

Should inpatients be adjusted by their complexity and severity for efficiency assessment? Evidence from Portugal

Diogo Cunha Ferreira · Rui Cunha Marques

Received: 2 October 2013 / Accepted: 21 May 2014 / Published online: 3 June 2014
© Springer Science+Business Media New York 2014

Abstract Hospital efficiency analysis depends largely on the model specifications. This study discusses the importance of the case-mix index (CMI) to homogenize the sample of inpatient discharges. It proposes a new index where they are classified by service, since it is usual to have lack of data to compute the CMI and this can influence the credibility of results. Data from the Portuguese national diagnosis-related group (DRG) database was utilized. Three different approaches are developed in this paper, based on locally convex order- m method as well as on translog functions. The first one correlates the efficiency with different inpatients weighting schemes, by using the Nadaraya-Watson method. The second approach compares different frontiers that have been computed using the different weighting schemes. Finally, by using bootstrap, the paper investigates whether the inclusion of severity/ complexity-related variables in the model statistically modifies the results. It has been shown that, under the Portuguese healthcare framework, if the model is environment corrected (which should include epidemiological and main political/ structural health reforms variables), then the severity adjustment of inpatients is pointless. The employment of an inpatient-weighting scheme, such as the CMI, may introduce significant frontier shift, thus its absence is not recommended in productivity evolution analyzes. The CMI shifts the efficiency frontier, but not the relative position of units against it (the last scenario if exogenous variables are present).

Keywords Case-mix index · Service-mix index · Internment services efficiency · Locally convex order- m method · Translog function

1 Introduction

The health care industry faces new challenges every day, such as new regulations or new technologies. Managers and policy makers need to respond to these challenges with sound performance evaluation [1]. In the health sector, the outputs (products) are difficult to define, since *health* is a complex concept, there is no traditional health market [2,3], and there is lack of homogeneity between the output measurements, as in case of inpatients [4]. Differences between hospitals are due to factors such as technology, complexity and severity differences and legal status. There are several ways to try to homogenize that information, e.g. using a service-mix index (for structures and facilities) or a case mix (for complexity and severity of patients). There is a special care in some studies to adjust the hospital production in terms of complexity, using the case mix index (CMI), especially for inpatients, e.g. [5–10]. Indeed, highly aggregated variables (such as gross measures of hospital services, e.g. the total number of inpatients) may, however, harm the efficiency comparisons against hospitals that treat a relatively high proportion of severely ill patients [11].

Nonetheless, only few studies have been performed to evaluate the implication of such index on the efficiency scores (ES). While some studies conclude that there are no differences in technical ESs whether the case mix is incorporated or not [12–15], others conclude that there are statistically significant differences in the ES distribution due to different case mix specifications [16–18]. However, other authors studied the limitations of case mix grouping and found that such instrument should be applied with caution both for monitoring

D. C. Ferreira (✉)
Instituto Superior Técnico (IST), University of Lisbon, Av. Rovisco Pais, 1049-001 Lisbon, Portugal
e-mail: diogo.cunha.ferreira@tecnico.ulisboa.pt

R. C. Marques
Centre for Urban and Regional Systems (CESUR), DECivil-IST, University of Lisbon, Av. Rovisco Pais, 1049-001 Lisbon, Portugal
e-mail: rui.marques@tecnico.ulisboa.pt

and financing purposes at the department level [19]. Concerning the Portuguese case, Dismuke and Sena [20] concluded that the inclusion of the diagnosis-related group (DRG) payment “has had a positive impact on productivity and technical efficiency of some commonly employed diagnostic technologies”. Furthermore, according to Yang and Reinke [21] both classification systems (ICD and DRG) generate similar case mix-based indexes.

Therefore, there is a noteworthy discussion about the influence of the CMI (as the most frequent risk adjustment way) on hospital efficiency. Nevertheless, the literature does not address conveniently the effect of environmental variables, which may be considered as good proxies of the complexity/risk/ severity of illness. Furthermore, most of them use deterministic efficiency methods, and it is well known that they may entail some problems (see *infra* for details).

The objective of this paper is threefold. It is intended to answer three different questions: (1) Does risk adjustment (through the CMI) really matter for efficiency assessment? (2) Can risk adjustment be replaced by a hospital service complexity-based index without major efficiency measurement changes? To do so, a new service-mix index is proposed and analyzed. (3) Can risk adjustment be done through the inclusion of some environmental variables (such as the morbidity)?

This study is organized as follows: after the CMI has been characterized in Section 2, a new index (service-mix based) is proposed in Section 3. An overview of the main methods is provided in Section 4 and the main results are presented in Section 5. The conclusions, results discussion and further research topics are given in the last section.

2 The case-mix index

Despite the existence of multiple services in a hospital, it might not be correct to include them all in the same model, due to the necessity of considering a lot of outputs (and even inputs). Then, only the internment department is considered in this study, due to lack of data available. The main outputs are the inpatients. However, they should be adjusted by risk because “not all patients arriving at the hospital require the same level of attention and service” [1, pp. 107. It seems to be clear that the inpatients risk adjustment is necessary [13,22–24], at least theoretically, although some authors (e.g. [25]) have noted that this adjustment may be a fallacy. Being the CMI the most common risk adjustment (for details, see *infra*), there are two main approaches to do that: (1) by multiplying the number of inpatients (admissions or discharges) by that index—given the way it is computed (see *infra*), this seems to be the most appropriate procedure, or (2) by considering the CMI as an extra (non-discretionary) variable, as discussed below (see [14]). Both approaches are tested

in this paper so as to assess the model sensitivity to this variable (or adjustment).

The most common case mix classification scheme is the DRG system, which is a set of patient case types established under a disease and injury classification system that identifies patients with similar illness conditions and processes of care (assigning the DRGs with a weighting factor that compares its costliness to the average for all DRGs under a case-mix funding scheme) [11]. The DRG system has been employed in Portugal since 1984 [20]. The DRG system reflects the distribution of patients based on morbidity, patient characteristics such as age, and aspects of the treatment received. Each DRG represents a class of patients with similar clinical conditions, hence requiring similar resource usage for treatment.

Associated with DRG groups, there is the CMI, which can be computed using the Eq. (2.1), where (1) it is assumed that two patients in the same DRG group are equivalent, i.e., the expenses are the same for both, and (2) the weight of the k -th DRG is the relationship between its average costs and the national costs, then it measures the expected costs for an average inpatient (of this group). The CMI reflects the complexity of the treatments provided and the level of resources consumed in a hospital [21].

$$CMI_{hospital\ k} = \frac{\sum_i(\text{equivalent inpatients } DRG_i)_k \times \text{weight } DRG_i}{\sum_i(\text{equivalent inpatients } DRG_i)_k} \quad (2.1)$$

The national CMI is unitary—any hospital with $CMI < 1$ (resp. > 1) treats, on average, patients with low (high) complexity. That is, the higher the CMI, the higher the average diseases complexity the hospital handles. In the Portuguese case, the CMI is only available for inpatient and ambulatory services, medical or surgical. In general, the CMI for surgical services is greater than the CMI for medical services.

3 The service-mix index: a new approach

Despite the theoretical importance of weighting the inpatients by their clinical complexity (morbidity), it is not unusual that the CMI values may be often non-available in hospital reports. Analogously to the CMI, one can define a new index, changing the patient classification into DRG to a patient classification by service¹—for this research purpose, all considered services belong to the internment production line, i.e., all specialty wards where patients are admitted for a period longer or shorter, including both medical and surgical services/specialties and intensive care units.

¹ This change of patient classification resembles the work of Yang and Reinke [21].

This new index is called Service-Mix Index (SMI) because it measures the complexity of hospitals by their services, meaning that if a hospital has services that consume more resources than the national average, then its SMI is greater than 1. This index allows homogenizing the hospital services and, furthermore, it has some advantages relative to the CMI: a) it can be computed with respect to any baseline (e.g., the global internment service or a lower/ higher hospital level)² and b) it requires much less data (only the costs and production), which are usually available. The SMI also takes into account the complexity of inpatients, if it is accepted that two patients treated in the same service have similar disease complexity.

Ozcan and Luke [26] and Ozcan [27] discuss the importance of some proxies of capital investments, namely the service-mix that measures the hospital services complexity. However, that complexity is frequently associated with the inpatients severity/ complexity/ risk—the higher the patient's illness severity, the higher the capital requirements in the hospital service where the patient is treated. For instance, an Intensive Care Unit (ICU) treats pathologies with severe clinical profiles, thus a priori it requires more inputs than other services. Consequently, a hospital with ICU services (and/ or other equally complex services) has a priori a higher CMI. Therefore, the SMI may be a similar appropriate inpatients weight, or an extra variable, as it will be discussed further.

The formula of the SMI is equal to the CMI, replacing the DRG weight by the service weight (sw), as indicated in Eq. (3.1). Weights could be estimated as follows. First, the geometric mean of unitary costs for the internment global service ($GMIGS$) is computed. Next, the geometric mean of unitary costs for each service (GMS_i) is determined and divided by $GMIGS$, i.e., $sw_i = GMS_i / GMIGS$. Then, it is easy to conclude that services with higher relative weights present higher costs. It may be observed, from Eq. (3.1) that the same patient can be counted more than once. Actually, one patient can be admitted to more than one service during the same hospital stay. However, for two different services, the same patient is considered twice. As a consequence, the SMI is an aggregate index of all services complexity.

$$SMI_{hospital\ k} = \frac{\sum_i (inpatients\ of\ service_i)_k \times sw_i}{\sum_i (inpatients\ of\ service_i)_k} \quad (3.1)$$

In Portugal, for instance, in 2009, the geometric mean of the unitary costs of all internment services was 2,776€ ($GMIGS$ per inpatient). Thirteen hospitals provide “Convalescence Units” services (CUS), such that the geometric mean

²Note that this is an advantage over the CMI since the latter one is computed concerning the national standard, which complicates some desirable international comparisons. Moreover, if the CMI is computed for the whole hospital, it hampers the analysis of efficiency of only one service or department (as the internment).

of their unitary costs was 5,566€ (GMS_{CUS} per patient), thus the weight of this kind of service is 2.01 ($sw_{CUS} = 5,566 / 2,776$), i.e., this service consumes, on average, about twice more resources than the national standard (taken as unitary). Meanwhile, 7 hospitals provide “Palliative Care Units” services (PCUS) such that $GMS_{PCUS} = 3,984€$ (per patient), then $sw_{PCUS} = 3,984 / 2,776 = 1.44$. Suppose that two hospitals provide only these two services— H_1 treats 100 inpatients in CUS and 200 inpatients in PCUS, thus $SMI_{Hospital1} = (100 \times 2.01 + 200 \times 1.44) / (100 + 200) = 1.63$; H_2 treats 500 inpatients in CUS and 50 in PCUS, thus $SMI_{Hospital2} = (500 \times 2.01 + 50 \times 1.44) / (500 + 50) = 1.96$; hospital H_2 has a higher SMI because it treats more inpatients than H_1 in the service that consumes more resources due to its implicit complexity (kind of technology and drugs, staff specialization level and inherent inpatients complexity).

The comparison of CMI and SMI can be observed in Fig. 1. A simple correlation was performed to assess the Pearson's coefficient $R = 0.7840$, which means that the linear equation describes the relationship between CMI and SMI with a strong correlation, i.e., there is a strong linear relationship between the global complexity of inpatients and the global complexity of the hospital internment services. Thus, the SMI could be considered as a good proxy of CMI, which means that the most complex patients are generally associated with the most complex services. The fact that the relationship between CMI and SMI is not perfectly linear results from the possible existence of a significantly broad spectrum of patients' complexities in the same service, especially regarding the less complex services, or eventually due to the fact that evidence based protocols are sometimes vague.

It should be noted that there are several definitions of service-mix in the literature. In particular, some studies (e.g. [1,28,29]) describe and use the service-mix as the non-weighted quantity of services provided by the hospital, from a list of possible 151 services identified by the American Hospital Association. Meanwhile, Chou et al. [30] develop a relative weighted service-mix since the non-weighted one does not correctly differentiate the hospitals, because there are services that require larger financial investment due to their inherent complexity and their patients' severity/ complexity (which is the basis of the new SMI developed in this paper). This weighted service-mix scoring is based on two different dimensions (clinical intensity and investment intensity). Nonetheless, such measure is categorical and one may argue that even between two similar services, the same scoring on clinical and/or investment intensities may result into biased measurements of the service complexity. Thus, the SMI computed by Eq. (3.1), being a continuous variable, seems to be more appropriate to differentiate the services concerning their complexities. It has a further advantage: unlike the standard service-mix measure (as previously defined), it is able to adjust the inpatients since they are included in the computation method.

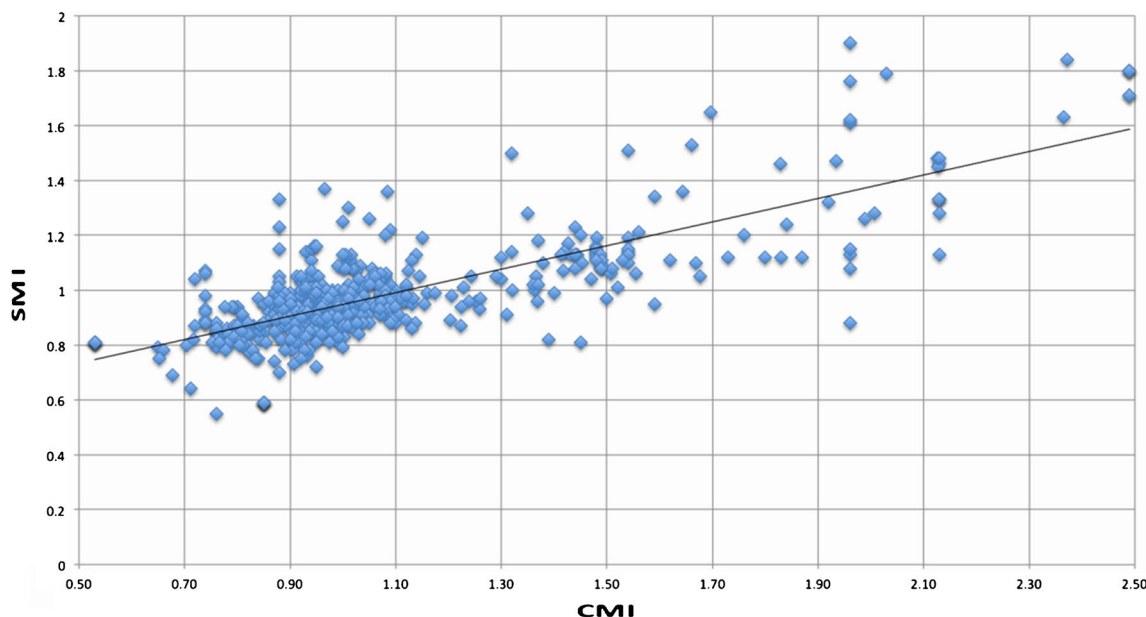


Fig. 1 Correlation between CMI and SMI: $R=0.7840$, $SMI=0.3750 \times CMI+0.6625$

Furthermore, the SMI (1) is comparable to the CMI and may replace it (at least, theoretically) when that one is not available, (2) it is easy to compute—one only needs the total costs and the production, per service, and (3) it is able to represent the capital inputs, as well as the standard approach [26,28].

4 Efficiency of Portuguese hospitals: the internment case

4.1 Methods

This paper is based on two different efficiency estimation approaches, which are based on frontier analysis. The locally convex non-parametric order- m is employed to achieve robust efficiency estimation, without strong frontier assumptions. This procedure is quite useful since it allows the inclusion of environmental effects in the model; furthermore, being a partial frontier method, it is much less sensitive to outliers and extreme data and to the curse of dimensionality than full frontier methods, such as DEA or FDH. Non-parametric methods only require few frontier assumptions (RTS, disposability of variables and convexity/non-convexity), thus they are not based on frontier functional specifications [31,32].

Let $X \in \mathbb{R}_+^p$ be the set of p inputs consumed to produce $Y \in \mathbb{R}_+^q$ outputs by n Decision Making Units (DMU), under the influence of $Z \in \mathbb{R}^s$ exogenous (or environmental) factors. The conditional, locally convex order- m ES, θ , of the unit $(x, y | z)$ projected against the set $(X, Y | Z)$ can be computed as follows (after an appropriate m has been defined by the user) [33]:

1. For a given y , draw a sample of size m with replacement, and with a probability $K((z-Z)/h)/\sum_{j=1}^n K((z-Z_j)/h)$,

among those X such that $Y \geq y$ —denote this sample by $\mathcal{L}=(X_{1,b} \dots X_{m,b})$;

2. Compute $\theta_b = \min_{\lambda, \theta} \left\{ \theta \left| \begin{array}{l} \theta x_i \geq \sum_{j=1}^m \lambda_j X_{ji}, i = 1, \dots, p; \\ y_r \leq \sum_{j=1}^m \lambda_j Y_{jr}, r = 1, \dots, q; \\ \sum_{j=1}^m \lambda_j = 1; \\ \forall_{j=1, \dots, m} \lambda_j \geq 0; \\ X \in \mathcal{L} \end{array} \right. \right\};^3$

3. Redo previous steps for $b=1 \dots B$, where B is large ($\gg 500$) that allows tuning the estimation quality⁴;
4. Finally, $\theta(x, y) \cong \frac{1}{B} \sum_{b=1}^B \theta_b$.

In the first step of the previous procedure, K represents a kernel function that works as a probability density function (*pdf*), and h is a bandwidth that is useful to restrict the entities to be selected in the first step of the previous procedure

³ The order- m assumes variable returns to scale (VRS), by default. This assumption is always consistent either the frontier exhibits VRS or not. As a matter of fact, even if the hospitals act on their own optimal scale, i.e., under constant returns to scale (CRS), the VRS assumption is reliable as well. In order to keep the method’s consistency, VRS is assumed for all approaches, although some authors argue that it is not a good choice if the Malmquist index is under analysis. This is not a problem if the technology exhibits CRS, i.e., if there are no significant scale economies, which is assumed here. This is an open question and the reader should be aware of it.

⁴ Note that this procedure assumes the resampling with reposition and following a probability density function (a kernel, in this case) from a finite or infinite sample. Hence, it may not be equivalent to some (even most commonly) procedures when one samples without reposition and from a finite population. Nevertheless, the stochastic nature of order- m presents clear advantages over the remaining, as explained before. The reproducibility of the results can be affected by the model’s choice and the reader must be aware of that.

(if $h \rightarrow \infty$, then the drawing process is equiprobable among all units, no matter what kernel function was chosen). The most common kernel function is the Gaussian one, and its optimal h is given by the normal reference rule as follows [34]:

$$h = 1.06 \cdot \min\{\sigma, R_{23}/1.34\} \cdot n^{-\frac{1}{5}} \quad (4.1.1)$$

In the previous equation, σ and R_{23} are the empirical standard deviation and the interquartile range of the n data points, respectively. In the case of multivariate Gaussian kernels (several environmental variables), it is assumed that the global *pdf* is the product of all *pdf*'s although some others methods have been proposed in the literature [32]. Indeed, there is no noticeable reason to choose one method in particular: the *pdf* product seems, however, to be the simplest procedure.

Since the fourth step of order- m has been concluded, one may fit the set of efficiency estimations by employing a parametric function [31].⁵ It is employed to strengthen the conclusions. The employment of both parametric and nonparametric measures in health care can be found in several studies, e.g. [35], although most of the health care analyses follow nonparametric approaches [36]: some examples include O'Neill [37], Hollingsworth et al. [38], Bjorkgren et al. [39] and Biorn et al. [40]. Following Daraio and Simar [32], the input-oriented translog function is given by:

$$\begin{aligned} \ln \frac{1}{\theta} = & \alpha_0 + \sum_{i=1}^p \alpha_i \cdot \ln X_i + \sum_{r=1}^q \beta_r \cdot \ln Y_r \\ & + \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \Gamma_{ij} \cdot \ln X_i \cdot \ln X_j \\ & + \frac{1}{2} \sum_{r=1}^q \sum_{w=1}^q \Gamma_{rw} \cdot \ln Y_r \cdot \ln Y_w \\ & + \frac{1}{2} \sum_{i=1}^p \sum_{r=1}^q \Gamma_{ir,1} \cdot \ln X_i \cdot \ln Y_r \\ & + \frac{1}{2} \sum_{r=1}^q \sum_{i=1}^p \Gamma_{ir,2} \cdot \ln X_i \cdot \ln Y_r \end{aligned} \quad (4.1.2)$$

s.t.

$$\sum_{i=1}^p \alpha_i = 1;$$

$$\sum_{r=1}^q \Gamma_{ir,1} = 0 \text{ (} q \text{ constraints);}$$

$$\sum_{j=1}^p \Gamma_{ij} = 0 \text{ (} p \text{ constraints)}$$

In the previous equation, $\alpha_0, \alpha_i, \beta_r, \Gamma_{ij}, \Gamma_{rw}, \Gamma_{ir,1}$ and $\Gamma_{ir,2}$ are the model coefficients that are computed in the least-squares sense subjected to $1 + p + q$ homogeneity (of degree one in X) constraints, which follow an input-oriented framework. Daraio and Simar [32] argue that it overcomes the difficulties of classical approximations [37,38], namely (1) the restrictive and non-realistic assumption of homoscedasticity, and (2) some problems of consistency between the error term and the dependent term. Indeed, the previous translog model uses the projected values on the non-parametric frontier and does not make any assumption on the error term, which results into more sensible results [32].

4.2 Modeling issues—an overview

In this paper, three different approaches are proposed and developed in the following sections. Although some methodological details are described there, some common issues may have already been defined elsewhere.

4.2.1 Data collection

All required data for this research is available in the official database of the Portuguese Ministry of Health, the Central Administration of Health Systems (CAHS), and in lawful annual reports of each hospital, for the period between 2002 and 2009 (8 years). Only this period has been considered due to the readiness of data.⁶

Moreover, only 49 hospitals (among all Portuguese units) could be selected from this database, since they were the ones that provided reliable information without gaps for all the years and all the variables. Maternities, oncology centers and psychology centers (specialized hospitals found in Portugal) were excluded because their own production process and structures are expected to be significantly different from those non-specialized units considered in this paper. No private hospitals have been included in the sample, as well. Therefore, 392 units, homogeneously distributed in 2002–2009, comprise this sample.

4.2.2 Variables

Data are evaluated under the most important hospital dimension—the internment department—given its expenditure volume (~50 %). Furthermore, the CMI is only available for this department, which is why no other departments have been chosen to this study, although all hospitals in this study provide outpatient services. The choice of variables followed the availability of data and the purpose of the paper, which focuses on the case of technical efficiency (no outcomes measures are included here, mainly due to absence of this kind of information in official databases).

⁵ Statistical properties and methods (such as bootstrapping) are detailed in [31,32].

⁶ This data collection has already been utilized by Ferreira and Marques [41]. Further variables (including environmental ones) and some model specifications have also been discussed there. Here, they are adapted and extended to the present case study.

The main output considered is the Inpatient Discharges (InpD) which is the total number of patients treated in any service from the internment department, within a year, adjusted by mortality rates. Since it is an internment department study, no other outputs have been used, such as outpatients or number of emergencies, although all 49 hospitals provide these services.

Since an economic outlook is desirable (given, at least, the theoretical financial unsustainability of the health system), most of the input variables employed here are expenses, updated to 2009 (by using the consumer price index):

- a) *Costs of goods sold and consumed* (CGSC), representing the drugs and clinical materials expenditures;
- b) *Supplies and external services* (SES), which include external labor outsourcing;
- c) *Staff Costs* (StaffC);
- d) *Other costs* (OthC), including depreciation and indirect costs related to the other services in the same hospital; and, finally
- e) *Hospital Days* (HpD), which correspond to the total number of days used by all inpatients within 1 year. The latter output may be considered as an extra input since it is necessary to produce InpD, along with the other inputs.

4.2.3 Adjusting for environment

To take into account the environmental effect, the conditional formulations are employed here. The Gaussian kernel and the bandwidth given by Eq. (4.1.1) are used in all approaches. The environment factors adopted were:

- a) *Population density*—PD—(inhabitants per km²), since in urban regions the secondary health care access is facilitated compared to rural areas, where the populations have typically lower levels of education and, therefore, a priori lower wealth levels; in this sample, $PD \in [14.80; 2811.30]$ and the coefficient of variation (CV) is 132.03 %, i.e., there is a significant heterogeneity in the PD variable;
- b) *Wealth index*—WI—(purchasing power, adjusted for inflation); $WI \in [63.82; 150.63]$, $CV = 19.20\%$;
- c) *Aging index*—AI—(number of older people, >65, per 100 youth, 0–14 years old) is included since expenses with health usually increase with aging, as predicted by the Grossman model, because diseases are generally more usual and complex; $AI \in [78.60; 228.80]$, $CV = 28.83\%$;
- d) *Morbidity*—M—measured by the total of patients discharged from hospitals by 1,000 inhabitants in the most frequent diseases groups in International Classification of Diseases (ICD-9), i.e., groups II, VII, VIII, IX, XI and XVII; $M \in [8.80; 982.71]$, $CV = 141.37\%$;
- e) *Corporatization status*—CS—categorical variable (SPA, SA, EPE)—until 2002, all Public Portuguese hospitals

- had belonged to the Administrative Public Sector (SPA), subjected to public/administrative law, but at the end of 2002 the legal status of some of them was changed: they were transformed into hospital enterprises (SA) with limited liabilities and became subjected to the commercial/private law [36,39,40]; however, all SA hospitals were transformed into corporate public entities (EPE) in 2005—in practical terms the difference between SA and EPE is that SA hospital are easier to be privatized (this change occurred only for political reasons when the new government come into force) [3,40]. This sample is composed of 216 SPA, 40 SA and 136 EPE hospitals, heterogeneously distributed among the 2002–2009 period; EPE hospitals tendentially treat populations with higher WI and AI (although these ones are negatively correlated); apart from that, there is no significant relationship between the CS and the other epidemiological variables (PD and W);
- f) *Merging status*—MS—categorical variable (SH, HC, LHU)—for efficiency gains, some singular hospitals (SH) have been merged with other SHs, creating hospital centers (HC), or with other SH and few primary health units, creating local health units (LHU) [3]; this sample (of 392 units) is composed by 269 SH, 106 HC and 17 LHU (all HC and LHU are EPE units), heterogeneously distributed among the 2002–2009 period;
 - g) *Year*—Y—categorical variable (2002–2009)—a pooled frontier is employed (see *infra*), gathering the information of those 8 years into a single meta-frontier; this variable only has significantly non-zero correlation coefficients with CS and MS because these reforms have gradually occurred in time—however, its lack of correlation with the epidemiological variables means that they do not significantly change with time.

The inclusion of the CS and the MS as exogenous variables allows filtering the effect of the political and structural reforms on hospital efficiency.⁷ These have been the most important ones in the Portuguese public health system in the past few years. Therefore, any conclusion drawn from environment-

⁷ In Portugal, the policy makers, in particular, the Ministry of Health, have several roles, but none of them include the specification of the patients mix. However, the hospitals may theoretically have a perverse choice of the patients, but this is neither allowed nor clear (indeed, there is no evidence of such situations). Policy makers roles include: a) the regulation of public and private services through the laws (note that only public hospitals have been analyzed in this paper due to lack of information concerning the private ones); b) the financing of public services through taxing (and sometimes some private medical procedures due to the public services congestion); and c) the services delivery, because doctors, nurses and others, working in public hospitals, are public servants. One may argue that some important structural hospital reforms (such as the corporatization and the merging of health units) may somehow affect the inpatients complexity measurement and influence efficiency assessment. However, the effects of these reforms on efficiency are corrected by using conditional measures.

corrected results does not depend on those important reforms. Note also that all the exogenous variables are, in general, good proxies of the services provided by the hospitals (e.g. large hospitals are usually HC providing services to richer populations with high PD and low AI), thus the comparison imposed by the method does not generate injustices among hospitals providing (resp. treating) different ranges of services (resp. inpatients).

It should be noted that the sample is quite homogenous, such as in Grosskopf and Valdmanis [14], because most of units lie near CMI~1 and SMI~1, while the exogenous variables present high heterogeneity levels (their coefficient of variation may reach 141 %). However, the CMI and the SMI present statistically significant Pearson's coefficients of correlation with the population density, the wealth index (being positively correlated with these ones), the aging index and the morbidity.⁸ There is no evidence of correlation between the CMI and the year or the corporatization status.

4.2.4 Models specification

The models adopted assume an input minimization emphasis, since (1) the Portuguese public hospitals (non-for-profit) should satisfy all the patients' needs at the least possible cost, with a standard quality of service [3], and (2) one expects that it is easier to manipulate resources rather than production. A DMU is technically efficient under that standpoint if it is not possible to reduce its inputs without worsening its outputs [45,46].

In this paper, the locally convex order- m method is employed using $m=49$ and $B=1,500$. Conditional (or environmental-corrected) models assume $Z = \{PD, WI, AI, M, CS, MS, Y\}$, while unconditional models only adopt $Z = \{Y\}$, due to the aforementioned reasons. Note that the convexity seems to be more appropriate than the non-convexity assumption since it allows the feasibility of some linear combinations between two or more observations [33], which is particularly important with cost-related inputs. Furthermore, the non-convexity assumption tends to overestimate the efficiency estimation [45] and to bias the efficiency spread analysis (see *infra*). Note that the global convexity assumption does not correctly handle conditional measures.

A pooled frontier (or meta-frontier) is employed, which projects units from different time moments in the same frontier—however, since the variable *year* is included as an exogenous variable and imposing $m=49$ (the size of the sample per year) and $h = \left[1.06 \cdot \min\{\sigma, R_{23}/1.34\} \cdot n^{-\frac{1}{5}}\right]/3$, it becomes clear that only units from the same year are chosen in the first order- m step, with a probability of 99.8 %. The time frontier shift effect is then prevented. This procedure, with above-mentioned restrictions, allows the increase of the sample size

(392 units=49×8) and then, the scale-up of the results discrimination, the decrease of effect of the curse of dimensionality as well as the reduction of the number of exercises (1 instead of 8).

However, there may be some evidence of the curse of dimensionality due to the choice of $m=49$.

In order to avoid that effect, to strengthen the results, and since there is no clear political or managerial reason to impose virtual weight restrictions, the following approach seems to be the most appropriate one [32]: the inputs (or even outputs) are merged, using the eigenvector components (corresponding to the largest eigenvalue of the matrix $X^T X$) as the weights. In this paper, for all approaches, $X_g = \alpha_1 \cdot CGSC + \alpha_2 \cdot SES + \alpha_3 \cdot StaffC + \alpha_4 \cdot OthC + \alpha_5 \cdot HpD$ with $\alpha = [0.1956, 0.0478, 0.6399, 0.7416, 0.0034]^T$. X_g has a significant correlation ($p < 5\%$) with each input indicating how well this one-dimensional variable represents the original inputs. Another examination that may be employed is to add up all capital-related inputs into a single variable (total costs), i.e., imposing $X_T = \sum_{k=1}^4 X_k = CGSC + SES + StaffC + OthC$.

For each approach, at least three different models are proposed to test the results robustness. Table 1 summarizes the variables considered in each model and approach. Note that this study analyses 60 different models (=15 conditional parametric+15 conditional non-parametric+15 unconditional parametric+15 unconditional non-parametric) in order to strengthen the conclusions. The InpD variable is always weighted by mortality rates, which was the best outcome that could be found in databases. Further sub-sections describe some details that are specific of each approach.

5 Empirical results of efficiency assessment—the case of Portuguese hospitals

5.1 First approach

5.1.1 Modeling issues

As suggested by Simões and Marques [46], this approach tries to regress the ratios $Q_1 = \theta(x, y|z \cup \{CMI\})/\theta(x, y|z)$ and $Q_2 = \theta(x, y|z \cup \{SMI\})/\theta(x, y|z)$ against the CMI and the SMI, respectively, to test whether these indexes are associated or not with high hospital performance. Simões and Marques [46] use the double-bootstrap approach to test the influence of the CMI in the Portuguese hospital performance and conclude that it negatively affects hospital performance. To compute $\theta(x, y|z \cup \{CMI\})$ one only has to include the CMI as an exogenous variable, discarding it when $\theta(x, y|z)$ is calculated.

These regressions shall be done by using the non-parametric Nadaraya-Watson method. Under an input-oriented framework, an exogenous variable Z is detrimental

⁸ The SMI is positively (negatively) correlated with the morbidity (aging index), while the CMI is negatively correlated with both.

Table 1 List of variables considered in each model and approach

Model Variables	First approach			Second approach									Third approach		
	1a	1b	1c	2a.1	2a.2	2a.3	2b.1	2b.2	2b.3	2c.1	2c.2	2c.3	3a	3b	3c
Inputs	CGSC; SES; StaffC; OthC; HpD	✓		✓	✓	✓							✓		
	X_g		✓				✓	✓	✓					✓	
	X_{7i} ; HpD			✓						✓	✓	✓			✓
	SMI												✓	✓	✓
Outputs	CMI-weighted InpD			✓			✓			✓					
	SMI-weighted InpD				✓			✓			✓				
	Non-weighted InpD	✓	✓	✓		✓			✓			✓	✓	✓	✓
	CMI												✓	✓	✓

(conductive) to efficiency if the smoothed nonparametric regression is increasing (decreasing) [32]. Clearly, if $Q_j=1$, no effect of Z on efficiency should be expected. In a statistical sense, one may argue that if the 95 %-confidence intervals of $Q_1=\theta(x,y|z \cup \{CMI\})/\theta(x,y|z)$, contain the unit, then there is no reason to believe that the CMI affects hospital productivity.

5.1.2 Main results

Several important results may be drawn from the analysis described above. First, there is no meaningful influence of the CMI on efficiency if the epidemiological pattern and important political/ structural health reforms variables are taken into account by the models and under the non-parametric framework, since the 95 % Nadaraya-Watson confidence intervals (NWCI) of Q_1 include the unit over all values of the CMI. Thus, according to these specifications, it is not expected that the CMI affects the hospital performance. Under the parametric framework, the CMI seems to be detrimental (conductive) to efficiency for low (high) values, i.e., when the hospital treats patients with lower complexities, on average, it is expected that this lower complexity is harmful to efficiency due to the existence of outsourced assets and human capital. All these results are consistent among models 1a, 1b and 1c. Figure 2 shows the non-parametric regression (against the CMI) for the model 1b, where the efficiencies have been environment-corrected and computed under the non-parametric (on top) and the parametric (on bottom) frameworks.

Second, if the model disregards the environment, CMI seems to affect efficiency. Under the parametric basis, CMI only affects hospital efficiency for low values (being favorable), i.e., $CMI < 1$; otherwise, the 95 % NWCI's contain the unit and there is no significant influence on hospital efficiency. However, if the model is non-parametric, then there is no consistency among the models. While model 1b points out that CMI is unfavorable to the production process, according

to model 1a the CMI is only unfavorable (favorable) for low (high) CMI values, i.e., $CMI < 1$ ($CMI > 1$). Meanwhile, model 1c indicates that CMI is unfavorable until $CMI=1$ and after $CMI=1.5$, being favorable to the efficiency between these two values.

Third, when Q_2 is regressed against the SMI, the results show some consistency among the models. In the case of conditional measures, (1) parametric models show that Q_2 is increasing with the SMI, thus this variable is adverse to hospital efficiency, while (2) non-parametric models indicate that it seems to have no influence on efficiency. On the other hand, in the case of non-environment-corrected measures, (1) the SMI seems to be detrimental for $SMI > 1.5$ when the model is parametric, but (2) it appears to have no influence on efficiency when the model is non-parametric. Note that here there is a higher consistency between conditional and non-conditional backgrounds.

5.2 Second approach

5.2.1 Modeling issues

This approach aims to examine the influence of both indexes on hospital efficiency assuming that they are InpD weights. To do so, the most appropriate procedure is to compare those different frontiers (obtained for CMI, SMI and NW schemes).

With the purpose of comparing two different frontiers, two different metrics are suggested, as follows [47]:

$$\Theta_{AB} = \sqrt[n]{\prod_{i=1}^n \frac{\delta^A(x_i^A, y_i^A)}{\delta^B(x_i^B, y_i^B)}} \tag{5.2.1}$$

$$\Phi_{AB} = \sqrt[2n]{\prod_{i=1}^n \frac{\delta^B(x_i^A, y_i^A) \cdot \delta^B(x_i^B, y_i^B)}{\delta^A(x_i^B, y_i^B) \cdot \delta^A(x_i^A, y_i^A)}}$$

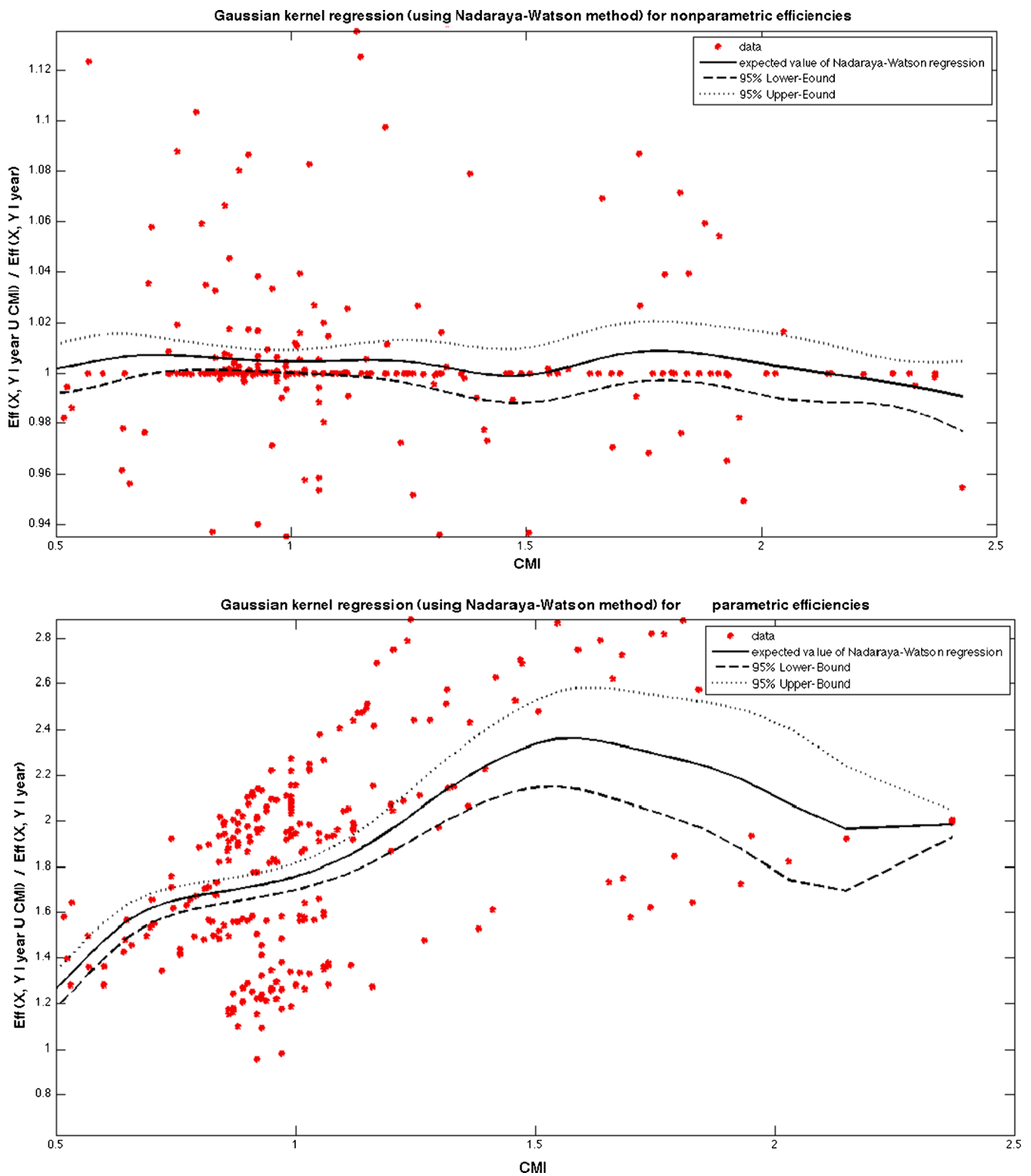


Fig. 2 Non-parametric Nadaraya-Watson regression (against the CMI) using the Gaussian kernel, for environment-corrected efficiency estimates (Model 1b): non-parametric (*top*) and parametric (*bottom*)

In previous eq., (1) $\delta^A(x_i^A, y_i^A)$ represents the Shephard's input distance of DMU characterized by (x_i^A, y_i^A) against the frontier A (usually, a cluster

frontier); (2) $\delta^B(x_i^A, y_i^A)$ represents the Shephard's input distance of DMU characterized by (x_i^A, y_i^A) against the frontier B; (3) Θ_{AB} denotes the efficiency spread between

clusters A and B—if $\Theta_{AB} < 1$, then DMUs belonging to cluster A have greater consistency in efficiency than DMUs from cluster B; and (4) Φ_{AB} designates the productivity gap between both clusters' frontiers—if $\Phi_{AB} < 1$, then cluster A frontier has, on average, higher levels of productivity than group B frontier. In this work, there is only one cluster, but three different frontiers. Using previous concepts in a straightforward way, one may perform their comparison. Clearly, if CMI and/or SMI have no effect on efficiency (productivity), then Θ_{AB} (resp. Φ_{AB}) is unitary. In a statistical viewpoint, it may be measured by the confidence intervals of Θ_{AB} and Φ_{AB} : if they contain the unit, then no meaningful differences are observable between those different models.

If any difference is detected by non-parametric methods, then it is totally attributed to inefficiency (which may be not true in practice), but one may argue that the detected inefficiency is the same whether the InpD are weighted by the CMI, by the SMI or none—that is, the analysis proposed allows disentangling the inefficiency from the index effect on performance if one states that the “true” efficiency is that one obtained by using the CMI as the InpD weighting scheme. In such a case, one may argue that we are under a *separability* condition case and possible deviations may be due to differences between the CMI and the SMI/ non-weighting schemes. No differences mean that providers follow evidence-based principles to guide patient care (standardized care) and act in the interest of their patients.

5.2.2 Main results

According to the results of this approach,⁹ if the model has been corrected by epidemiological pattern and important political/ structural health reforms variables, then there is no noteworthy efficiency spread (Θ_{AB}) among those different InpD weighting schemes. That is, efficiencies do not change by using the CMI or the SMI or none of them. It is noticeable that (1) both parametric and non-parametric efficiencies lead to quite similar results, being the translog function a good choice, and (2) there is no remarkable presence of the curse of dimensionality since all models, 2a, 2b and 2c, yield alike values of Θ_{AB} . Notwithstanding, unconditional models generate statistically significant deviations of Θ_{AB} from the unit, except when the SMI model is compared with the NW model. Actually, the

employment of the CMI seems to generate efficiencies with higher consistency than the ones obtained in the other models. Once again, there is a high consistency among parametric and non-parametric measures.

Even though conditional measures have shown that there is no efficiency spread among those different metrics, i.e., the relative position of the units against their own frontier does not change whether the InpD are risk adjusted or not. The same does not happen in the case of the relative positions of the different frontiers (one for each weighting scheme) where conditional and non-conditional measures show similar trends, as opposed to the parametric *vs* non-parametric dichotomy. The CMI seems to strongly decrease (increase) the productivity of the units when compared to the NW and the SMI models in non-parametric (parametric) frameworks. This effect is strengthened in non-conditional backgrounds. One should remark that in conditional parametric models, the CMI and the SMI's frontiers tend to overlap, which means that under these conditions, it is not relevant to choose between one index or the other.

5.3 Third approach

5.3.1 Modeling issues

In this approach, it is assumed that any possible differences between the case mix and service intensity indices are solely due to inefficiency, and there is no explicit hospital-wide intention to game the CMI or the SMI for financial (reimbursement-related) purposes. In such cases, if the CMI is truly meant to measure illness severity, it should be an extra non-discretionary output in the model, according to Grosskopf and Valdmanis [14] and Garavaglia et al. [48], because it is based on patient conditions and other factors external to the hospital. On the other hand, the severity of services provided should be a discretionary input [29], given that the hospital determines when and sometimes how to treat certain conditions when evidence based protocols are vague. Since the models are input-oriented, the discretionary status of those extra variables does not change the model—all inputs are discretionary and the outputs non-discretionary, by default.

However, by using the bootstrap it is possible to test whether the inclusion of the CMI and the SMI as variables is meaningful to the DMU efficiency analysis. As a matter of fact, the inclusion of a variable keeps or increases the ESs, but never the opposite. Let $\theta_{CMI\&SMI}$ be the ES using the CMI and the SMI as previously defined; let also be θ_{CMI} , θ_{SMI} and θ the ESs obtained only with the CMI but without the SMI, only with the

⁹ In order to save space, the results can be provided by the authors if requested.

SMI but without the CMI, and without any of these variables, respectively. Thus, the succeeding five

hypotheses sets are imposed (the p -value is computed through Eq. (5.3.1)):

$$\begin{aligned} H_{0(1)} : \theta_{CMI\&SMI} &= \theta; & H_{0(2)} : \theta_{CMI\&SMI} &= \theta_{CMI}; & H_{0(3)} : \theta_{CMI\&SMI} &= \theta_{SMI}; & H_{0(4)} : \theta_{CMI} &= \theta; \\ H_{1(1)} : \theta_{CMI\&SMI} &> \theta; & H_{1(2)} : \theta_{CMI\&SMI} &> \theta_{CMI}; & H_{1(3)} : \theta_{CMI\&SMI} &> \theta_{SMI}; & H_{1(4)} : \theta_{CMI} &> \theta; \\ H_{0(5)} : \theta_{SMI} &= \theta; & & & & & & \\ H_{1(5)} : \theta_{SMI} &> \theta. & & & & & & \end{aligned}$$

$$p = \frac{1}{B} \cdot \sum_{k=1}^B \mathbb{I} \left(T_k^{bootstrap} \leq T_k^{observed} \right) \quad (5.3.1)$$

In the previous equation, \mathbb{I} is the indicator function and it is equal to the number of true occurrences on its argument. Meanwhile, $T_{k,j}$ is a statistical test that may be computed using, for instance, the following equations¹⁰:

$$\begin{aligned} T_{k,1} &= \frac{\sum_{i=1}^n \{\theta_{Model\ 1}(x_i, y_i)\}}{\sum_{i=1}^n \{\theta_{Model\ 2}(x_i, y_i)\}} \\ T_{k,2} &= \frac{(\prod_{i=1}^n \{\theta_{Model\ 1}(x_i, y_i)\})^{\frac{1}{n}}}{(\prod_{i=1}^n \{\theta_{Model\ 2}(x_i, y_i)\})^{\frac{1}{n}}} \\ T_{k,3} &= \frac{median\{\theta_{Model\ 1}(x_i, y_i)\}}{median\{\theta_{Model\ 2}(x_i, y_i)\}} \end{aligned} \quad (4.5.2)$$

5.3.2 Main results

The results of this approach are peremptory, as shown in Table 2 (for the case of Model 3b). In terms of efficiency measurements, there is no significant differences whether those two variables are included or not in models. This is in line with the efficiency spread analysis previously performed and its results, concerning the conditional framework. One should note that there is a considerable consistency among parametric and non-parametric measures, except for few special circumstances. Although it is not clear whether the SMI (as formulated in this paper) would be an input or an output, one would expect that if it is an output (unlikely this approach states), it would return

similar results because the SMI and the CMI are highly comparable and present quite similar values.

6 Discussion and conclusions

The results attained point out towards one simple but important conclusion: both the CMI and the SMI seem to be pointless in the model if and only if (1) it takes into account epidemiological pattern and other environmental variables,¹¹ and (2) one does not perform any cluster comparison analysis (e.g. by using the Malmquist index) because the CMI tends to considerably change the frontier, but not the relative position of units against that. This means that a time evolution analysis (or others somehow related) should account for the inpatients complexity, either through the CMI, the SMI or the environmental first-step adjustment, instead, since the latter group of units is truly comparable. Therefore, the inpatients adjustment has an influence on hospitals productivity.¹² Otherwise, the model would certainly be biased/ mis-specified and the results would be wrongly obtained/ addressed. Thus, the kind of analysis and the chosen model (parametric/ non-parametric; conditional/ non-conditional; weighting schemes/ extra variables) really matter because they tend to significantly influence the results. Nonetheless, the results suggest that the CMI does not actually reflect the hospital costs. One strong limitation of this study is that CMI data is quite homogenous, which might be the result of considerable epidemiology homogeneity within Portugal (because it is a small country), and the reader should be aware of that. Even so, this may introduce important changes in efficiency assessment if not corrected by the environment. Nevertheless, hospitals may become

¹⁰ It should be noted that $\theta_{Model\ 1} \leq \theta_{Model\ 2}$. For example, regarding $H_{0(1)}$, one has $\theta_{Model\ 1} = \theta$, while $\theta_{Model\ 2} = \theta_{CMI\&SMI}$. Otherwise, the p -value should be adapted to $p = \frac{2}{B} \cdot \min \left\{ \sum_{k=1}^B \mathbb{I} \left(T_k^{bootstrap} \leq T_k^{observed} \right), \sum_{k=1}^B \mathbb{I} \left(T_k^{bootstrap} \geq T_k^{observed} \right) \right\}$, which is particularly important when the efficiency is parametrically computed. In such a case, it is not evident whether the inclusion of a new variable increases or not the efficiency.

¹¹ Note that non-environment-corrected models are usually biased and do not correctly estimate the efficiency. Apparently, the environment correction also adjusts the inpatients (and perhaps outpatients and others).

¹² Productivity and efficiency are not necessarily the same thing. Usually, the productivity refers to two different frontiers [45].

Table 2 *p*-values for those five different hypotheses sets, three different statistical tests, T_k , using bootstrap (Model 3b)

		Model 3b			
		Conditional		Unconditional	
		Non-parametric	Parametric	Non-parametric	Parametric
$H_{0(1)}$ vs $H_{1(1)}$	T_1	0.486	0.940	0.530	0.976
	T_2	0.764	0.952	0.706	0.980
	T_3	<i>0.000</i>	0.928	0.986	0.896
$H_{0(2)}$ vs $H_{1(2)}$	T_1	0.480	0.964	0.494	1.000
	T_2	0.650	0.996	0.756	0.992
	T_3	<i>0.000</i>	0.572	<i>0.000</i>	0.880
$H_{0(3)}$ vs $H_{1(3)}$	T_1	0.498	0.992	0.484	0.948
	T_2	0.552	0.996	0.082	0.956
	T_3	<i>0.000</i>	0.896	0.352	0.748
$H_{0(4)}$ vs $H_{1(4)}$	T_1	0.502	0.988	0.510	1.000
	T_2	0.564	0.976	0.484	0.976
	T_3	<i>0.000</i>	0.696	0.986	0.928
$H_{0(1)}$ vs $H_{1(5)}$	T_1	0.490	0.976	0.552	0.996
	T_2	0.746	0.980	0.974	0.956
	T_3	<i>0.000</i>	0.916	0.986	0.920

Note: *p*-values below 0.05 are displayed as italic entries; the null hypothesis is rejected if and only if, at least, two of three statistical tests present $p < 0.05$

efficient or increase their efficiency without necessarily increasing the average inpatients illness by choosing those ones with higher diseases complexities, which is in line with [12–15]. This is an important, debatable but stimulating result. It is debatable because the results seem to be model- and hospital realities-dependent, although a lot of authors have obtained similar conclusions, as previously stated.

One should also note that the high correlation among CMI and SMI as well as the similar efficiency results obtained through these different adjustment schemes may not occur in other countries, where the CMI may present a broader spectrum of values and/or if there are services where the inpatients have a larger broad of complexities than in the Portuguese case—indeed, the service notion is not standardized at all across countries as a common unit of hospital organization. Clearly, the CMI and the SMI are only comparable if and only if the last one is computed with the same reference than the CMI, in this case, the hospital admission. A statistical test is then required to test whether the influence of one index on efficiency is higher/ smaller than the other.

At the level of the facility, there is a tremendous amount of patient aggregation. Hence, many differences in clinical practice (which are based on patients who present with very different conditions, each of which was managed differently by the patient prior to inpatient care) would average out. If the hospitals that are being compared are truly comparable, and if the efficient frontier is solely based upon comparable hospitals, then CMI/SMI should not matter, regardless of how they are

measured. In essence, at an aggregate level, comparable hospitals treat comparable patients. So if one identifies the “right” set of hospitals against which to create the efficient frontier (the choice of environmental variables) these differences effectively mitigate any need to control for illness severity and/or service mix, at the level of the facility as the DMU. Two main implications arise from this logic.

Firstly, it does not necessarily imply that such adjustments are “meaningless”. Rather, the use of some CMI/SMI adjustment forms the basis for a test of model misspecification, especially at the level of the facility. CMI/SMI data should be collected and used as a robustness check. If the results are dependent on the inclusion of such variables, the researcher, in all likelihood, has made some other error in his analysis. Secondly, the choice of Z is *per se* a hot topic and has several implications on results. A wrong/ inappropriate/ incomplete choice of Z (and, possibly the bandwidth and the kernel functions, as well) generally biases the final results and, therefore, the arising conclusions [40]. Moreover, the frequently employed two-step methodology (e.g. by using a truncated Tobit model or a double bootstrap method, as proposed by Simar and Wilson [49]), without an a priori units environmental adjustment, inherently assumes a separability condition on the production process, which is not necessarily true [32,33,40]. In such a case, the results in the first step are biased because units are not compared with truly comparable units [50]. Future analysis should take into account other exogenous variables, other kernel functions and other bandwidths for a robustness check.

Although the results point out that those indexes seem to be needless to assess hospital efficiency, the SMI may replace the CMI as InpD weight, due to theoretical reasons, or as an extra variable to describe the hospital resources. This index is easy to compute, it is comparable to the CMI, which is not usually available, and may be computed taking any baseline (a service, a department, a hospital or a country). Thus, the SMI, as proposed in this paper, may be useful for international comparisons. Nevertheless, we should be cautious in concluding that CMI and SMI indexes measure the same thing. While the SMI is a cost-based measure, since it is based on expenses associated with specific types of patient care, the CMI, being a DRG-based measure, is revenue based measure, because prospective payment system reimbursement models are based on the DRG units. There is certainly quite a bit of overlap, especially for non-profit firms (as happens in the present paper), who essentially equate revenues and accounting expenses. However, if the researcher is using variables, DEA orientations, or extensions of VRS technical efficiency (revenue efficiency, cost efficiency, etc.), it may be more appropriate to use one over the other. One would use the most appropriate measure as the basis for the robustness test. Hence, these results should be compared with other obtained through suchlike methods, but using other realities (countries with different health systems, for-profit units, etc). For instance, the sample used by Grosskopf and Valdmanis [14] has a quite different nature compared to ours (including the significant time shift between both samples), thus they are not directly comparable. Notwithstanding, by using similar methods (less robust however), similar conclusions are drawn.

This paper is important for a) scholars/ practitioners, because it is a hot topic on hospital efficiency measurement and contributes with several important questions and answers, and for b) policy-makers and hospital managers, since it has been concluded that, under the environment influence, the effect of inpatients severity/complexity (assumed to being quite related to service complexity) seems to be meaningless for efficiency assessment, which has important implications on hospital management and new political reforms. E.g. this contributes to the decrease of the perverse inpatient choice probability by the hospital as a function of the case mix. That is, efficiency is not a function of the output complexity and heterogeneity, but rather of the inputs consumption. Therefore, the “case mix game” (and specially the so-called “DRG creep”, where the lower severity inpatients are coded as sicker) has no sense if hospitals want to increase their own efficiency. Furthermore, this conclusion has an important influence on hospitals reimbursement by the Ministry of Health,

since the increase of the case mix (or eventually the service mix) does not necessarily translate itself into technical efficiency improvements, in line with this paper results. One may also suggest that the prospective payment should be done through a function of the environmental conditions, instead of the CMI.

One should strongly recognize that this analysis takes place at the level of the facility (non-for-profit hospitals), and more specifically over inpatient services. If efficiency analyses are completed at a much more local level, say, individual operating units (intensive care, medical/surgical, physical medicine, etc.), it may be not necessary to control for patient heterogeneity, as well as organizational differences (for example, separate intensive care and medical/surgical units vs combined units) across hospitals. The more specific and narrowly defined are the operating units in the analysis, the less likely are differences in patient heterogeneity (no matter how they are adjusted: CMI or SMI) to average out. It would be incumbent upon future research to replicate this study in these more “local” areas of production to determine whether or not CMI/SMI adjustments are necessary.

Also, it must be acknowledged that this is an introductory study of efficiency in the Portuguese health sector and the influence of the severity adjustment of inpatients. All techniques and models here presents should be viewed as complementary. Then, the models employed should be target of a further research. Apart from the previous model specifications, new models should consider, for instance, (1) the HpD as an output (as frequently pointed out in the literature), (2) new inputs (as doctors, nurses and beds), (3) other SMI formulations (see *supra*), (4) other CMI formulations (by changing the DRG grouping schemes by the ICD ones, e.g. [21]), or (5) other model orientations (such as directional models). These points should be included for a robustness analysis. Furthermore, one may test the local convexity and/or the efficiency estimates adjustment by other parametric functions. Finally, the models should include outcome (quality) measures [48], because there might be some efficiency/quality trade-offs that are not accounted for in this paper models and may change due to the presence of such outcomes information. It would have some political, economic and managerial implications on hospitals because the tendentious “case mix game” may affect the health care quality. Because the outputs volume may affect the quality (in general, it puts down the outcomes), the information concerning the illness severity may explain the outcomes, like the mortality rate or the increase in the length of stay, being a powerful indicator for the hospital management and higher entities, such as the Ministry of Health.

Acknowledgments We would like to thank to two anonymous referees who kindly and significantly have improved the paper's quality, clarity and structure, due to their beneficial comments. We would like also to thank to Prof. Yasar Ozcan for his valuable comments, specially regarding the new SMI development. Any errors are our own responsibility.

References

- Ozcan YA (2008) Health care benchmarking and performance evaluation: an assessment using data envelopment analysis (DEA). Springer, New York
- Jacobs R, Smith P, Street A (2006) Measuring efficiency in health care: analytic techniques and health policy. Cambridge University Press, Cambridge
- Marques RC, Carvalho P (2013) Estimating the efficiency of Portuguese hospitals using an appropriate production technology. *Int Trans Oper Res* 20:233–249
- Tatchell M (1983) Measuring hospital output: a review of the service-mix and case-mix approaches. *Soc Sci Med* 17(13):871–883
- Grosskopf S, Valdmanis V (1987) Measuring hospital performance: a nonparametric approach. *J Health Econ* 6:89
- Burgess JF, Wilson PW (1995) Decomposing hospital productivity changes 1985–1988: a nonparametric Malmquist approach. *J Prod Anal* 6(4):343–363
- Dalmau-Atarrodona E, Puig-Junoy J (1998) Market structure and hospital efficiency: evaluating potential effects of deregulation in a National Health Service. *Rev Ind Organ* 13(4):447–466
- Sahin I, Ozcan YA (2000) Public sector hospital efficiency for provincial markets in Turkey. *J Med Syst* 24(6):307–320
- Wan TTH (2000) The impact of the prospective payment system on technical efficiency of hospitals. *J Med Syst* 24(3):159–172
- Hofmarcher MM, Paterson I, Riedel M (2002) Measuring hospital efficiency in Austria—a DEA approach. *Health Care Manag Sci* 5:7–14
- Peacock S, Chan C, Mangolini M, Johansen D (2001) Techniques for measuring efficiency in health services: Staff working paper. Productivity Commission
- Chilingerian JA (1995) Evaluating physician efficiency in hospitals: a multivariate analysis of best practices. *Eur J Oper Res* 80(3):548–574
- Rosko MD, Chilingerian JA (1999) Estimating hospital inefficiency: does case mix matter? *J Med Syst* 23(1):57–71
- Grosskopf S, Valdmanis V (1993) Evaluating hospital performance with case-mix adjusted outputs. *Med Care* 31:525–532
- Vitikainen K, Street A, Linna M (2009) Estimation of hospital efficiency—do different definitions and case mix measures for hospital output affect the results? *Health Policy* 89(2):149–159
- Claudio P (2014) Severity of illness in the case-mix specification and performance: a study for Italian public hospitals. *J Hosp Adm* 3(1):23–33
- Becker ER, Steinwald B (1981) Determinants of hospital case mix complexity. *Health Serv Res* 16:4
- Rosen AK, Loveland SA, Rakovski CC, Christiansen CL, Berlowitz DR (2003) Do different case-mix measures affect assessments of provider efficiency? Lessons from the Department of Veterans Affairs. *J Ambul Care Manag* 26:229–242
- Magnussen J, Nyland K (2008) Measuring efficiency in clinical departments. *Health Policy* 87(1):1–7
- Dismuke CE, Sena V (1999) Has DRG payment influenced the technical efficiency and productivity of diagnostic technologies in Portuguese public hospitals? An empirical analysis using parametric and non-parametric methods. *Health Care Manag Sci* 2(2):107–116
- Yang CM, Reinke W (2006) Feasibility and validity of international classification of diseases based case mix indices. *BMC Health Serv Res* 6:125
- Iezzoni LI (1997) The risks of risk adjustment. *J Am Med Assoc* 278(19):1600–1607
- Iezzoni LI (2009) Risk adjustment for performance measurement. In: Smith PC, Mossialos E, Papanicolas I, Leatherman S (eds) *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge University Press, Cambridge, pp 251–285
- Chilingerian JA, Sherman HD (2011) Health care applications: from hospitals to physicians. From productive efficiency to quality frontiers. In: Cooper WW, Seiford LM, Zhu J (eds) *Handbook on data envelopment analysis*, 2nd edn. Kluwer, Norwell, pp 445–494
- Lilford R, Mohammed MA, Spiegelhalter D, Thomson R (2004) Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 363(9415):1147–1154
- Ozcan YA, Luke RD (1993) A national study of the efficiency of hospitals in urban markets. *Health Serv Res* 28(6):719–739
- Ozcan YA (1993) Sensitivity analysis of hospital efficiency under alternative output/input and peer groups: a review. *Knowl Policy* 5(4):1–29
- Lobo MSC, Ozcan YA, Lins MPE, Silva ACM, Fiszman R (2014) Teaching hospitals in Brazil: findings on determinants for efficiency. *Int J Healthcare Manag* 7(1):60–68
- Ozcan YA, Lins ME, Lobo MSC, Silva ACM, Fiszman R, Pereira BB (2010) Evaluating the performance of Brazilian university hospitals. *Ann Oper Res* 178:247–261
- Chou T-H, Ozcan YA, White KR (2012) Technical and scale efficiencies of catholic hospitals: does a system value of stewardship matter? In: Tãnfani E, Testi A (eds) *Advanced decision making methods applied to health care*. International Series in Operations Research & Management Science. Springer Science Publisher, pp 83–101
- Florens JP, Simar L (2005) Parametric approximations of nonparametric frontier. *J Econ* 124:91–116
- Daraio C, Simar L (2007) Advanced robust and nonparametric methods in efficiency analysis: methodology and applications. Springer Science + Business Media, New York
- Daraio C, Simar L (2007) Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. *J Prod Anal* 28(1–2):13–32
- Silverman B (1986) *Density estimation for statistics and data analysis*. Chapman & Hall, London
- Hollingsworth B (2003) Non-parametric and parametric applications measuring efficiency in health care. *Health Care Manag Sci* 6(4):203–218
- Rego G, Nunes R, Costa J (2010) The challenge of corporatisation: the experience of Portuguese public hospitals. *Eur J Health Econ* 11:367–381
- O'Neill L (1998) Multifactor efficiency in data envelopment analysis with an application to urban hospitals. *Health Care Manag Sci* 1(1):19–27
- Hollingsworth B, Dawson PJ, Manidakis N (1999) Efficiency measurement of health care: a review of non-parametric methods and applications. *Health Care Manag Sci* 2(3):161–172
- Bjorkgren MA, Hakkinen U, Linna M (2001) Measuring efficiency of long-term care units in Finland. *Health Care Manag Sci* 4(3):193–200
- Biorn E, Hagen TP, Iversen T, Magnussen J (2003) The effect of activity-based financing on hospital efficiency: a panel data analysis of DEA efficiency scores 1992–2000. *Health Care Manag Sci* 6(4):271–284

41. Ferreira D, Marques RC (2014) Did the corporatization of Portuguese hospitals significantly change their productivity? *Eur J Health Econ*. doi:10.1007/s10198-014-0574-8
42. Grosskopf S, Hayes K, Taylor L, Weber W (1997) Budget constrained frontier measures of fiscal equity and efficiency in schooling. *Rev Econ Stat* 79:459–474
43. Coelli T (2000) On the econometric estimation of the distance function representation of a production technology. Working Paper
44. Afonso A, Fernandes S (2008) Assessing hospital efficiency: Non-parametric evidence for Portugal. Working Paper
45. Witte K, Marques RC (2011) Big and beautiful? On non-parametrically measuring scale economies in non-convex technologies. *J Prod Anal* 35:213–226
46. Simões P, Marques RC (2009) Performance and congestion analysis of the Portuguese hospital services. *Cent Eur J Oper Res* 19(1):39–63
47. Camanho AS, Dyson RG (2006) Data envelopment analysis and Malmquist indices for measuring group performance. *J Prod Anal* 26(1):35–49
48. Garavaglia G, Lettieri E, Agasisti T, Lopez S (2011) Efficiency and quality of care in nursing homes: an Italian case study. *Health Care Manag Sci* 14(1):22–35
49. Simar L, Wilson PW (2007) Estimation and inference in two-stage semi-parametric models of production processes. *J Econ* 136:31–64
50. Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS, Shale EA (2001) Pitfalls and protocols in DEA. *Eur J Oper Res* 132(2):245–259