the Hölder mean it is obvious that $\langle \mu^k, \log \mathbf{X}^{k(\ell)} \rangle \geq \min_{j \in \Omega^{k(\ell)}} \log x^{j(\ell)} = \log \min_{j \in \Omega^{k(\ell)}} x^{j(\ell)}$ and $\langle \mu^k, \log 1/\mathbf{Y}^{k(\ell)} \rangle \geq \min_{j \in \Omega^{k(\ell)}} \log 1/y^{j(\ell)} = \log \min_{j \in \Omega^{k(\ell)}} 1/y^{j(\ell)}$. Therefore,

$$\begin{aligned} \log x^{k(\ell)} - d_1^k \beta^k &\geq \log \min_{j \in \Omega^{k(\ell)}} x^{j(\ell)} \wedge \log 1/y^{k(\ell)} - d_2^k \beta^k \geq \log \min_{j \in \Omega^{k(\ell)}} 1/y^{j(\ell)} \iff \\ \iff \beta^k &\leq \frac{\log x^{k(\ell)} - \log \min_{j \in \Omega^{k(\ell)}} x^{j(\ell)}}{d_1^k} \wedge \beta^k \leq \frac{\log 1/y^{k(\ell)} - \log \min_{j \in \Omega^{k(\ell)}} 1/y^{j(\ell)}}{d_2^k} \iff \\ \iff \beta^k &\leq \underbrace{\frac{1}{d_1^k} \log \frac{x^{k(\ell)}}{\min_{j \in \Omega^{k(\ell)}} x^{j(\ell)}}}_{=\mathcal{X}^{k(\ell)}} \wedge \beta^k \leq \underbrace{\frac{1}{d_2^k} \log \frac{1}{y^{k(\ell)} \min_{j \in \Omega^{k(\ell)}} 1/y^{j(\ell)}}}_{=\mathcal{Y}^{k(\ell)}} \iff \beta^k \\ &\leq \min\{\mathcal{X}^{k(\ell)}, \mathcal{Y}^{k(\ell)}\}. \end{aligned}$$

□

² $A \succ B$ means that A outperforms (envelops) B.

Because of Proposition 2 and the maximization of β^k in the objective function, we know that $\beta^k = \min\{\mathcal{X}^{k(\ell)}, \mathcal{Y}^{k(\ell)}\}$. Hence, the log-DEA primal can be simplified as follows:

$$\begin{aligned} &[\text{log-DEA primal (simplified)}] \quad \text{maximize } s_1^k + s_2^k \\ &\quad \text{subject to:} \end{aligned} \tag{17a}$$

$$\langle \boldsymbol{\mu}^k, \log \mathbf{X}^{k(\ell)} \rangle + s_1^k = \log x^{k(\ell)} - d_1^k \min\{\mathcal{X}^{k(\ell)}, \mathcal{Y}^{k(\ell)}\}, \tag{17b}$$

$$\langle \boldsymbol{\mu}^k, \log \frac{\mathbf{1}}{\mathbf{Y}^{k(\ell)}} \rangle + s_2^k = \log \frac{1}{y^{k(\ell)}} - d_2^k \min\{\mathcal{X}^{k(\ell)}, \mathcal{Y}^{k(\ell)}\}, \tag{17c}$$

$$\langle \sqrt{\boldsymbol{\mu}^k}, \sqrt{\boldsymbol{\mu}^k} \rangle = 1, \tag{17d}$$

$$\boldsymbol{\mu}^k \geq 0, \tag{17e}$$

$$s_1^k, s_2^k \geq 0. \tag{17f}$$

Proposition 3 (Scale invariance) *If \mathbf{D}^k does not depend on data, then the log-DEA primal model is scale invariant.*

Proof Let us assume that \mathbf{D}^k does not depend on data. Consider the transformation of $x^{k(\ell)}$ using a scalar $\xi \in \mathbb{R}$: $x^{k(\ell)} \mapsto \xi x^{k(\ell)}$. The first constraint of log-DEA primal becomes:

$$\begin{aligned} &\langle \boldsymbol{\mu}^k, \log \xi \mathbf{X}^{k(\ell)} \rangle + d_1^k \beta^k + s_1^k = \log \xi x^{k(\ell)} \Leftrightarrow \langle \boldsymbol{\mu}^k, \log \xi \rangle + \langle \boldsymbol{\mu}^k, \log \mathbf{X}^{k(\ell)} \rangle + d_1^k \beta^k \\ &\quad + s_1^k = \log \xi + \log x^{k(\ell)} \Leftrightarrow \\ &\Leftrightarrow \log \xi + \langle \boldsymbol{\mu}^k, \log \mathbf{X}^{k(\ell)} \rangle + d_1^k \beta^k + s_1^k = \log \xi + \log x^{k(\ell)} \Leftrightarrow \langle \boldsymbol{\mu}^k, \log \mathbf{X}^{k(\ell)} \rangle \\ &\quad + d_1^k \beta^k + s_1^k = \log x^{k(\ell)}, \end{aligned}$$

where $\langle \boldsymbol{\mu}^k, \log \xi \mathbf{X}^{k(\ell)} \rangle = \langle \boldsymbol{\mu}^k, \log \xi \rangle + \langle \boldsymbol{\mu}^k, \log \mathbf{X}^{k(\ell)} \rangle$ results from the distributive property of the inner product and $\langle \boldsymbol{\mu}^k, \log \xi \rangle = \log \xi$ results from the condition $\langle \sqrt{\boldsymbol{\mu}^k}, \sqrt{\boldsymbol{\mu}^k} \rangle = 1$. Thus, the first constraint of log-DEA primal is recovered. However, this is only true if \mathbf{D}^k does not depend on data; otherwise, the equivalences above do not necessarily hold. Since the same does apply for the output $y^{k(\ell)}$, we conclude this proof. \square

Proposition 4 (Convexity) *The frontier constructed by log-DEA primal model is not convex in the original space of variables. This is because $\log \mathcal{T}^{(\ell)}$ is (normal) convex, meaning that $\mathcal{T}^{(\ell)}$ is log-convex.*

Proof It is not difficult to conclude that, in the log-space, the frontier is convex. Now, let us consider a linear facet of the frontier in the log-space. Hence, we have

$\log y = a \log x + b$ or, equivalently, $y = x^a \exp b$, which is not linear. That is, in the original space of variables, the constructed frontier is not convex. \square

Proposition 5 (Strict monotonicity) *Log-DEA is strictly monotone. $\theta^{k(\ell)}$ strictly decreases with the increase of slacks.*

Proof The objective function of log-DEA strictly increases with slacks, and so does $\log(1/\theta^{k(\ell)})$. Because of the monotonicity property of logarithmic functions, $1/\theta^{k(\ell)}$ also increases with slacks, which means that $\theta^{k(\ell)}$ decreases with the increase of slacks. In fact, we have $\theta^{k(\ell)} = \exp(-\beta^k(d_1^k + d_2^k) - s_1^k - s_2^k)$. This is in line with the efficiency requirement property. \square

Proposition 6 (Directional nature) *Log-DEA is a radial directional model.*

Proof Log-DEA is a radial model because of the parameter β^k that imposes the projection of hospital k in the frontier constructed using $\Omega^{k(\ell)}$. It is directional because of the vector \mathbf{D}^k , which controls for the path used to project k (Fukuyama and Weber 2017). \square

Proposition 7 (Translation invariance) *Log-DEA is not translation invariant.*

Proof This proposition results from the fact that the logarithmic function does not verify the distributive property, i.e., $\log(a + b) \neq \log a + \log b, \forall a, b \in \mathbb{C}$. \square

Proposition 8 (Log-translation invariance) *Log-DEA is log-translation invariant.*

Proof This is because the log-DEA model is scale invariant. Indeed, scale invariance or log-translation invariance are tantamount. \square

To introduce the next proposition, we need the dual version of log-DEA primal, which results from the duality in linear programming. We first note that β^k is an unrestricted variable of log-DEA, which means that it can be transformed into two nonnegative variables: $\beta^k = \bar{\beta}^k - \underline{\beta}^k$. Additionally, each constraint of the primal model (except for the nonnegativity of some variables) is modelled by an equation, meaning that two nonnegative variables are required to model each constraint: u^+, u^- for inputs, v^+, v^- for outputs, and w^+, w^- for the convexity constraint. Hence, we get the following set of dual constraints:

$$\log x^{j(\ell)}(u^+ - u^-) + \log \frac{1}{y^{j(\ell)}}(v^+ - v^-) + (w^+ - w^-) \geq 0, j \in \Omega^{k(\ell)}, \tag{18a}$$

$$u^+ - u^- \geq \varepsilon, \tag{18b}$$

$$v^+ - v^- \geq \varepsilon, \tag{18c}$$

$$d_1^k(u^+ - u^-) + d_2^k(v^+ - v^-) \geq 1, \tag{18d}$$

$$-d_1^k(u^+ - u^-) - d_2^k(v^+ - v^-) \geq -1, \tag{18e}$$

$$u^+, u^-, v^+, v^-, w^+, w^- \geq 0. \tag{18f}$$

The objective function is, obviously,

$$[\text{log-DEA dual}] \text{ minimize } \log x^{k(\ell)}(u^+ - u^-) + \log \frac{1}{y^{k(\ell)}}(v^+ - v^-) + (w^+ - w^-). \tag{19}$$

From constraints (18d) and (18e), we conclude that $d_1^k(u^+ - u^-) + d_2^k(v^+ - v^-) = 1$ or, equivalently, $u^+ - u^- = \frac{1-d_2^k(v^+-v^-)}{d_1^k}$, which allows us to simplify the dual model. Since $u^+ - u^- \geq \varepsilon$, we have $v^+ - v^- \leq \frac{1-\varepsilon d_1^k}{d_2^k}$. From constraint (18c), we conclude that:

$$\varepsilon \leq v^+ - v^- \leq \frac{1 - \varepsilon d_1^k}{d_2^k}.$$

Furthermore, the first constraint can be rewritten as:

$$\underbrace{\log \left((x^{j(\ell)})^{d_2^k/d_1^k} y^{j(\ell)} \right)}_{=\eta^j} (v^+ - v^-) - (w^+ - w^-) \leq \frac{1}{d_1^k} \log x^{j(\ell)}.$$

Now, we observe that $v^+ - v^- = v$ and $w^+ - w^- = w$, where v and w are both unrestricted by definition. Let

$$\eta^j = \frac{1}{d_1^k} \log \left((y^{j(\ell)})^{d_2^k} (x^{j(\ell)})^{d_1^k} \right), \text{ for } j \in \Omega. \tag{20}$$

This means that the log-DEA dual model becomes:

$$[\text{log-DEA dual}] \text{ minimize } \eta^k v - w \tag{21a}$$

subject to:

$$\eta^j v - w \leq \frac{1}{d_1^k} \log x^{j(\ell)}, j \in \Omega^{k(\ell)}, \tag{21b}$$

$$\varepsilon \leq v \leq \frac{1 - \varepsilon d_1^k}{d_2^k}, \tag{21c}$$

$$w \text{ unrestricted.} \tag{21d}$$

Proposition 9 (Multiple solutions) *If $k \in \Omega^{k(\ell)}$, then the log-DEA dual model violates the desired property of solutions uniqueness.*

Proof To prove that the log-DEA dual model provides multiple solutions it is sufficient to observe that the hyperplane created by the constraint (21b) and the one

constructed by the objective function are parallel if $k \in \Omega^{k(\ell)}$. Indeed, when $j = k$, we have $\eta^j v - w = \eta^k v - w$: the latter is, precisely, the objective function. \square

3.5 The case of multiple optima

Should the dual have multiple optima (Proposition 9), then every optimal basic solution to the primal is degenerate. This causes problems in the definition of optimal paid prices. In the absence of degeneracy, two different simplex tableaus in canonical form give two different solutions. Otherwise, it is possible that there are two different sets of basic variables (coefficients μ^k) giving the same solution (efficiency score). Since we need unique μ^k , we must develop a linear program that returns these coefficients verifying such a property.

Based on proposition 9, if one desires/needs only the efficiency score, then there is no need for further model improvements. In general, restrictions over dual variables are adopted to include managerial/policy preferences (cf. Shimshak et al. (2009), Podinovski (2016) and Podinovski and Bouzdine-Chameeva (2015, 2016) for a further discussion), but they tend to influence the efficient frontier shape. That is, efficiency scores change due to the inclusion of multiplier constraints. Intrinsically, these restrictions also minimize (but do not avoid) the multiple solutions problem (Cooper et al. 2007), at the cost of both efficiency scores changing and their meaning potential loss. To avoid this serious problem, we follow the approach of Sueyoshi and Sekitani (2009) to develop a multiplicative log-DEA model, which does not violate the desired property of solutions uniqueness. The next two definitions help us on this goal.

Definition 9 (*Complementary slackness condition*) The well-known complementary slackness theorem of linear programming can be applied to both log-DEA primal and log-DEA dual. The following conditions must be obeyed for every optima of both log-DEA models, should $u = u^+ - u^-$, $v = v^+ - v^-$, $w = w^+ - w^-$ are three free in sign variables such that u and v are related via $u = (1 - d_2^k v)/d_1^k$. Let $\eta^j = \frac{1}{d_1^k} \log \left((y^{j(\ell)})^{d_1^k} (x^{j(\ell)})^{d_2^k} \right)$ for $j \in \Omega$.

$$\mu_j^k \left(u \log x^{k(\ell)} + v \log \frac{1}{y^{k(\ell)}} + w \right) = 0 \implies \mu_j^k - \eta^j v + w > -\frac{1}{d_1^k} \log x^{j(\ell)}, j \in \Omega^{k(\ell)}, \tag{22a}$$

$$\beta^k \underbrace{(d_1^k u + d_2^k v - 1)}_{=0} = 0 \implies \beta^k \text{ is free in sign}, \tag{22b}$$

$$s_1^k (u - \varepsilon) = 0 \implies s_1^k + u > \varepsilon \implies s_1^k - \frac{d_2^k}{d_1^k} v > \varepsilon - \frac{1}{d_1^k}, \tag{22c}$$

$$s_2^k(v - \varepsilon) = 0 \implies s_2^k + v > \varepsilon. \tag{22d}$$

These are the strong complementary slackness conditions associated with the log-DEA model.

Definition 10 (*Unique Solution-based log-DEA*) The following linear program provides solutions for log-DEA model that obey to the desirable uniqueness property.

$$\begin{aligned} \text{[US-log-DEA] maximize } & \alpha \\ \text{subject to:} & \end{aligned} \tag{23a}$$

$$\langle \boldsymbol{\mu}^k, \log \mathbf{X}^{k(\ell)} \rangle + d_1^k \beta^k + s_1^k = \log x^{k(\ell)}, \tag{23b}$$

$$\langle \boldsymbol{\mu}^k, \log \frac{\mathbf{1}}{\mathbf{Y}^{k(\ell)}} \rangle + d_2^k \beta^k + s_2^k = \log \frac{1}{y^{k(\ell)}}, \tag{23c}$$

$$\langle \sqrt{\boldsymbol{\mu}^k}, \sqrt{\boldsymbol{\mu}^k} \rangle = 1, \tag{23d}$$

$$\beta^k + \varepsilon(s_1^k + s_2^k) - \eta^k v + w = 0, \tag{23e}$$

$$\eta^j v - w \leq \frac{1}{d_1^k} \log x^{j(\ell)}, j \in \Omega^{k(\ell)}, \tag{23f}$$

$$v \leq \frac{1 - \varepsilon d_1^k}{d_2^k}, \tag{23g}$$

$$- \mu_j^k + \eta^j v - w + \alpha \leq \frac{1}{d_1^k} \log x^{j(\ell)}, j \in \Omega^{k(\ell)}, \tag{23h}$$

$$- s_1^k + \frac{d_2^k}{d_1^k} v + \alpha \leq \frac{1}{d_1^k} - \varepsilon, \tag{23i}$$

$$- s_2^k - v + \alpha \leq -\varepsilon, \tag{23j}$$

$$\boldsymbol{\mu}^k, s_1^k, s_2^k, \alpha \geq 0, \tag{23k}$$

$$v \geq \varepsilon, \tag{23l}$$

$$\beta^k, w \text{ unrestricted.} \tag{23m}$$

The US-log-DEA merges the dual, the primal, and the strong complementary slackness conditions (9), and forces that the objectives of both dual and primal log-DEA models are equal (23e). If a variable x verifies $x > a$, where $a \in \mathbb{R}$, then there

is a nonnegative quantity α such that $x - \alpha \geq a$. The objective of US-log-DEA is to maximize the parameter α that transforms the strong complementary slackness conditions into feasible linear conditions (23h–23j). This means that the optimal solution of US-log-DEA, $(\mu^{k*}, \beta^{k*}, s_1^{k*}, s_2^{k*}, \nu^*, w^*, \alpha^*)$, verifies the strong complementary slackness conditions (9). US-log-DEA solves the problem of multiple projections in log-DEA dual as it restricts multipliers so that the projection becomes unique (Sueyoshi and Sekitani 2009). That is, restricting multipliers ν and w does not solve the problem of multiple solutions. Furthermore, it results from the linear programming theory and from the duality theorem that optima of US-log-DEA are also optima of both log-DEA dual and primal. Therefore, the efficiency score $\theta^{k(\ell)}$ remains unchanged. But since we need coefficients μ^k to define optimal prices, we must save the outcomes of US-log-DEA for such a purpose. It is needless to say that, by proposition 9, if k does not belong to $\Omega^{k(\ell)}$, then it is not necessary the use of US-log-DEA as the problem of multiple solutions vanishes in that case.

3.6 The problem of infeasibility in linear programming models

We have developed a log-DEA model to assess the coefficients related to the optimal price to be paid to hospital k . In some circumstances, the log-DEA primal can be infeasible because of the choice of the directional vector \mathbf{D}^k . If $k \in \Omega^{k(\ell)}$, then the linear model cannot be infeasible. This is because k is enveloped by the frontier constructed using data associated with that subset of Ω . However, k may not belong to $\Omega^{k(\ell)}$ because of its quality levels. Two scenarios are then suitable: k is enveloped by the frontier related to $\Omega^{k(\ell)}$ —and, in such a case, there is not an infeasibility problem—or k is super-efficient regarding such a frontier. Only in the latter scenario infeasibility may occur. However, being super-efficient is not a sufficient condition to generate infeasibility, vide Proposition 10. We thus have to look at the log-DEA primal and search for a vector \mathbf{D}^k linear transformation to ensure that the model becomes feasible. Naturally, if k belongs to $\Omega^{k(\ell)}$, then a multiple solutions problem arise and the US-log-DEA is compulsory. Nevertheless, in such a case, the linear problem cannot be infeasible. If $k \notin \Omega^{k(\ell)}$, then the multiple solutions problem vanishes. Hence, it is enough to consider the log-DEA primal.

Proposition 10 (Infeasibility of log-DEA primal) *Let $\mathbf{D}^k = (\log x^{k(\ell)}, \log y^{k(\ell)})$ (Chambers et al. 1996, 1998). If $y^{k(\ell)} > 1$, $y^{j(\ell)} > 1$ for all $j \in \Omega^{k(\ell)}$, and $\langle \mu^k, \log \mathbf{X}^{k(\ell)} \rangle > 2 \log x^{k(\ell)}$, then log-DEA primal is infeasible.*

Proof This proof follows Ray (2008). Under the assumption of $\mathbf{D}^k = (\log x^{k(\ell)}, \log y^{k(\ell)})$ (Chambers et al. 1996, 1998), we obtain $\langle \mu^k, \log \mathbf{X}^{k(\ell)} \rangle \leq (1 - \beta^k) \log x^{k(\ell)}$. If $\langle \mu^k, \log \mathbf{X}^{k(\ell)} \rangle > 2 \log x^{k(\ell)}$, then $(1 - \beta^k) \log x^{k(\ell)} > 2 \log x^{k(\ell)} \Leftrightarrow 1 - \beta^k > 2 \Leftrightarrow \beta^k < -1$. Likewise, we have $\langle \mu^k, \log \mathbf{Y}^{k(\ell)} \rangle \geq (1 + \beta^k) \log y^{k(\ell)}$. Now, we know that $\beta^k < -1$ or $1 + \beta^k < 0$. If $y^{j(\ell)} > 1$ for all $j \in \Omega^{k(\ell)}$, then $\langle \mu^k, \log \mathbf{Y}^{k(\ell)} \rangle$ must be positive. However, given that $y^{k(\ell)}$ is also larger than 1, the right hand side of the inequality associated with $\mathbf{Y}^{k(\ell)}$ becomes

negative because $1 + \beta^k < 0$. Therein lies the problem: there is no combination of (nonnegative coefficients) μ_j^k for $j \in \Omega^{k(\ell)}$ obeying to $\langle \sqrt{\mu^k}, \sqrt{\mu^k} \rangle = 1$ and simultaneously satisfying $\mathbf{Y}^{k(\ell)} < 0$. Hence, the log-DEA primal model becomes infeasible. \square

Proposition 11 (Affine transformation of \mathbf{D}^k to remove infeasibility) *Let $\mathcal{D}^{k(1)} = \delta_0^{k(1)} + \delta_1^{k(1)} d_1^k$ and $\mathcal{D}^{k(2)} = \delta_0^{k(2)} + \delta_1^{k(2)} d_2^k$ be the affine transformation of the components of vector $\mathbf{D}^k = (d_1^k, d_2^k)$, with $\delta_0^{k(1)}, \delta_1^{k(1)}, \delta_0^{k(2)}, \delta_1^{k(2)} \geq 0$. Let $m_Y = \min \mathbf{Y}^{k(\ell)}$ and $\nabla^k = \max \mathbf{X}^{k(\ell)} - \min \mathbf{X}^{k(\ell)} + \epsilon$, with $\epsilon > 0$. If the following condition is met by the coefficients of the affine transformation of \mathbf{D}^k , then the infeasibility problem vanishes.*

$$m_Y \mathcal{D}^{k(1)} \geq \nabla^k \mathcal{D}^{k(2)} \iff m_Y (\delta_0^{k(1)} + \delta_1^{k(1)} d_1^k) \geq \nabla^k (\delta_0^{k(2)} + \delta_1^{k(2)} d_2^k) \tag{24}$$

We can select any values for the coefficients of the affine transformation, as long as they obey to the relationship in (24).

Proof (Inspired on Chen et al. 2013) The log-DEA primal is feasible if and only if all of its constraints are obeyed. With the affine transformation of \mathbf{D}^k , those constraints should be verified. Let us ignore slacks s_1^k and s_2^k and start by the constraint over the input:

$$\langle \mu^k, \log \mathbf{X}^{k(\ell)} \rangle + \mathcal{D}^{k(1)} \beta^k \leq \log x^{k(\ell)} \iff \beta^k \leq \frac{\log x^{k(\ell)} - \langle \mu^k, \log \mathbf{X}^{k(\ell)} \rangle}{\delta_0^{k(1)} + \delta_1^{k(1)} d_1^k}.$$

From the constraint over the output, we have:

$$\langle \mu^k, \log \frac{1}{\mathbf{Y}^{k(\ell)}} \rangle + \mathcal{D}^{k(2)} \beta^k \leq \log \frac{1}{y^{k(\ell)}} \iff \langle \mu^k, \log \mathbf{Y}^{k(\ell)} \rangle \geq \mathcal{D}^{k(2)} \beta^k + \log y^{k(\ell)}.$$

The output target is, of course, $y^{k(\ell)} = \langle \mu^k, \log \mathbf{Y}^{k(\ell)} \rangle$, and this should be nonnegative, i.e.

$$\mathcal{D}^{k(2)} \beta^k + \log y^{k(\ell)} \geq 0 \iff \beta^k \geq \frac{\log \frac{1}{y^{k(\ell)}}}{\delta_0^{k(2)} + \delta_1^{k(2)} d_2^k}.$$

Hence, the following is obvious:

$$\frac{\log \frac{1}{y^{k(\ell)}}}{\delta_0^{k(2)} + \delta_1^{k(2)} d_2^k} \leq \beta^k \leq \frac{\log x^{k(\ell)} - \langle \mu^k, \log \mathbf{X}^{k(\ell)} \rangle}{\delta_0^{k(1)} + \delta_1^{k(1)} d_1^k} \Rightarrow \frac{\log y^{k(\ell)}}{\delta_0^{k(2)} + \delta_1^{k(2)} d_2^k} \geq \frac{\langle \mu^k, \log \mathbf{X}^{k(\ell)} \rangle - \log x^{k(\ell)}}{\delta_0^{k(1)} + \delta_1^{k(1)} d_1^k}.$$

Unfortunately, we do not know the value of $\langle \mu^k, \log \mathbf{X}^{k(\ell)} \rangle$, but it is clear that $\langle \mu^k, \log \mathbf{X}^{k(\ell)} \rangle - \log x^{k(\ell)} \leq \nabla^k$. Therefore, it is sufficient to impose:

$$m_Y \mathcal{D}^{k(1)} \geq \nabla^k \mathcal{D}^{k(2)} \iff m_Y (\delta_0^{k(1)} + \delta_1^{k(1)} d_1^k) \geq \nabla^k (\delta_0^{k(2)} + \delta_1^{k(2)} d_2^k),$$

which concludes this proof. \square

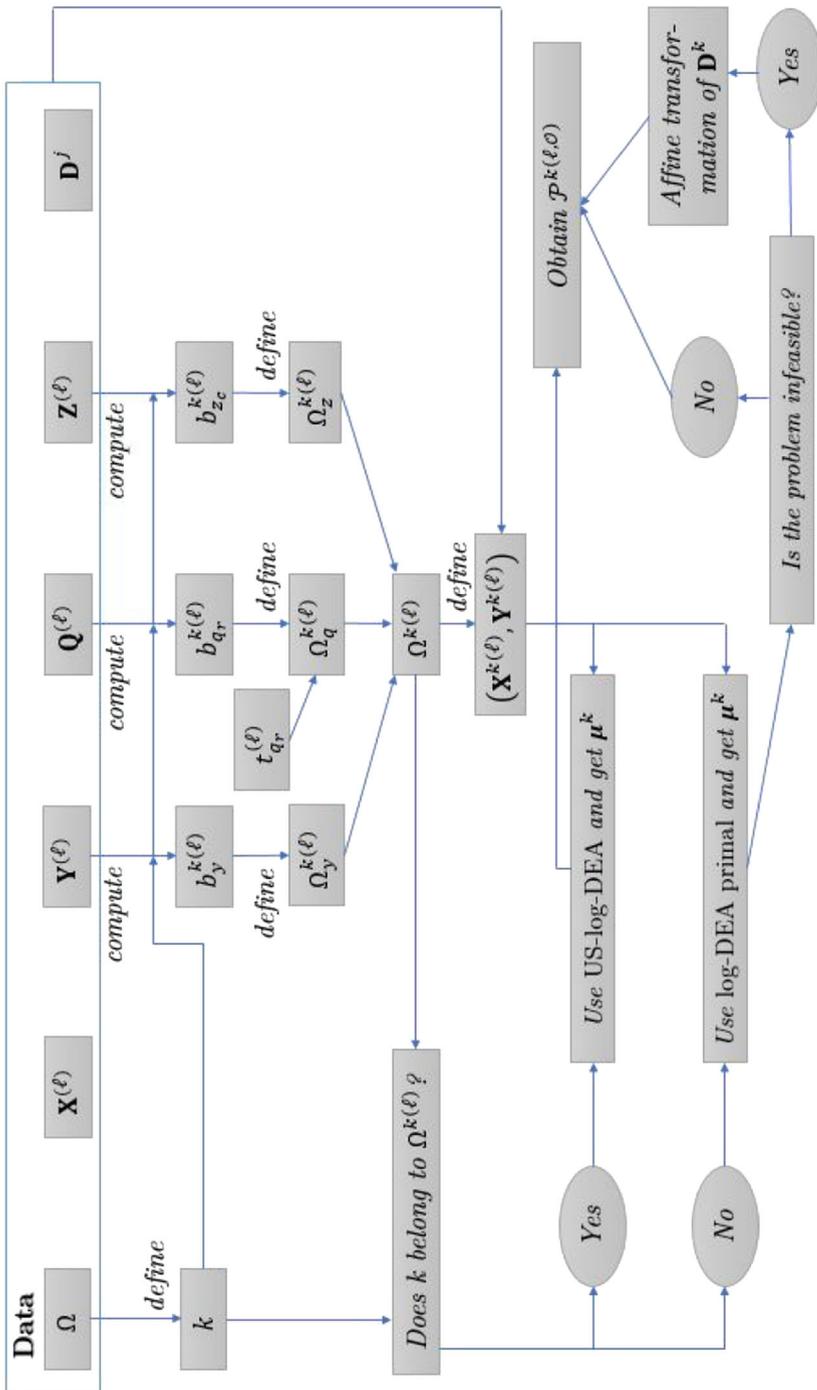


Fig. 3 Flowchart

3.7 On optimizing the best practice tariffs: a step-by-step procedure

Figure 3 presents the flowchart of the procedure adopted to optimize the paid price to hospitals by their services. Steps are as follows³:

Step 1 Define data.

Step 2 Define $k \in \Omega$ as the hospital whose payment we want to optimize.

Step 3 Define bandwidths, as suggested in Sect. 2.4.1.

Step 4 Construct the comparability sets, according to Definitions 1–3.

Step 5 Construct the overall comparability set, $\Omega^{k(\ell)}$, by intersecting the sets achieved in Step 4—vide Definition 4.

Step 6 Using the entries of $\Omega^{k(\ell)}$, construct $\mathbf{X}^{k(\ell)}$ and $\mathbf{Y}^{k(\ell)}$ —details at the end of Subsection 2.4.2.

Step 7 Check if k belongs to $\Omega^{k(\ell)}$.

Step 7.1 If $k \in \Omega^{k(\ell)}$, use the US-log-DEA model (23) and obtain $\boldsymbol{\mu}^k$ —its coefficients are the unique solutions associated with the benchmarks of k .

Step 7.2 If $k \notin \Omega^{k(\ell)}$, use the log-DEA primal model (14). Check if the log-DEA primal is infeasible.

Step 7.2.1 If the model is infeasible, transform \mathbf{D}^k using an affine transformation obeying to Proposition 11. The infeasibility problem vanishes. Obtain $\boldsymbol{\mu}^k$ whose components are unique for k .

Step 7.2.2 If the model is not infeasible, just obtain $\boldsymbol{\mu}^k$.

Step 8 Using $\boldsymbol{\mu}^k, \mathbf{X}^{k(\ell)}, y^{k(\ell)}$, and Eq. 8, obtain the optimal payment $\mathcal{P}^{k(\ell,0)}$ associated with k .

4 Efficiency, optimal payments, and cost savings

The definition of optimal tariffs to be paid foresees some cost savings to the commissioner. This section explains how optimal payments are related to the technical efficiency, as previously detailed, and how cost savings can be generated.

4.1 Efficiency and optimal payments

We can easily relate the efficiency score with the optimal payment due to k .

Proposition 12 *The optimal payment due to k ($\mathcal{O} = 0$) can be written as follows:*

$$\mathcal{P}^{k(\ell,0)} = \frac{1}{y^{k(\ell)}} \theta^{k(\ell)} \mathcal{H}(\mathbf{Y}^{k(\ell)}, \boldsymbol{\mu}^k, 0) \hat{\mathcal{P}}^{k(\ell)}, \quad (25)$$

where $\hat{\mathcal{P}}^{k(\ell)} = x^{k(\ell)} / y^{k(\ell)}$ is the current unitary cost of k in service/DRG $\ell \in \Psi$.

Proof This proposition is obvious after merging Eqs. (8) and (16). \square

³ The model will be made available upon request.

According to Eq. (25), the optimal payment $\mathcal{P}^{k(\ell,0)}$ depends on coefficients $\boldsymbol{\mu}^k$ of the log-linear combination; hence, it can be non-unique unless the US-log-DEA model is used. However, $\theta^{k(\ell)}$ is unique and can be achieved simply by using the log-DEA primal and the relationship between the efficiency score and β^k .

Proposition 13 *If hospital k is efficient regarding $\Omega^{k(\ell)}$ in the Pareto sense (absence of slacks), then $\mathcal{P}^{k(\ell,0)} = \hat{\mathcal{P}}^{k(\ell)}$.*

Proof If hospital k is efficient regarding $\Omega^{k(\ell)}$ in the Pareto sense (absence of slacks), then $\theta^{k(\ell)} = 1$ and $\boldsymbol{\mu}^k$ contains one and only one entry that is different from zero, i.e., $\mathcal{H}(\mathbf{Y}^{k(\ell)}, \boldsymbol{\mu}^k, 0) = y^{k(\ell)}$, which implies $\mathcal{P}^{k(\ell,0)} = \hat{\mathcal{P}}^{k(\ell)}$ from Eq. (25). \square

Proposition 14 $\mathcal{P}^{k(\ell,0)} < \hat{\mathcal{P}}^{k(\ell)} \iff \theta^{k(\ell)} < y^{k(\ell)} / \mathcal{H}(\mathbf{Y}^{k(\ell)}, \boldsymbol{\mu}^k, 0)$.

Proof The proposition results directly from (25). Let hospital k be inefficient. Thus, there may exist an inefficiency source in the side of outputs, say s_2^k . Should it be zero, we have $\mathcal{H}(\mathbf{Y}^{k(\ell)}, \boldsymbol{\mu}^k, 0) = y^{k(\ell)}$. Yet, because of inputs, k is inefficient by hypothesis, and it is clear that $\theta^{k(\ell)} < 1$ and $\mathcal{P}^{k(\ell,0)} < \hat{\mathcal{P}}^{k(\ell)}$. That is, the optimal payment is smaller than the current unitary cost incurred by k . Should s_2^k be larger than zero, we have $\mathcal{H}(\mathbf{Y}^{k(\ell)}, \boldsymbol{\mu}^k, 0) > y^{k(\ell)}$. Therefore, we get $\mathcal{P}^{k(\ell,0)} < \hat{\mathcal{P}}^{k(\ell)}$ if and only if $\theta^{k(\ell)} < y^{k(\ell)} / \mathcal{H}(\mathbf{Y}^{k(\ell)}, \boldsymbol{\mu}^k, 0)$. \square

4.2 Cost savings

Based on both current and optimal costs, we can easily define the cost savings for hospital $k \in \Omega$ and service/DRG $\ell \in \Psi$.

Definition 11 (*Cost savings for hospital k and service/DRG ℓ , $\Delta^{k(\ell)}$)* Cost savings are defined as the (positive) difference between the current and the optimal operational expenditures of hospital k for the service/DRG ℓ :

$$\Delta^{k(\ell,\mathcal{O})} = x^{k(\ell)} - \mathcal{H}(\mathbf{X}^{k(\ell)}, \boldsymbol{\mu}^k, \mathcal{O}) = x^{k(\ell)} - \left\langle \boldsymbol{\mu}^k, \left(\mathbf{X}^{k(\ell)}\right)^{\mathcal{O}} \right\rangle^{1/\mathcal{O}} \underbrace{=}_{(8)} x^{k(\ell)} - y^{k(\ell)} \mathcal{P}^{k(\ell,\mathcal{O})}. \tag{26}$$

As claimed in Definition 11, there are cost savings if and only if $\Delta^{k(\ell)}$ is strictly positive, i.e., $x^{k(\ell)} > \mathcal{H}(\mathbf{X}^{k(\ell)}, \boldsymbol{\mu}^k, \mathcal{O})$.

Proposition 15 *Cost savings associated with k ($\mathcal{O} = 0$) can be written as follows:*

$$\Delta^{k(\ell,\mathcal{O})} = x^{k(\ell)} \left(1 - \frac{1}{y^{k(\ell)}} \theta^{k(\ell)} \mathcal{H}(\mathbf{Y}^{k(\ell)}, \boldsymbol{\mu}^k, 0) \right). \tag{27}$$

Proof It is straightforward after inserting (25) into (26). □

Proposition 16 *If k is technically efficient regarding $\Omega^{k(\ell)}$ and in the Pareto sense, then $\Delta^{k(\ell, \mathcal{O})} = 0$.*

Proof If k is efficient in the Pareto sense, then $\mathcal{H}(\mathbf{Y}^{k(\ell)}, \boldsymbol{\mu}^k, \mathcal{O}) = y^{k(\ell)} \wedge \mathcal{H}(\mathbf{X}^{k(\ell)}, \boldsymbol{\mu}^k, \mathcal{O}) = x^{k(\ell)} \iff \theta^{k(\ell)} = 1 \implies \Delta^{k(\ell, \mathcal{O})} = 0$. The reciprocal is not necessarily true. □

Proposition 17 $\Delta^{k(\ell, 0)} = 0 \iff \mathcal{P}^{k(\ell, 0)} = \hat{\mathcal{P}}^{k(\ell)}$.

Proof This proposition results from both propositions 13 and 16. □

Proposition 18 $\Delta^{k(\ell, 0)} > 0 \iff \mathcal{P}^{k(\ell, 0)} < \hat{\mathcal{P}}^{k(\ell)}$.

Proof □

Proposition 19 *Let $t_{q_r}^{\uparrow(\ell)}$ and $t_{q_r}^{\downarrow(\ell)}$ be two thresholds related to the $r^{(\ell)}$ th quality dimension. Cost savings associated with them are $\hat{\Delta}^{k(\ell, 0)}$ and $\tilde{\Delta}^{k(\ell, 0)}$, respectively. If $t_{q_r}^{\downarrow(\ell)} < t_{q_r}^{\uparrow(\ell)}$ (ceteris paribus), then $\hat{\Delta}^{k(\ell, 0)} \geq \tilde{\Delta}^{k(\ell, 0)}$.*

Proof Thresholds are defined as the minimum acceptable quality levels. Let us decrease a threshold related to a given quality dimension, keeping all remaining parameters unchanged, i.e. from $t_{q_r}^{\uparrow(\ell)}$ to $t_{q_r}^{\downarrow(\ell)}$. Let $\#A$ be the cardinality of a set A . It is clear that $\#\hat{\Omega}_q^{k(\ell)} \geq \#\tilde{\Omega}_q^{k(\ell)}$, $\#\hat{\Omega}_y^{k(\ell)} \geq \#\tilde{\Omega}_y^{k(\ell)}$, and $\#\hat{\Omega}_z^{k(\ell)} \geq \#\tilde{\Omega}_z^{k(\ell)}$. By construction, $\#\hat{\Omega}^{k(\ell)} \geq \#\tilde{\Omega}^{k(\ell)}$. Note that we have to impose \geq instead of $>$ because a decrease on thresholds does not necessarily increase the cardinality of a set. The probability of k become dominated by the new comparability set increases as a consequence of the broadening of $\Omega^{k(\ell)}$.⁴ Therefore, the efficiency of hospital k does not increase, and, by Proposition 15, cost savings do not decrease. □

Proposition 20 *Let $b_{q_r}^{\uparrow k(\ell)}$ and $b_{q_r}^{\downarrow k(\ell)}$ be two bandwidths associated with the $r^{(\ell)}$ th quality dimension. Cost savings related to them are $\hat{\Delta}^{k(\ell, 0)}$ and $\tilde{\Delta}^{k(\ell, 0)}$, respectively. If $b_{q_r}^{\downarrow k(\ell)} > b_{q_r}^{\uparrow k(\ell)}$ (ceteris paribus), then $\hat{\Delta}^{k(\ell, 0)} \geq \tilde{\Delta}^{k(\ell, 0)}$.*

Proof As in previous proposition. □

Based on previous propositions, and given the user-defined thresholds $t^{(\ell)} = \{t_1^{(\ell)}, \dots, t_{q_r}^{(\ell)}, \dots, t_{q_R}^{(\ell)}\}$, we can identify several scenarios:

- Hospital k belongs to $\Omega^{k(\ell)}$ and exhibits the smallest unitary cost; hence, $\mathcal{P}^{k(\ell, 0)} = x^{k(\ell)} / y^{k(\ell)}$ and $\Delta^{k(\ell, 0)} = 0$.

⁴ This is the so-called curse of dimensionality, although it certainly cannot be understood as a curse in the present case.

- Hospitals k and j both belong to $\Omega^{k(\ell)}$ but $x^{k(\ell)}/y^{k(\ell)} > x^{j(\ell)}/y^{j(\ell)}$; hence, $\mathcal{P}^{k(\ell,0)} < x^{k(\ell)}/y^{k(\ell)}$ and $\Delta^{k(\ell,0)} > 0$.
- Hospital k does not belong to $\Omega^{k(\ell)}$ because of its poor quality levels ($q_{r^{(\ell)}}^{k(\ell)} < t_{q_r^{(\ell)}}^{(\ell)}$):
 - There is, at least, a hospital j in $\Omega^{k(\ell)}$ such that $x^{k(\ell)}/y^{k(\ell)} > x^{j(\ell)}/y^{j(\ell)}$; hence, $\Delta^{k(\ell,0)} > 0$.
 - All hospitals in $\Omega^{k(\ell)}$ have unitary costs larger than $x^{k(\ell)}/y^{k(\ell)}$; hence, $\Delta^{k(\ell,0)} < 0$.

In the last case, the user (or group of stakeholders) should decide whether quality threshold(s) is(are) too large and must be decreased to raise cost savings, or threshold(s) must be kept and low-quality hospitals must be financed by $\mathcal{P}^{k(\ell,0)} > x^{k(\ell)}/y^{k(\ell)}$ and the difference $|\Delta^{k(\ell,0)}|$ must be solely devoted to quality improvements.

5 The application of the new tool to the Portuguese public hospitals

5.1 An overview about the Portuguese health care system

Contracting in health care aims to establish a mechanism for resources allocation in line with each health care service provision and the corresponding population needs. These ones include to ensure the quality, efficiency, and effectiveness of health care (Shimshak et al. 2009). The contracting process in Portuguese health care services arises from the relationship between different stakeholders, including (1) public funding sources (the Portuguese State, through taxes), (2) the regulator (an independent Regulatory Agency for Health and the Ministry of Health), and (3) the health care providers (hospitals, either public or private, primary health care centres, and tertiary health care centres) (Sakellarides 2010). Figure 4 shows the financial flow in the Portuguese (NHS). The analysis of this section is centred on public hospitals. Most of their financing volume comes from taxes, collected from citizens and companies. Health care providers are public though autonomous entities, implying that there must be a monitoring model holding them responsible for their weak performance and inducing the so-called transparency for the population those hospitals serve.

Contracting mechanisms try to allocate resources to health care providers in an efficient and effective way, through a clear and well-designed strategy. However, in most European countries there is nothing like a set of indicators allowing the evaluation of that strategy impact on the population health status which is a quite difficult task (Figueras et al. 2005; Lu et al. 2003). Accordingly, such a strategy requires a previous fieldwork (negotiation phase) that should be fulfilled by a contracting-specific team from the Ministry of Health. This strategy must be in line with some directions defined by the Portuguese Central Administration of Health

Portugal in the past few years, leading to capital injections with money from the 2011 Memorandum of Understanding between the Portuguese Government and the European Commission (on behalf of the Eurogroup), the European Central Bank and the International Monetary Fund (Nunes and Ferreira 2019a).

In Portugal, contracts between the (central) Government and each health care provider encompass a set of dimensions including the provisions type, volume and duration, referencing networks, human resources and facilities, monitoring schemes, performance prizes, and prices. Worth to mention, contracts should also contain quality-related terms, including penalties for poor quality. That is, contracting health care services is paramount to promote a user's focused action and to improve quality and efficiency, not only in Portugal but also in any other country.

Prices are perhaps one of the most important variables within a contract for health care provision. As a matter of fact, statutory payments to hospitals regard their production in terms of DRG and services (emergencies, medical appointments, inpatient services, ...). These prices are usually defined according to the smallest unitary costs (in the previous year) among providers. Some adjustments regarding the average complexity of treated patients can also be included. Nonetheless, quality is clearly neglected in such a framework. We remark that health care services, like any other service (either public or not), may exhibit considerable levels of inefficiency, which means those unitary costs necessarily include an inefficiency parcel. Furthermore, providers also exhibit heterogeneous technologies (e.g., assets). In a period featured by resources scarcity, financing hospitals through an inefficiency-based mechanism that disregards their underlying technologies is totally nonsense, requiring a considerable reformulation, justifying, in theory, the adoption of our approach.

5.2 Contracting in the Portuguese NHS

Several New Public Management-based health policies have been extensively applied to the Portuguese public hospitals. Particularly, their management model (legal status) has observed a few changes since the beginning of 2003, moving from the Administrative Public Sector (under the public/administrative law) to the State Business Sector (under the commercial/private law) in 2005 (Ferreira and Marques 2015). This was the so-called *corporatisation reform*. Currently, most of Portuguese hospitals belong to the State Business Sector. One of the most relevant issues underlying this reform regards the hospitals financing model. While in the past such a financing was based on a retrospective model and prices (or tariffs) derived from each hospital history, nowadays this financing stems from the negotiation phase and the resulting contract. There is a hospital-specific budget that is negotiated between each hospital's Administrative Council and the Ministry of Health. Budget is negotiated in terms of production and quality, in line with some other European countries like Finland, Ireland, France, and England. In Portugal, paid unitary price (tariff) results from the national average of DRGs or services' costs. Patients are assigned to a specific DRG and/or to a specific service. In any case, tariffs are tabled, which means they are the same for all hospitals, regardless their performance. In other words, the current mechanism is basically a case-based

payment, which encompasses several disadvantages: first, it may prompt the *supplier-induced demand* that, in turn, includes unwanted activity; second, more complex cases can be disregarded or treated negligently, which reduces the overall quality of care; third, providers are encouraged to *up-code* classification of patients into a more highly reimbursed group; fourth, the introduction of quality-raising (though cost-increasing) technologies is totally discouraged (Marshall et al. 2014).

Valente (2010) identifies several contracting objectives: (1) to control health care provision-related expenses that could jeopardise the system sustainability in the medium to long term; (2) to promote providers' efficiency, by simultaneously increasing their production and quality, and decreasing their resources consumption; (3) to ensure high quality levels of care; and (4) to promote the hospital managers' accountability. As a matter of fact, besides production and quality that must be ensured, there are several contract-specific objectives that must be fulfilled under pain of penalties.

The contracting process in the Portuguese NHS has seen considerable changes over the past 13 years, as it must fit to reality, available resources, needs, and technological developments. These dynamics and the usual process redefinition and re-adaptation promote an adjustment mechanism of the hospitals' performance (either financial or not) program, encouraging the objectives for which the contract was created. Currently, the contracting process between the Portuguese public hospitals and the Ministry of Health integrates a triennial strategic planning process merging the existing forecasting documents, i.e., the Business Plan, the Performance Plan, and the Adjustment Plan and Financial Statements.

5.3 Data, sample, hypotheses and methodological issues

To exemplify the new tool, we use monthly data of twenty nine hospitals and their activity in 2016 and 2017, which results on 696 observations. Following Ferreira and Marques (2017), the sample does not include two out of the four public–private partnership hospitals (because of data gaps), three highly differentiated hospitals as oncology centres (because of their production technology that significantly differs from the one of general hospitals), and local health units (since they result from the vertical merging of a general hospital and several primary health care units). Data are available in the official source website (<http://benchmarking.acss.min-saude.pt/>). Available (operational) expenditures refer to the whole activity of the hospital, hence we can assess the optimal payment by patient if our measure of output is the number of standard patients.

Selected variables are exhibited in Fig. 5. Data are available on-line.⁵ Quality indicators refer to patients' clinical safety, care appropriateness and timeliness, and access to health care. In this case, we have followed the variables adopted by Ferreira and Marques (2018b), taking into account the data availability for all the 696 observations. There is a maximum legal time for a secondary health care response to the citizens searching for (nonurgent) medical appointments and for surgeries. Overstepping such a time indicates poor access to the health care services.

⁵ <https://drive.google.com/drive/folders/1I0qxCQcKp2SZRBkoF4zu1WxGRvjZkXF?usp=sharing>.

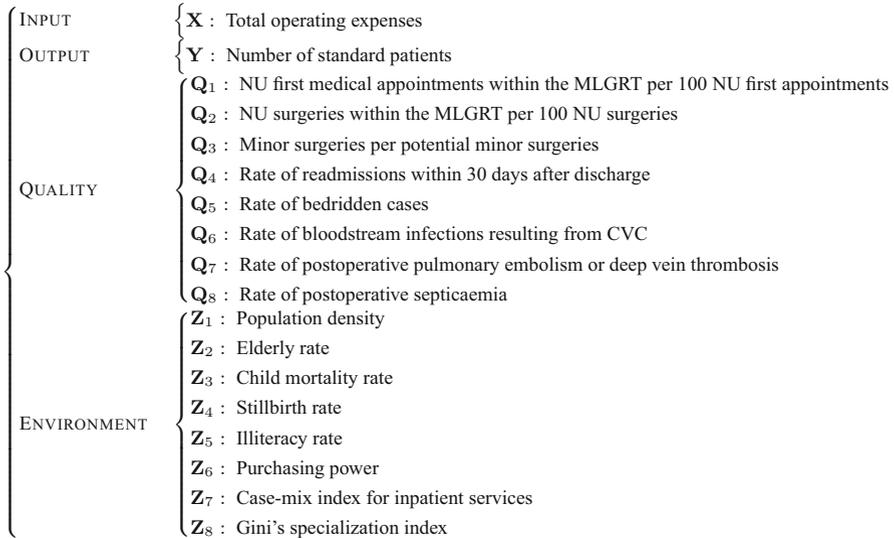


Fig. 5 Input, output, quality, and environment. *NU* nonurgent, *MLGRT* maximum legal guaranteed response time, *CVC* central venous catheter

This is a phenomenon revealed by small values of both **Q**₁ and **Q**₂. Minor surgeries should be made in ambulatory services, otherwise the patient can be subject to more complex procedures and to hospital acquired severe infections. The costs of surgeries undertaken in the operating theatre are, of course, larger than the others. Minor surgeries accomplished in ambulatory services reveal good care appropriateness and efficiency levels. Hence, the larger **Q**₃, the better. In opposition, high values of **Q**₄ up to **Q**₈ reveal poor levels of care appropriateness and clinical safety.

As we have noted before, our methodology assumes that the utility function related to each quality level should be monotonically increasing, i.e., $\forall q_{r^{(\ell)}}^{k^{(\ell)}} \geq q_{r^{(\ell)}}^{j^{(\ell)}} \implies U_{r^{(\ell)}}(k) \geq U_{r^{(\ell)}}(j), k, j \in \Omega$. This means that quality dimensions **Q**₄ up to **Q**₈ must be transformed. Additionally, the definition of thresholds can be facilitated if all quality dimensions are rescaled to the interval [0,100]%. For instance, rates of bloodstream infections resulting from central venous catheter are expressed in cases per 100,000 surgeries. Rescaling quality dimensions also reduce the odds of having empty comparability sets and, hence, reducing the number of interactions with the decision maker(s). We use the following rescaling procedure:

$$\tilde{q}_{r^{(\ell)}}^{k^{(\ell)}} = \begin{cases} 100 \frac{q_{r^{(\ell)}}^{k^{(\ell)}} - \min \mathbf{Q}_{r^{(\ell)}}}{\max \mathbf{Q}_{r^{(\ell)}} - \min \mathbf{Q}_{r^{(\ell)}}}, & \text{if } r^{(\ell)} = 1, 2, 3, \\ 100 \frac{\max \mathbf{Q}_{r^{(\ell)}} - q_{r^{(\ell)}}^{k^{(\ell)}}}{\max \mathbf{Q}_{r^{(\ell)}} - \min \mathbf{Q}_{r^{(\ell)}}}, & \text{otherwise.} \end{cases} \tag{28}$$

Therefore, the larger $\tilde{q}_{r^{(\ell)}}^{k^{(\ell)}}$, the better the utility and the performance of the hospital.

Table 1 Statistics associated with the rescaled quality variables

	Q ₁	Q ₂	Q ₃	Q ₄	Q ₅	Q ₆	Q ₇	Q ₈
Percentile 25%	41.33	42.19	76.20	30.51	92.38	100.00	98.72	75.47
Median	52.61	57.73	80.89	38.84	100.00	100.00	100.00	89.78
Average (\bar{x})	54.52	59.25	80.10	39.34	94.87	96.29	99.16	85.87
Percentile 75%	65.89	78.62	85.35	46.33	100.00	100.00	100.00	100.00
Standard deviation ($\hat{\sigma}$)	20.37	23.76	9.23	15.12	9.73	11.87	3.93	17.17
Global bandwidth (Triweight kernel)	19.96	23.28	9.05	14.82	9.53	11.63	3.85	16.82
Coefficient of variation ($\hat{\sigma}/\bar{x}$)	37.36	40.10	11.53	38.44	10.25	12.33	3.97	19.99
Kurtosis	2.82	2.16	20.98	5.26	19.27	22.59	582.92	4.96
Skewness	0.28	0.01	- 2.62	0.79	- 3.09	- 4.12	- 23.13	- 1.36

The hospital $j \in \Omega$ with $\hat{q}_{r^{(\ell)}}^{j(\ell)} = 100$ is the best performer in the quality dimension $r^{(\ell)} \in \Gamma^{(\ell)}$. In opposition, $\hat{q}_{r^{(\ell)}}^{j'(\ell)} = 0$ identifies the hospital $j' \in \Omega$ with the worst performance in the very same quality dimension. Table 1 presents the basic statistics associated with the rescaled quality variables. Hereinafter, we will consider the global bandwidths computed by using the triweight kernel, i.e., $b_{q_r}^{j(\ell)} = b_{q_r}^{j+1(\ell)} = 3.62\hat{\sigma}_{q_r}J^{-1/5} = 0.9799\hat{\sigma}_{q_r}, \forall j = 1, \dots, J - 1$, with $J = 696$. We will also use the average of the distributions to represent the user predefined thresholds, $t_{q_r}^{(\ell)} = \frac{1}{J} \sum_{j=1}^J \hat{q}_{r^{(\ell)}}^{j(\ell)}$. We observe that averages and medians are similar; thus, we have roughly half of the sample (348 observations) above each threshold. Of course, there is a correspondence between the rescaled and the original threshold. Indeed, take the case of $r^{(\ell)} = 1$; if $t_{q_r}^{(\ell)} = 54.52$, then $t_{q_r}^{(\ell)} = t_{q_r}^{(\ell)} \frac{\max Q_1 - \min Q_1}{100} + \min Q_1 = 54.52 \frac{99.99 - 34.51}{100} + 34.51 = 70.21$, which is (hypothetically) the minimum acceptable number of nonurgent first medical appointments within the maximum legal guaranteed response time per 100 first appointments.

Environment is also paramount to derive optimal payments to hospitals. We have selected six dimensions that intend to capture the population’s demography and two more to explain the complexity and specialization degree of each hospital. The case-mix index relates the costs associated with each DRG and the quantity of patients of the very same group. The larger the number of patients of higher complexity, the larger the case-mix index. This one is usually rescaled such that the national average is set as 1; hence, hospitals with a case-mix larger than 1 usually handle with more complex cases. Meanwhile, the Gini’s specialization degree, as proposed by Daidone and D’Amico (2009) and Lindlbauer and Schreyögg (2014), is related to the hospital efficiency and ranges between 0 (low specialization) and 1 (high specialization). Table 2 provides the basic statistics related to environmental variables. As before, we have adopted the triweight kernel to compute global bandwidths. In the case of environment dimensions, no data transformations are required.

Table 2 Statistics associated with the environment variables

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
Percentile 25%	88.40	16.92	2.40	30.77	4.00	85.50	0.74	62.63
Median	233.00	20.59	2.70	33.33	5.00	92.00	0.83	64.91
Average	489.04	20.46	3.06	35.59	5.40	94.75	0.87	64.27
Percentile 75%	649.45	23.14	3.20	35.71	6.30	99.30	0.97	65.86
Standard deviation	567.53	3.81	1.18	10.79	2.18	12.86	0.19	2.32
Global bandwidth (Triweight kernel)	556.13	3.74	1.16	10.58	2.13	12.60	0.18	2.27
Coefficient of variation	116.05	18.64	38.71	30.33	40.32	13.57	21.35	3.60
Kurtosis	5.98	2.68	6.49	6.86	3.90	3.81	3.59	3.71
Skewness	1.79	0.62	1.77	1.80	1.22	1.03	1.15	- 0.83

Optimization models were developed and solved through the integration of the package IBM ILOG CPLEX[®] Optimization Studio, v12.6.3, and the MATLAB[®] software. CPLEX[®] is useful for linear programming, whereas MATLAB[®] easily handles with matrixes manipulation, for-/while- cycles, and if-else environments, as the ones exhibited in the flowchart of Fig. 3. The algorithm can be delivered upon request.

5.4 Results and discussion

Once the overall set has been defined for each hospital, based on its corresponding size-, quality-, and environment-based comparability sets, we can use the US-log-DEA to optimize operational expenses and, by construction, paid prices.

Let us consider the case of observation $k = 100$, for which we have

$$\Omega_z^{k(\ell)} = \{97, 98, \dots, 100, \dots, 119, 120, 385, 386, \dots, 407, 408\}$$

and

$$\Omega_q^{k(\ell)} = \{37, 126, 127, 128, 139, 143, 241, 575, 685, 689, 691\}.$$

Since $k \notin \Omega_q^{k(\ell)}$, we state that its quality levels are below the minimum required quality defined by the stakeholder(s). It is clear that $k \in \Omega_z^{k(\ell)}$ because of the definition of the environment-related comparability set. As $\Omega_z^{k(\ell)} \cap \Omega_q^{k(\ell)} = \emptyset$, no optimization procedure can be executed. Hence, we have to slightly change some parameters (bandwidths and/or thresholds) to get a nonempty overall comparability set. So, while that set remains empty, we iteratively increase the bandwidths by 5% and to reduce the thresholds by the same rate, until the emptiness problem vanishes. In the case of $k = 100$, with a single iteration, we got $\Omega^{k(\ell)} = \{126, 127, 128, 141, 142, 143\}$. These observations (hospitals) are still better performers than k in terms of quality. Because $k \notin \Omega^{k(\ell)}$ we do not have to use the US-log-DEA, as there is not a problem of multiple solutions. Therefore, we

consider the log-DEA primal, which is computationally faster. The log-linear programming model is as follows:

$$\begin{aligned}
 & \text{[log-DEA primal] maximize } \beta^k + \varepsilon(s_1^k + s_2^k) \\
 & \text{subject to:}
 \end{aligned}
 \tag{29a}$$

$$[14.27, 15.24, 15.18, 15.25, 15.32, 15.36, 15.33]^\top \boldsymbol{\mu}^k + 15.75\beta^k + s_1^k = 15.75,
 \tag{29b}$$

$$- [7.06, 7.09, 6.88, 7.18, 7.12, 7.23, 7.24]^\top \boldsymbol{\mu}^k + 7.64\beta^k + s_2^k = -7.64,
 \tag{29c}$$

$$\langle \sqrt{\boldsymbol{\mu}^k}, \sqrt{\boldsymbol{\mu}^k} \rangle = 1,
 \tag{29d}$$

$$\boldsymbol{\mu}^k \geq 0,
 \tag{29e}$$

$$s_1^k, s_2^k \geq 0.
 \tag{29f}$$

The solution of this problem is $\boldsymbol{\mu}^{k*} = [0, 0, 0, 0, 0, 1]^\top$, which is related to $\beta^k = -0.0524$; hence, k is super-efficient regarding $\Omega^{k(\ell)}$ eventually because of its poor quality. Yet, being super-efficient does not mean that the optimal paid price should be larger than the current unitary cost. In fact, we get $\mathcal{P}^{k(\ell, 0)} = 2183$, which compares with $\hat{\mathcal{P}}^{k(\ell)} = 3336$. Thus, the optimal payment is about 35% smaller than the current unitary cost of k .

By executing the procedure proposed in Sect. 3.7, we obtained the optimal payments per standard patient. The optimal quantity of resources for the 2 years of study was €6,290,336,455, against €8,967,878,333 of consumed resources, which means that nearly 30% of them were wasted. The average payment was €1862 per standard patient, whereas the average current unitary cost is about €3052 (1.64 times the average optimal payment). Figure 6 exhibits the histogram and some parametric probability density functions associated with optimal payments.⁶ Figure 7 does the same for the current unitary costs related to standard patients. Optimal payments seem to be well fitted by a generalized extreme value distribution with $\xi = 0.58[0.46, 0.70]$ (shape parameter), $\sigma = 322.90[292.73, 356.18]$ (scale parameter), and $\mu = 1460.60[1428.39, 1492.81]$ (location parameter).⁷ Meanwhile, current unitary costs are modelled by the t location-scale probability distribution, with $\xi = 3.61[2.71, 4.81]$ (shape), $\sigma = 451.49[410.56, 496.50]$ (scale), and $\mu = 3012.15[2971.61, 3052.69]$ (location).

Optimal payments Optimal payments are well modelled by a generalized extreme value distribution (GEV) with support $p \in [903.88, +\infty[$ and c.d.f. $\text{Prob}(\mathcal{P}^{i(\ell, 0)} \leq p) = \exp\{-v(p)\}$, with $v(p) = 53, 987/(p - 901.98)^{1/0.58}$.

⁶ Plots have been obtained through the program ‘allfitdist.m’ available for MATLAB: <https://github.com/dcherman/tools/blob/master/misc/allfitdist.m>.

⁷ Because the shape parameter is positive, the distribution belongs to the Fréchet family of distributions.

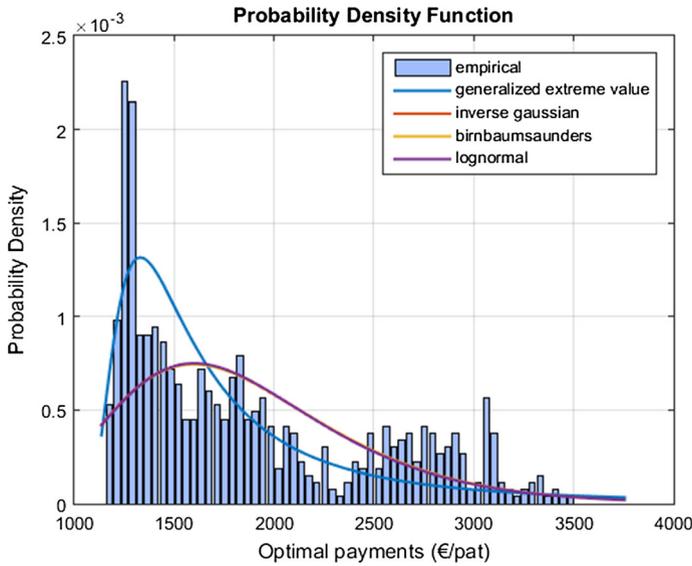


Fig. 6 Histogram of optimal payments

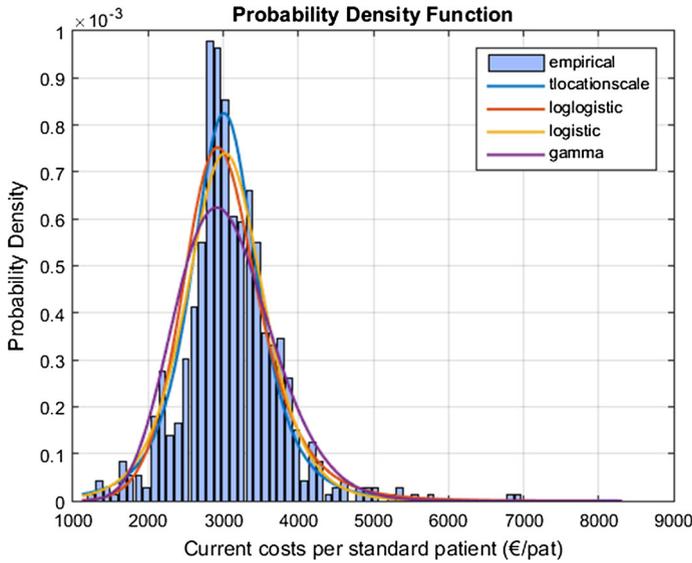


Fig. 7 Histogram of current unitary costs

Accordingly, the mean and the median of the distribution are, respectively, $\mathbb{E}(\mathcal{P}^{i(\ell,0)}) = \mu + \frac{\sigma}{\xi}(\Gamma(1 - \xi) - 1) \approx 2079$ and $\text{median} = \mu + \frac{\sigma}{\xi}(-1 + (\ln 2)^{-\xi}) \approx 1593$ (€ per patient). The variance of this distribution is infinite because the expected shape parameter is larger than 0.5.

Current unitary costs The t location-shape probability density function (with parameters ξ, σ , and μ) fits well the distribution of current unitary costs. As such, the random variable $(\hat{\mathcal{P}}^{j(\ell)} - \mu)/\sigma$ follows a Student's t distribution, with support $\hat{p} \in \mathbb{R}$, mean and median 0, and standard deviation $\sqrt{\xi/(\xi - 2)}$. Therefore, the current unitary costs have mean = $\mathbb{E}(\hat{p}) = \mu = 3012$, median = $\mu = 3012$ (€ per patient), and standard deviation = $\sigma\sqrt{\xi/(\xi - 2)} \approx 676.06$. The c.d.f. associated with the original distribution is

$$\begin{aligned} \text{Prob}(\hat{\mathcal{P}}^{j(\ell)} \leq \hat{p}) &= \frac{\Gamma(\frac{\xi+1}{2})}{\sqrt{\xi}\pi\Gamma(\frac{\xi}{2})} \int_{-\infty}^{(\hat{p}-\mu)/\sigma} du \left(1 + \frac{u^2}{\xi}\right)^{-\frac{\xi+1}{2}} \\ &= \frac{1}{2} + \frac{\hat{p} - \mu}{\sigma} \frac{\Gamma(\frac{\xi+1}{2})}{\sqrt{\xi}\pi\Gamma(\frac{\xi}{2})} F_1^2\left(\frac{1}{2}, \frac{\xi + 1}{2}; \frac{3}{2}; -\frac{(\hat{p} - \mu)^2}{\sigma^2\xi}\right), \end{aligned}$$

where F_1^2 denotes the hypergeometric function. Provided that $\xi = 3.61, \sigma = 451.49$, and $\mu = 3012.15$, we can simplify the c.d.f. as

$$\text{Prob}(\hat{\mathcal{P}}^{j(\ell)} \leq \hat{p}) \approx \frac{1}{2} + \frac{\hat{p} - 3012.15}{1211.70} F_1^2\left(0.5, 2.3050; 1.5; \frac{(\hat{p} - 3012.15)^2}{735874}\right).$$

We can, then, compare $\text{Prob}(\hat{\mathcal{P}}^{j(\ell)} \leq \hat{p})$ with $\text{Prob}(\mathcal{P}^{j(\ell,0)} \leq p)$, and verify after which value of p the probability of the optimal tariff surpasses the current unitary cost. Figure 8 plots the c.d.f. of both optimal tariffs and current unitary costs. Because of the Bolzano–Weierstrass theorem and the monotonic nature of c.d.f., we know that $D(p) = \text{Prob}(\hat{\mathcal{P}}^{j(\ell)} \leq p) - \text{Prob}(\mathcal{P}^{j(\ell,0)} \leq p) = 0$ has a solution in the interval $p \in [3500, 4500]$ €. More precisely, the solution is close to € 3993, associated with $\text{Prob}(\mathcal{P}^{j(\ell,0)} \leq p) = 0.9485$. This means that there is a small probability of 5% of observing hospitals whose optimal tariff is larger than their unitary costs. This result clearly state the advantage of the proposed framework: in 95% of the cases, the model was capable of achieving a tariff lower than the current unitary cost at the same time that important dimensions like quality and operational environment are accounted for. This is an important issue as sustainability should ensure that quality is not hampered.

We also used the stepwise forward multiple regression method to relate optimal payments to environment dimensions:

$$\mathcal{P}^{j(\ell,0)} = 236 \frac{z_1^{j(\ell)}}{\sigma_1} + 76 \frac{z_3^{j(\ell)}}{\sigma_3} + 92 \frac{z_5^{j(\ell)}}{\sigma_5} + 73 \frac{z_6^{j(\ell)}}{\sigma_6} + 367 \frac{z_7^{j(\ell)}}{\sigma_7} + 99 \frac{z_8^{j(\ell)}}{\sigma_8}, j \in \Omega,$$

where $\sigma_{c^{(\ell)}}$ is the standard deviation of environment dimension $c^{(\ell)}$. This regression verifies $R^2 \approx 50\%$, meaning that half of the data variance remains unexplained by the linear model. Note that including second-order terms into the linear model will only improve the coefficient of determination by 3%. Because explanatory variables

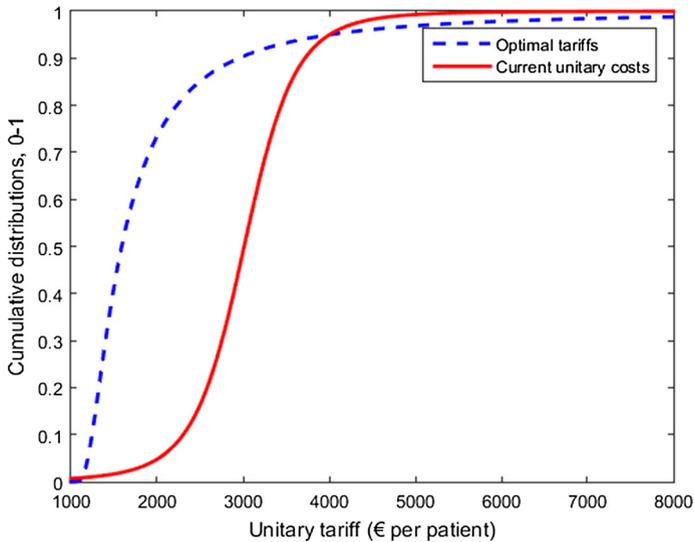


Fig. 8 Cumulative distributions of optimal tariffs and current unitary costs

were standardized, coefficients help us on defining the most important predictors for optimal payments and it becomes clear that these ones are the population density, the illiteracy rate, the case-mix index (complexity), and the Gini's specialization degree. Illiteracy is negatively correlated with the remaining three dimensions. Furthermore, we may foresee that more differentiated facilities also handle with more complex cases and are located in urban regions (mostly in the coastline), which, in turn, are more densely populated and whose citizens have higher literary levels. Those providers are, according to our multiple linear model, the ones whose payment will be the largest.

As we have pointed out before, payments are currently set as the smallest unitary cost within the group in which the hospital is positioned. It takes into account neither quality nor the environment, a pitfall on defining optimal payments. There are four groups of hospitals in Portugal (B, C, D, and E).⁸ Table 3 presents the comparison between the current procedure to fix payments and our approach. As we can see, our approach allows substantial cost savings but these ones represent only 54% of savings that could be achieved through the current procedure. We remark, however, that such a level of cost savings result from ignoring quality dimensions (which usually are costly as they require investments, Karagiannis and Velentzas 2012) and from unfair comparisons (because environment and technology are not considered in the current procedure). Although hospitals are currently clustered based on dimensions related to size, it is not sufficient to make adequate comparisons. It is also interesting to observe that, according to our approach, the potential cost savings decrease with the increase of both size and complexity of the

⁸ There is another one, F, which is composed of oncology centres. This paper does not consider this group because the production technology of these health care providers is quite particular.

Table 3 Comparison between the current procedure to fix payments and our approach

Group	Price (current procedure)	Price (our approach), average	Cost savings (current procedure)/€ 10 ⁶	Cost savings (our approach)/€ 10 ⁶
B	1338	1494	211	195
C	1619	1678	864	832
D	1136	1836	1482	833
E	1843	2633	1674	447
Total			4231	2307

health care providers (in opposition to the current approach). One could think that the larger the hospital, the more resources it could save if it would be efficient; however, this is not linear because these hospitals are also the ones receiving and treating the most complex and resources consuming patients, meaning that the margin for cost savings also decreases given their complex structures—this is, indeed, in line with our approach which adjusts the performance of health care providers to their internal and external operational environment. That is, our approach still contributes to the Portuguese health care system's sustainability, although with smaller cost saving levels when compared to the currently adopted methodology (which, in turn, seems to induce misconducts within the field as an outcome of absurd cost constraints).

Of course, these results may strongly depend on parameters like the threshold and the bandwidths. Thresholds can be understood as the minimum levels of quality to be a potential benchmark. Such levels depend on the stakeholder(s) point of view. We may expect that the smaller the threshold, the larger the cost savings because underinvestments on quality are allowed, which is in line with Proposition 19. Likewise, the definition of bandwidths may play a relevant role on cost savings, see Proposition 20. The extreme case of these two propositions occurs when thresholds are $t_{q_r}^{(\ell)} = 0$ for all $r^{(\ell)} \in \Gamma^{(\ell)}$ and $b_y^{k(\ell)}, b_{q_r}^{k(\ell)}, b_{z_c}^{k(\ell)} \mapsto +\infty$. In this case, $\Omega^{k(\ell)} \mapsto \Omega$, so it is sufficient to apply US-log-DEA altogether the frontier constructed using the entire sample to obtain optimal payments which, in turn, disregard size, environment, and quality. In our empirical application, this results on cost savings of roughly €3145 million, a value that is obviously above the cost savings using our approach with thresholds set by the median of quality indicators, but still below the cost savings according to the current procedure adopted by the Portuguese Ministry of Health. We, then, suspect that the approach adopted by the Ministry is too conservative and induces the health care system's underfunding. Probably because of that several authors have reported a considerable missing of resources in the Portuguese health system (Nunes et al. 2019; Nunes and Ferreira 2019b). Accordingly, our approach is more flexible, contributing for the system's sustainability at the same time that quality of delivered services is ensured and comparisons for benchmarking purposes are fair, being based on both external and internal conditions in which hospitals operate.

6 Concluding remarks and the economic impact to the NHS

This paper proposes and explores a new tool that achieves the best tariffs set for a health care provider. This achievement clearly encompasses both components of quality and financial sustainability, and finances those providers based on their past performance and on the existence of potential benchmarks, which in turn spend fewer resources to treat a comparable quantity of patients, show higher levels of quality, and act on similar environment (inner/outer). Our approach can optimize payments in several payment schemes extensively adopted worldwide.

The proposed tool has several advantages. As a matter of fact, it has a clear impact on the health care financial system, since costs are reduced as a function of providers' performance. Considerable potential cost savings arise as a result of the application of the new tool. Additionally, potential benchmarks cannot show quality levels below the stakeholder(s) expectations. It is not clear whether higher quality levels do imply or not higher costs. Although high quality-based facilities are usually expensive, in the long-run it may contribute to the reduction of hospital days, postoperative complications, inner-hospital infections, readmissions, and deaths. Furthermore, high quality hospitals tend to attract more patients. Since *money follows the patient*, the competition between providers is expected to improve their performance and to reduce wasted resources. Competition is commonly associated with performance improvements. Due to the competition, activity tends to increase; assuming that it is cost-effective and appropriate, it becomes another advantage of the proposed strategy (Geissler et al. 2011). If scope and scale economies can be exploited, then the more patients treated, the lower the provider cost per patient. If the provider is able to join both cost savings due to scale economies and high quality standards, it becomes a potential benchmark in our framework. If an inefficient hospital achieves a tariff lower than its current unitary cost, then it has room to improve its practices, once there is, at least, one unit producing roughly the same amount of outputs, with fewer resources and better quality. That is, the inefficient one should learn from its benchmarks' practices. By reducing the payment for such a hospital, it can increase its income by increasing its production. But then again, the *invisible hand of competition* will impose that production increasing is possible if and only if quality is improved. In other words, the reduction of payments for some hospitals will force their production increase on the one hand, and reduce the total costs and wastefulness in the long-run if scale economies are correctly explored, on the other hand.

Additionally, low-quality units may receive a tariff higher than their current cost per patient, which should be seen as an incentive to improve their quality by introducing quality-raising through cost-increasing new technologies (Quentin et al. 2011; Marshall et al. 2014). Even for very high quality thresholds, this scenario can happen for most (if not all) units if the provider with the best quality standards also exhibits significant signs of technical inefficiency. In such a case, the stakeholder(s) must decide whether the threshold is in fact too high, or an incentive to quality improvement across the entire sample shall be *placed on the table*. If the last scenario is adopted and both quality and efficiency are not positively related, the

financial sustainability can be threatened. This means that the stakeholder(s) has a prominent role, by defining a priori which criteria must be utilised to define *quality* and how to define an appropriate threshold. This one can be perhaps optimized to maximise potential cost savings, which is left for a further research.

As in the case-based payment scheme, several disadvantages or difficulties can be identified for the new financing tool. Yet, one must be aware that there is not an available flawless reimbursement method. We can only try to minimize the shortcomings of available schemes. Most of those disadvantages identified for the new scheme (our proposal) can be reduced by implementing strong contract management tools. For instance, State's financial control can be improved if volume of care is specified in contract. Quality standards are also set. If they are not accomplished, pre-defined penalties can be applied in further contracts, reducing the new budget. Also, selecting the least complex and severe patients must also be prevented in contracts. However, some disadvantages still remain. Indeed, if payment is made according to DRGs, the incorrect patients' registry is likely in practice and can be even promoted through the new tool. For instance, a patient from a specific DRG is assigned to a different DRG group (Vita 2001), simply because contracted volumes must be accomplished and/or the income must be increased to overcome some mismanagement scenarios. It is not so simple to mitigate such a situation, especially because there is a resistance on medical group as well as a pre-defined DRG quantity to be *produced*. Registry systems are different among providers, thus they must be updated and uniformed, and an integrated system should be created. It must cross and relate data from different levels of care (primary, secondary, tertiary), and detect potential bias on expected treatments for a specific patient. Another alternative would be increasing the contracted production range, which would mitigate the aforementioned misregistration.

Finally, by creating a database with the production of all hospitals, costs and quality, for several years, it is possible to refine the potential best practice search for each entity, as few years may ensure a larger range of possible optimal payments. For instance, a hospital may present an optimal current unitary cost, which could be non-optimal in the past. Thus, its corresponding tariff will be lower than its current cost per patient. Accordingly, we expect a convergence of health care providers' performance in the long-run.

Acknowledgements This research was generously supported by the project 'hSNS: Portuguese public hospital performance assessment using a multi-criteria decision analysis framework' (PTDC/EGE-OG/30546/2017), funded by the Portuguese Foundation for Science and Technology (FCT, from the Portuguese abbreviation of Fundação para a Ciência e a Tecnologia) (Grant No. 02/SAICT/2017/30546). The first author also acknowledges the financial support by FCT: Grant SFRH/BD/113038/2015. The authors would like to express their gratitude to two anonymous referees who kindly and significantly have improved this paper's quality, clarity, and structure, due to their beneficial comments. The contents of the paper are the authors' own responsibility. The usual disclaimer applies.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The

images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agrell PJ, Tind J (2001) A dual approach to nonconvex frontier models. *J Prod Anal* 16:129–147
- Bádin L, Daraio C, Simar L (2010) Optimal bandwidth selection for conditional efficiency measures: a data-driven approach. *Eur J Oper Res* 201(2):633–640
- Banker RD, Maindiratta A (1986) Piecewise loglinear estimation of efficient production surfaces. *Manag Sci* 32:126–135
- Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manag Sci* 30(9):1078–1092
- Barros PP, Machado SR, Simões J (2011) Portugal: health system review. *Health Syst Transit* 13(4):1–140
- Chambers RG, Chung Y, Färe R (1996) Benefit and distance functions. *J Econ Theory* 70(2):407–419
- Chambers RG, Chung Y, Färe R (1998) Profit, directional distance functions, and nerlovian efficiency. *J Optim Theory Appl* 98(2):351–364
- Chang KP, Guh YY (1995) Piecewise loglinear frontiers and log efficiency measures. *Comput Oper Res* 22(10):1031–1037
- Charnes A, Cooper WW, Rhodes E (1979) Measuring the efficiency of decision-making units. *Eur J Oper Res* 3(4):201–209
- Charnes A, Cooper WW, Seiford LM, Stutz J (1982) A multiplicative model for efficiency analysis. *Socio-Econ Plan Sci* 16:223–224
- Chen Y, Du J, Huo J (2013) Super-efficiency based on a modified directional distance function. *Omega Int J Manag Sci* 41:621–625
- Cherchye L, Moesen W, Rogge N, Puyenbroeck TV (2011) Constructing composite indicators with imprecise data: a proposal. *Expert Syst Appl* 38(9):10940–10949
- Cooper WW, Ruiz JL, Sirvent I (2007) Choosing weights from alternative optimal solutions of dual multiplier models in DEA. *Eur J Oper Res* 180:443–458
- Cordero JM, Polo C, Santín D (2018) Assessment of new methods for incorporating contextual variables into efficiency measures: a Monte Carlo simulation. *Oper Res*. <https://doi.org/10.1007/s12351-018-0413-2>
- Daidone S, D'Amico F (2009) Technical efficiency, specialization and ownership form: evidences from a pooling of Italian hospitals. *J Prod Anal* 32:203–216
- Daraio C, Simar L (2007) *Advanced robust and nonparametric methods in efficiency analysis*. Springer, New York
- Davoodi A, Zarepisheh M, Rezai HZ (2015) The nearest MPSS pattern in data envelopment analysis. *Ann Oper Res* 226(1):163–176
- Deprins D, Simar L, Tulkens H (1984) Measuring labor efficiency in post offices. In: Marchand M, Pestieau P, Tulkens H (eds) *The performance of public enterprises: concepts and measurements*. Elsevier, Amsterdam, pp 247–263
- Emrouznejad A, Cabanda E (2010) An aggregate measure of financial ratios using a multiplicative DEA model. *Int J Financ Serv Manag* 4:114–126
- Ferreira DC, Marques RC (2015) Did the corporatization of Portuguese hospitals significantly change their productivity? *Eur J Health Econ* 16(3):289–303
- Ferreira DC, Marques RC (2017) A step forward on order- α robust nonparametric method: inclusion of weight restrictions, convexity and non-variable returns to scale. *Oper Res*. <https://doi.org/10.1007/s12351-017-0370-1>
- Ferreira DC, Marques RC (2018a) Identifying congestion levels, sources and determinants on intensive care units: the Portuguese case. *Health Care Manag Sci* 21:348–375

- Ferreira DC, Marques RC (2018b) Do quality and access to hospital services impact on their technical efficiency? *Omega Int J Oper Res*. <https://doi.org/10.1016/j.omega.2018.07.010>
- Ferreira DC, Marques RC, Nunes AM (2017) Economies of scope in the health sector: the case of Portuguese hospitals. *Eur J Oper Res* 266(2):716–735
- Figueras J, Robinson R, Jakubowski E (2005) Purchasing to improve health systems performance. European observatory on health systems and policies series. Open University Press, Berkshire
- Friesner DL, Rosenman R (2009) Do hospitals practice cream skimming? *Health Serv Manag Res* 22(1):39–49
- Fukuyama H, Weber WL (2017) Directional distance DEA models. In: Tone K (ed) *Advances in DEA theory and applications: with extensions to forecasting models*, chapter 3. Wiley, Hoboken, pp 20–29
- Galizzi MM, Miraldo M (2011) The effects of hospitals' governance on optimal contracts: bargaining vs. contracting. *J Health Econ* 30(2):408–424
- Geissler A, Quentin W, Scheller-Kreinsen D, Busse R (2011) Introduction to DRGs in Europe: common objectives across different hospital systems. In: Busse R, Geissler A, Quentin W, Wiley M (eds) *Diagnosis-related groups in Europe: Moving towards transparency, efficiency and quality in hospitals*. European observatory on health systems and policies series. Open University Press, New York
- Hall P, Racine JS, Li Q (2004) Cross-validation and the estimation of conditional probability densities. *J Am Stat Assoc* 99(486):1015–1026
- Henderson JM, Quandt RE (1980) *Microeconomic theory: a mathematical approach*. McGraw-Hill, New York
- Hollingsworth B (2008) The measurement of efficiency and productivity of health care delivery. *Health Econ* 45:1107–1128
- Johannessen KA, Kittelsen SAC, Hagen TP (2017) Assessing physician productivity following norwegian hospital reform: a panel and data envelopment analysis. *Soc Sci Med* 175:117–126
- Kao C (2017) *Network data envelopment analysis: foundations and extensions*. International series in operations research and management science, vol 240. Springer International Publishing, Berlin
- Karagiannis R, Velentzas K (2012) Productivity and quality changes in Greek public hospitals. *Oper Res* 12(1):15–27
- Li Q, Racine JS (2007) *Nonparametric econometrics: theory and practice*. Princeton University Press, New York
- Lindlbauer I, Schreyögg J (2014) The relationship between hospital specialization and hospital efficiency: do different measures of specialization lead to different results? *Health Care Manag Sci* 17:365–378
- Lu M, Ma C, Yuan L (2003) Risk selection and matching in performance-based contracting. *Health Econ* 12:339–354
- Marshall L, Charlesworth A, Hurst J (2014) The NHS payment system: evolving policy and emerging evidence. Nuffield Trust, pp 1–44. <https://www.nuffieldtrust.org.uk/research/the-nhs-payment-system-evolving-policy-and-emerging-evidence>
- Mckee M, Healy J (2002) *Hospitals in a changing Europe*. Open University Press, Buckingham
- Mehdiloozad M, Sahoo BK, Roshdi I (2014) A generalized multiplicative directional distance function for efficiency measurement in DEA. *Eur J Oper Res* 232(3):679–688
- Nunes AM, Ferreira DC (2019a) The health care reform in Portugal: outcomes from both the new public management and the economic crisis. *Int J Health Plan Manag* 34(1):196–215
- Nunes AM, Ferreira DC (2019b) Reforms in the Portuguese health care sector: challenges and proposals. *Int J Health Plan Manag* 34(1):e21–e33
- Nunes AM, Ferreira DC, Campos AF (2019) Financial crisis in Portugal: effects in the health care sector. *Int J Health Serv* 49(2):237–259
- Podinovski VV (2016) Optimal weights in dea models with weight restrictions. *Eur J Oper Res* 254(3):916–924
- Podinovski VV, Bouzdine-Chameeva T (2015) Consistent weight restrictions in data envelopment analysis. *Eur J Oper Res* 244(1):916–924
- Podinovski VV, Bouzdine-Chameeva T (2016) On single-stage dea models with weight restrictions. *Eur J Oper Res* 248(3):1044–1050
- Portela MCAS, Thanassoulis E (2006) Malmquist indexes using a geometric distance function (GDF): application to a sample of Portuguese bank branches. *J Prod Anal* 25:25–41
- Quentin W, Scheller-Kreinsen D, Busse R (2011) Technological innovation in DRG-based hospital payment across Europe. In: Busse R, Geissler A, Quentin W, Wiley M (eds) *Diagnosis-related*

- groups in Europe: moving towards transparency, efficiency and quality in hospitals. European observatory on health systems and policies series. Open University Press, New York
- Ray SC (2008) The directional distance function and measurement of super-efficiency: an application to airlines data. *J Oper Res Soc* 59(6):788–797
- Sakellariades C (2010) Repensando a contratualização. *Revista Portuguesa de Saúde Pública* 9:5–6
- Seiford LM, Zhu J (1998) On piecewise loglinear frontiers and log efficiency measures. *Comput Oper Res* 25(5):389–395
- Shimshak DG, Lenard ML, Klimberg RK (2009) Incorporating quality into data envelopment analysis of nursing home performance: a case study. *Omega Int J Manag Sci* 37(3):672–685
- Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London
- Street A, O'Reilly J, Ward P, Mason A (2011) DRG-based hospital payment and efficiency: theory, evidence and challenges. In: Busse R, Geissler A, Quentin W, Wiley M (eds) *Diagnosis-related groups in Europe: moving towards transparency, efficiency and quality in hospitals*. European observatory on health systems and policies series. McGraw-Hill Education, Maidenhead
- Suayoshi T, Chang Y (1989) Efficient algorithm for additive and multiplicative models in data envelopment analysis. *Oper Res Lett* 8:205–213
- Sueyoshi T, Sekitani K (2009) An occurrence of multiple projections in dea-based measurement of technical efficiency: theoretical comparison among DEA models from desirable properties. *Eur J Oper Res* 196(2):764–794
- Tiedemann T, Francksen T, Latacz-Lohmann U (2011) Assessing the performance of german bundesliga football players: a non-parametric metafrontier approach. *Cent Eur J Oper Res* 19:571–587
- Tofallis C (2014) On constructing a composite indicator with multiplicative aggregation and the avoidance of zero weights in DEA. *J Oper Res Soc* 65(5):791–792
- Valadkhani A, Roshdi I, Smyth R (2016) A multiplicative environmental DEA approach to measure efficiency changes in the world's major polluters. *Energy Econ* 54:363–375
- Valente M (2010) Contratualização em contexto hospitalar. *Revista Portuguesa de Saúde Pública* 9:25–39
- Vita MG (2001) Regulatory restrictions on selective contracting: an empirical analysis of any-willing-provider regulations. *J Health Econ* 20(6):955–966
- Zarepisheh M, Khorram E, Jahanshahloo GR (2010) Returns to scale in multiplicative models in data envelopment analysis. *Ann Oper Res* 173:195–206

Affiliations

Diogo Cunha Ferreira^{1,2}  · Rui Cunha Marques¹ · Alexandre Morais Nunes^{1,2}

Rui Cunha Marques
rui.marques@tecnico.ulisboa.pt

Alexandre Morais Nunes
alexandre.m.nunes@tecnico.ulisboa.pt; anunes@iscsp.ulisboa.pt

¹ CERIS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

² CAPP, Centre for Administration and Public Policies, Instituto Superior de Ciências Sociais e Políticas, Universidade de Lisboa, Lisbon, Portugal

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.