Evaluating hospital performance with plant capacity utilization and machine learning

Abstract: This study extends the measurement of plant capacity utilization by incorporating undesirable outputs. We select indicators through feature selection in machine learning and also introduce an undesirable output for assessment in these models. By defining and applying four plant capacity concepts, we analyze plant capacity utilization in health institutions in 31 provinces in China over the last 11 years (2009 to 2019). This paper has two main contributions. First, we propose a refined by-production hospital technology by introducing the mortality rate into the performance evaluation of public hospitals. Second, we expand the measures of plant capacity utilization with undesirable outputs. The preliminary results show that after the introduction of the death rate, the long-run output-oriented plant capacity utilization of medical institutions is significantly impacted. Furthermore, we found a high level of long-run input-oriented plant capacity utilization tends to increase mortality.

Keywords: Plant capacity utilization; undesirable output; machine learning; health care

1. Introduction

Over the past 30 years, there have been numerous empirical studies on efficiency and productivity. In one of the first, Johansen (1968, p.52) proposed the concept of plant capacity, pointing out that "without limiting the variable production factors, the largest number of existing factories and equipment can be output per unit time is plant capacity." Based on Johansen's (1968) definition, Färe (1984) established the important conditions for the existence of plant capacity and argued that the concept of plant capacity cannot be derived from fashionable parametric production function specifications. Therefore, a nonparametric frontier method was adopted by Färe et al. (1989a and 1989b), who both argued plant capacity in terms of inputs and outputs by using two output-oriented efficiency measures.

After that, no progress was made on any method of plant capacity. Although nonparametric frontier efficiency and productivity measurement have been widely used, only the output-oriented plant capacity concept is generally applied. However, recent research has introduced some breakthroughs. Cesaroni et al. (2017) made a new definition of input-oriented plant capacity utilization (PCU) measuring approach based on two input-oriented efficiency measures.

Next, according to Cesaroni et al. (2019), a type of long-run (LR) output- and input-oriented plant capacity concept had been developed in recent years. Instead of simply changing variable inputs, they assumed all input dimensions can be changed at once. They also reinterpreted the plant capacity concept by only focusing on the change of variable inputs as a short-run concept. Where the long-run output-oriented plant capacity utilization rate is calculated as a ratio of the standard production technology to the same technology without any input limitation; the long-run input-oriented plant capacity utilization rate is calculated by input efficiency measures between a given level of output and its null level. Additionally, Kerstens et al. (2019) compared experience-oriented concepts of technical and economic competence and also compared the concept of capacity between convex and non-convex technologies. Briec et al. (2020) determined a generalized, encompassing formulation regarding input efficiency measures, which includes and thus links four well-known input efficiency measures. On this basis, Kerstens and Van de Woestyne (2021) tested the influence of the convexity assumption on cost estimation and used nonparametric criteria for technical and cost functions.

Measuring efficiency and productivity change has become a popular approach for performance evaluation from sectorial to firm levels. Some studies have used this method, for example, those focusing on the electric (Jamasb & Pollitt, 2001), banking (Sherman & Ladino, 1995), fishery (Felthoven, 2002), and insurance sectors (Cummins & Weiss, 2000). Ghaderi et al. (2006) and Azad et al. (2012) used cutting-edge data analysis methods to assess hospital efficiency. Eriksson et al. (2017) reviewed the relationship between inpatient capacity strain and health outcomes, and evaluated the pros and cons of health system interventions to improve the quality of inpatient care during periods of capacity strain. Mohamadi E et al. (2020) used extended data enveloping analysis to measure the technical efficiency and productivity of 568 public hospitals, and reported the results of a national study aimed at measuring the changes in technical efficiency and productivity of Iranian public hospitals from 2012 to 2016. However, most studies ignore undesirable or bad outputs in evaluations. This might lead to biased results as some potential social losses were not taken into account to the production technology.

The novelty of this paper is to refine the measure of plant capacity utilization by incorporating the undesirable output (e.g., mortality rate) into the hospital production technology, which consists of two sub-technologies. We compare and summarize short-run and long-run PCU measurements using the classic distance functions and directional distance functions. This study has two main contributions. First, we introduce undesirable output into the public hospital evaluation. A refined by-production with linking constraint is applied to model hospital production technology. Second, we expand the short-run and long-run plant capacity utilization with undesirable outputs for evaluating hospital performance.

The empirical data we used consisted of the capacity of medical institutions in 31 provinces in China from 2009 to 2019. As a developing country, ensuring the health of all its citizens is still a major problem in China. Chinese authorities need to take full advantage of the existing hospital capacity to ensure national health. In the face of such challenges, Chinese authorities need to more effectively use and improve existing hospital capacity to adequately treat patients. Following Eriksson et al. (2017), hospital capacity is increasingly under pressure as mortality rates rise and health outcomes deteriorate globally. Therefore, in this study, we introduce the death rate as an undesirable output. Since capacity accumulation needs to be replaced in this modeling strategy, we suggest the LR plant capacity concept is especially suitable in this regard.

The structure of this study is as follows. In Section 2, we explain the plant capacity concept in the context of the medical sector according to relevant literature and clarify the connection between plant capacity and mortality. Sections 3 and 4 introduce the technology and efficiency measurements (nonparametric estimation of the directional distance function, DDF), provide clear definitions of output- and input-oriented SR and LR plant capacity concepts, and discuss the nonparametric frontier techniques used to assess the capacity concepts of these plants. Section 5 describes the 11-year data of 31 Chinese provinces. Section 6 discusses the empirical results and clarifies the relationship between the short-run (SR) and LR plant capacity concepts. Finally, Section 7 concludes the whole study.

2. Hospital plant capacity in economic literature

China is the most populous country in the world and thus faces problems related to providing health services to its large population. National health is an important symbol of a country's capacity for sustainable development, and health has become an increasingly important issue in the international community. In 2018, a key decision was made to prioritize Chinese citizens' health and implement the "Healthy China Strategy," a national policy aimed to promote people's health in which medical and health institutions are meant to play a very important role. To more clearly understand the state of medical and health institutions in China, we use the nonparametric frontier method to analyze the plant capacity of public hospitals in 31 provinces from 2009 to 2019.

The hospital efficiency, productivity, and capacity utilization have attracted much attention of researchers. Färe et al. (1989a) used this approach to analyze hospitals in Michigan and evaluate plant capacity, they adopted a nonparametric (linear programming) framework to determine capacity according to observed input and output and the best practice results of all hospitals. Later, Magnussen and Rivers Mobley (1999) compared hospitals in Norway and California and determined the relationship between capacity utilization, production efficiency, and empty bed cost. They also compared the capacity utilization of beds in four different market environments. Färe et al. (1989b) analyzed the capacity of hospitals in different environments. Leleu et al. (2006) focused on the plant capacity of public hospitals in Tunisia, using a nonparametric approach to measure services and estimate capacity utilization.

Valdmanis et al. (2010) calculated the capacity of hospitals in Florida and used DEA to measure capacity in the frontier environment, combined information about hospital capacity, and determined the characteristics of discharged patients and financial performance. Panagiotis et al. (2015) studied the efficiency of Greek public hospitals and summarized their statistical characteristics. They combined random DEA with Bayesian analysis to enhance the results' stability by improving parameter accuracy and reducing outlier impact. Burdett and Kozan (2016) examined the multi-standard method of hospital capacity analysis and designed a multi-objective method to analyze the working ability of hospitals. Arfa et al. (2017) used a dual-DEA model to measure the capacity utilization of public hospitals in Tunisia, while Kerstens et al. (2019) mainly studied plant capacity and accessibility, exploration, and remedial measures. Boussemart et al. (2020) decomposed the aggregate productivity change of Chinese public hospitals into technological progress, technical, scale and mix efficiency changes. Shen and Valdmanis (2020) analyzed the efficiency and productivity changes for Chinese public provincial hospitals. These studies ignore undesirable outputs (mortality) in the production technology or without investigating the nexus between hospital performance and death rate.

The relationship between the hospital performance and mortality rate also attracts much attention. Hollingsworth (2003) and Pelone et al. (2015), had applied advanced technology to analyze the efficiency and productivity of hospitals and other medical facilities in relation to care services and death rate. However, there is almost no conclusive evidence that a connection exists between efficiency and productivity components and care services and death rate. Kuntz et al. (2015) argued that German hospital capacity and its expansion resources allocation have an impact on the mortality. During the COVID-19 epidemic, Kerstens and Shen (2021) investigated hospital capacity utilization at the city level in Hubei province of China and they argued that the plant capacity measure was related to the mortality rate. Similarly, in the wider literature on economic and operation management, there is scant clear evidence regarding the relationship between medical operational decision-making and mortality.

Examining the economic and medical literature, we find that efficiency and productivity analysis is applied to hospitals and other healthcare facilities by using cutting-edge technologies; studies have also focused on the relationship between hospital capacity and mortality. Rosko et al. (2008) analyzed 20 studies on US hospital inefficiency using stochastic frontier analysis; they compared the results of those using methods of best practice to those using previous methods in hospital studies to determine the effectiveness of stochastic frontier analysis in estimating cost efficiency. Moreover, Lapichino et al. (2004) examined the relationship between intensive care unit (ICU) productivity and mortality by measuring ICU-specific risk adjusted productivity, they found mortality is closely related to high utilization rate at the individual-disease level. Although the capacity pressure measurement methods used in ICU and non-ICU environments differ, Eriksson et al. (2017) found that hospital capacity pressure was related to increased patient mortality in nine of 12 studies in ICU environments and in 18 of 30 studies in non-ICU environments. Therefore, based on these studies, we can infer a direct positive correlation among plant capacity utilization and the death rate.

In summary, we argue that the undesirable output, namely, mortality cannot be ignored in the evaluation for hospitals and healthcare sectors. In this study, we investigate the nexus between mortality and health capacity utilization based on the balanced panel data of public hospitals in 31 provinces of China from 2009 to 2019.

We adopt a nonparametric approach to estimate distance functions and introduce short-run and long-run capacity measurements with undesirable outputs.

3. Methodology

3.1 Feature selection in machine learning

The first concern for machine learning is treating data, in which extracting and expressing feature vectors are the two keys of data processing. A feature refers to an attribute that describes a pattern. We can define feature selection in machine learning as the determination of a set of features from which a subset is selected to optimize the evaluation standard. Research on feature selection has received scholarly attention in the field of machine learning since the 1990s. For example, Langley et al. (2004) showed that sample complexity increases exponentially with irrelevant features, which is a property shared by other algorithms. Furthermore, Zhang et al. (2004) discussed important issues regarding feature selection and systematically introduced three structures combining feature selection and learning algorithms. In an in-depth analysis of feature selection methods in machine learning, Cui et al. (2018) provided potential future directions.

Feature engineering is important in machine learning and is inseparable from feature selection. Dimensionality reduction is one of the most important problems encountered when comparing the dimensionality of existing data. Dimension reduction refers to reducing the number of bits of data while preserving the nature of the data. We understand the dimension reduction operation as a mapping relationship, and thus there are many available technologies for processing. Sun et al. (2019) proposed that data dimensionality reduction methods can be divided into three categories from either a linear or nonlinear perspective. After dimensionality reduction, linear relationships can still be maintained between data in low-dimensional space using methods such as linear discriminant analysis (LDA), locality preserving projection (LPP), and so on.

Regarding supervised and unsupervised methods, these are differentiated by whether category label information exists in the data sample, such as principal component analysis (PCA), LPP, and LDA. In terms of global vs. local methods, local methods only consider the local information of the sample set, that is, the relationship between data points and critical points, such as Laplacian Eigenmaps and LPP, and global algorithms include PCA and LDA, among others.

In this paper, we focus on PCA technology, which was first proposed by Pearson (1901). Hotelling (1933) further completed the mathematical basis of PCA, which is essentially the Karhunen-Loève transform.

PCA is a commonly used dimension reduction technology. PCA reduces V-dimensional features to K-dimensional features via dimensionality reduction, that is, high-dimensional data are mapped to low-dimensional space and expected to retain more features of the original data with fewer data dimensions to maximize variance on the projected data dimensions. To achieve the maximum variance after data mapping, we set the V-dimensional vector V as the mapping vector as follows:

$$\max_{V} \frac{1}{i-1} \sum_{i=1}^{I} (D^T (X_i - \overline{X}))^2 , \qquad (1)$$

where I is index data quantity (I provinces), X_i refers to the ith vector expression of data, and $\overline{X}$ represents the average of all data instances. We define V as a matrix including all mapping vectors and obtain the following optimization objective functions by changing linear algebra:

$$\min_{V} tr(D^T AD), s.t. D^T D = W , \qquad (2)$$

where tr is the trace of the matrix,

$$A = \frac{1}{i-1} \sum_{i=1}^{I} (X_i - \overline{X})(X_i - \overline{X})^T . \qquad (3)$$

A is the data covariance matrix, each eigenvector of the covariance matrix is a projection plane, and the eigenvalue corresponding to each eigenvector is the variance of the original feature on the projection plane. As k projection surfaces generate k new features, we need to select the projection with a large variance to project the original feature, that is, to select the feature vector with a large eigenvalue. Thus, the output of PCA is $Y = V'X$.

3.2 Technology: Distance functions and efficiency measures

First, basic symbols and techniques are defined. Given the N-dimensional input vector $x \in R_+^N$, the M-dimensional desirable (good) output vector $y \in R_+^M$, and the L-dimensional undesirable (bad) output $z \in R_+^L$, we define the technology T as $T = \{(x, y, z) | \text{x can produce y and z}\}$. Identically, the output set related to T can be expressed as all outputs generated from a given level of input: $O(x) = \{y, z | (x, y, z) \in T\}$.

This technology meets some basic economic assumptions, as shown below:

(T.1) $(0,0,0) \in T$, and if $(0, y, z) \in T$, then $y = 0, z = 0$.

(T.2) T is a closed subset of $R_+^N \times R_+^M \times R_+^L$.

(T.3) For each input $x \in R_+^N$, T is bounded.

(T.4) If $(x, y, z) \in T$ and $(x', y', z') \in R_+^N \times R_+^M \times R_+^L$, then $(x', -y', z') \geq (x, -y, z) \Rightarrow (x', y', z') \in T$.

(T.5) T is convex.

It should be noted that not all axioms can exist in an analysis simultaneously, and the specific scale income hypothesis is not added. Next, the input vector can be classified as variable and fixed, namely $x = (x^f, x^v), x^v \in R_+^{N_v}$ and $x^f \in R_+^{N_f}$ with $N = N_v + N_f$. Färe et al. (1989a) showed that a fixed input is associated with some outputs, and an output is also associated with some fixed inputs.

In the short run, we can define SR technology $SR^f = \{(x^f, y, z) \in R_+^N \times R_+^M \times R_+^L \mid (x^f, x^v)$ can produce y and $z\}$, input set $P^f(y,z) = \{x^f \mid (x^f, y, z) \in SR^f\}$ and output set $O(x^f) = \{y, z \mid (x^f, y, z) \in SR^f\}$. Note that technology $T \in R^{N+M+L}$ can project into the subspace $T \in R^{N_f+M+L}$, from which we obtain technology SR^f. Similarly, they can be applied to the input set $P^f(y,z)$ and output set $O(x^f)$.

Next, we explain the distance function and efficiency measure. First, the input distance function contains all input characteristics, and we can define the radial input efficiency measures with considering undesirable outputs as:

$$DF_i(x, y, z) = \min\{\lambda : \lambda \geq 0, \lambda x \in P(y, z)\}, \tag{4}$$

where λ denotes that the input at a given output level may be proportionally reduced. The main property of $DF_i(x,y,z)$ is that it ranges from 0 to 1($0 \leq DF_i(x,y,z) \leq 1$), meaning the best practice is $DF_i(x,y,z) = 1$.

Similarly, the output distance function contains complete characteristics of $O(x)$, one can describe the radial output efficiency measures as:

$$DF_o(x, y, z) = \max\{\theta : \theta \geq 0, \theta y \text{ and } z/\theta \in O(x)\}. \tag{5}$$

The index of θ to measure technical efficiency shows that the maximum proportion of output expansion can be achieved at a given input level. The main property of $DF_o(x,y,z)$ is greater than (or equal to) 1. At the production frontier, efficient DMU is $DF_o(x,y,z)=1$, and inefficient units are located within the possibility of the output set $DF_o(x,y,z)>1$.

We use $DF_o^f(x^f,y,z)$ to represent the radial output efficiency of output set $O(x^f)$, and the definition of the efficiency measure is $DF_o^f(x^f,y,z)=\max\{\theta:\theta\geq 0,\theta y \text{ and } z/\theta\in O(x^f)\}$. Next, we denote $DF_o(y,z)=\max\{\theta:\theta\geq 0,\theta y \text{ and } z/\theta\in O(x)\}$. Comparing $DF_o(y,z)$ and $DF_o^f(x^f,y,z)$, the former does not rely on the specific input vector x.

Additionally, we need to pay attention to some special definitions. First, $P(0,0)=\{x:(x,0,0)\in T\}$ refers to the input set with zero output. Second, $DF_i^{SR}(x^f,x^v,y,z)=\min\{\lambda:\lambda\geq 0,(x^f,\lambda x^v)\in P(y,z)\}$ is the input efficiency measure of the sub-vector reducing variable input in the input-oriented evaluation. Finally, $DF_i^{SR}(x^f,x^v,0,0)=\min\{\lambda:\lambda\geq 0,(x^f,\lambda x^v)\in P(0,0)\}$ is the sub-vector input efficiency measure with variable input reduced and tangent output zero in the input-oriented evaluation.

Chambers et al. (1996) introduced the directional distance function (DDF) that may have input-oriented or / and output-oriented measures. The direction vectors can be denoted as $g=(-g_x,g_y,-g_z),g_x\in R_+^N,g_y\in R_+^M,g_z\in R_+^L$. Thus, the input and output DDF can be defined as follows:

$$\begin{aligned} DDF_i &= \max\{\gamma:(x-\gamma g_x)\in T\} \\ DDF_o &= \max\{\mu:(y+\mu g_y,z-\mu g_z)\in T\}, \end{aligned} \tag{6}$$

where γ and μ are inefficiencies for input and output DDFs. The relationship between the input-oriented DDF and distance function is expressed as $1-\gamma=\lambda$. Similarly, for the output-oriented DDF and distance function, we can get $\mu+1=\theta$.

3.3 Plant capacity utilization with undesirable outputs

3.3.1. Short-run plant capacity utilization

First, according to Färe et al. (1989a) and Färe et al. (1989b), we introduce undesirable output z and define short-run plant capacity utilization as $PCU_o^{SR}(x,x^f,y,z)$. $PCU_o^{SR}(x,x^f,y,z)$ can be expressed as the ratio of a conventional production technology $DF_o^{SR}(x,y,z)$ to the same technology $DF_o^{SR}(x^f,y,z)$ that does not limit the use of variable inputs. Then, we express it as a SR output-oriented DDFs $DDF_o^{SR}(x,y,z)$ and $DDF_o^{SR}(x^f,y,z)$, namely:

$$PCU_o^{SR}(x,x^f,y,z)=\frac{DF_o^{SR}(x,y,z)}{DF_o^{SR}(x^f,y,z)}=\frac{DDF_o^{SR}(x,y,z)+1}{DDF_o^{SR}(x^f,y,z)+1}. \quad (7)$$

The difference between $DF_o^{SR}(x,y,z)$ and $DF_o^{SR}(x^f,y,z)$ is whether there are variable inputs. Notice that $0<PCU_o^{SR}(x,x^f,y,z)\leq 1$, since $1\leq DF_o^{SR}(x,y,z)\leq DF_o^{SR}(x^f,y,z)$. Therefore, $PCU_o^{SR}(x,x^f,y,z)$ has no lower limit. It is less than 1 when the maximum output of a given input is compared to the maximum output of a sample that may have an infinite number of variable inputs. This explains that there are some relations between the existing effective output quantity and the maximum number of effective outputs. We can obtain the SR output-oriented decomposition by Färe et al. (1989a):

$$DF_o^{SR}(x,y,z)=DF_o^{SR}(x^f,y,z)\cdot PCU_o^{SR}(x,x^f,y,z). \quad (8)$$

In Equation (8), $DF_o^{SR}(x,y,z)$ is divided into two plant capacity measures: $DF_o^{SR}(x^f,y,z)$ is biased, and $PCU_o^{SR}(x,x^f,y,z)$ is unbiased (Färe et al., 1989a; Shen et al., 2022).

Following Cesaroni et al. (2017), we obtain the definition of SR input-oriented plant capacity utilization $PCU_i^{SR}(x,x^f,y,z)$ after introducing undesirable output, we define it as the ratio of the production technology to reduce variable input $DF_i^{SR}(x^f,x^v,y,z)$ to the same technology aimed at zero output $DF_i^{SR}(x^f,x^v,0,0)$. $PCU_i^{SR}(x,x^f,y,z)$ can be expressed by the sum of the short-run input-oriented DDFs $DDF_i^{SR}(x^f,x^v,y,z)$ and $DDF_i^{SR}(x^f,x^v,0,0)$. The equation is as follows:

$$PCU_i^{SR}(x,x^f,y,z)=\frac{DF_i^{SR}(x^f,x^v,y,z)}{DF_i^{SR}(x^f,x^v,0,0)}=\frac{1-DDF_i^{SR}(x^f,x^v,y,z)}{1-DDF_i^{SR}(x^f,x^v,0,0)}. \quad (9)$$

The two sub-vector input efficiency measures $DF_i^{SR}(x^f,x^v,y,z)$ and $DF_i^{SR}(x^f,x^v,0,0)$ only reduce the variable input, the output of the latter is zero. We can see $PCU_i^{SR}(x,x^f,y,z)\geq 1$, $0<DF_i^{SR}(x^f,x^v,0,0)\leq DF_i^{SR}(x^f,x^v,y,z)\leq 1$ .Therefore, $PCU_i^{SR}(x,x^f,y,z)$ takes one unit as the lower limit, but there is no upper limit. It compares the quantitative minimum variable input with the minimum variable input at the beginning of production, which is greater than one unit at the beginning of production. It explains that when the number of variable inputs at the beginning of production increases, the number of current variable inputs also increases, and then the current output can be produced. Cesaroni et al. (2019) provide a decomposition of the SR input-oriented measure:

$$DF_i^{SR}(x^f,x^v,y,z)=DF_i^{SR}(x^f,x^v,0,0)\cdot PCU_i^{SR}(x,x^f,y,z). \quad (10)$$

As in Equation (10), $DF_i^{SR}(x^f,x^v,y,z)$ is also divided into biased and unbiased plant capacity measures, that is, $DF_i^{SR}(x^f,x^v,0,0)$ and $PCU_i^{SR}(x,x^f,y,z)$.

In this study, we introduce the death rate as the undesirable output in medical and health institutions (public hospitals). Following Murty et al., (2012) and Baležentis et al. (2021), we introduce mortality rate into the hospital production technology which contains two sub-technologies: with and without desirable outputs, *Z* is the death (rate) of the undesirable output. The first sub-technology model the normal production process without considering the mortality rate while the second one is initialized to model the undesirable output. We obtain the measurements of short-run PCU as shown in Table 1.

Table 1 Short-run plant capacity utilization measures

Without *Z*	Measure	Notation	Scope
Output-oriented	Biased	$DF_o^{SR}(x^f, y)$	$[1, +\infty)$
	Unbiased	$PCU_o^{SR}(x, x^f, y)$	$(0,1]$
Input-oriented	Biased	$DF_i^{SR}(x^f, x^v, 0)$	$(0,1]$
	Unbiased	$PCU_i^{SR}(x, x^f, y)$	$[1, +\infty)$

With *Z*	Measure	Notation	Scope
Output-oriented	Biased	$DF_o^{SR}(x^f, y, z)$	$[1, +\infty)$
	Unbiased	$PCU_o^{SR}(x, x^f, y, z)$	$(0,1]$
Input-oriented	Biased	$DF_i^{SR}(x^f, x^v, 0, 0)$	$(0,1]$
	Unbiased	$PCU_i^{SR}(x, x^f, y, z)$	$[1, +\infty)$

3.3.2 Long-run plant capacity utilization

In this section, to facilitate brevity, we do not differentiate between variable and fixed inputs. We introduce undesirable output according to the new definition from Cesaroni et al. (2019) and obtain LR output-oriented plant capacity utilization $PCU_o^{LR}(x, y, z)$. $PCU_o^{LR}(x, y, z)$ is a ratio of the standard production technology

$DF_o^{LR}(x,y,z)$ to the same technology without any input limitation $DF_o^{LR}(y,z)$. It also can be represented by the long-run output-oriented DDFs $DDF_o^{LR}(x,y,z)$ and $DDF_o^{LR}(y,z)$, as shown in the following equation:

$$PCU_o^{LR}(x,y,z) = \frac{DF_o^{LR}(x,y,z)}{DF_o^{LR}(y,z)} = \frac{DDF_o^{LR}(x,y,z)+1}{DDF_o^{LR}(y,z)+1}. \tag{11}$$

$DF_o^{LR}(x,y,z)$ and $DF_o^{LR}(y,z)$ contain all the input and output technology, and Notice that $0 < PCU_o^{LR}(x,y,z) \le 1$, since $1 \le DF_o^{LR}(x,y,z) \le DF_o^{LR}(y,z)$. Observation shows that $PCU_o^{LR}(x,y,z)$ has a fixed upper limit but no lower limit. $DF_o^{LR}(x,y,z)$ represents the maximum output under a given input, and $DF_o^{LR}(y,z)$ represents the maximum output without fixed input. The scale of comparison between the two is smaller than one unit. Here, we note that the molecules of Equations (7) and (11) are the same because the fixed input can be adjusted. Thus, we can adjust inputs to maximize output. The same explanation is also applicable to SR samples. According to Cesaroni et al. (2019), we can decompose the LR output-oriented measure as follows:

$$DF_o^{LR}(x,y,z) = DF_o^{LR}(y,z) \cdot PCU_o^{LR}(x,y,z), \tag{12}$$

where $DF_o^{LR}(x,y,z)$ is divided into two different plant capacity measures: $DF_o^{LR}(y,z)$ is biased, and $PCU_o^{LR}(x,y,z)$ is unbiased.

Following Cesaroni et al. (2019), we introduce the undesirable output of mortality. $PCU_i^{LR}(x,y,z)$ is defined as the ratio between input efficiency measures at a given output level $DF_i^{LR}(x,y,z)$ and the null output level $DF_i^{LR}(x,0,0)$. Then, we use the long-run input-oriented DDFs $DDF_i^{LR}(x,y,z)$ and $DDF_i^{LR}(x,0,0)$ to express as follows:

$$PCU_i^{LR}(x,y,z)=\frac{DF_i^{LR}(x,y,z)}{DF_i^{LR}(x,0,0)}=\frac{1-DDF_i^{LR}(x,y,z)}{1-DDF_i^{LR}(x,0,0)}, \quad (13)$$

where $DF_i^{LR}(x,y,z)$ aims to reduce all inputs, and $DF_i^{LR}(x,0,0)$ is an efficiency measure assessed at the null output level. What should be noticed is the precondition: $DF_i(x,0,0)=\min\{\lambda:\lambda\geq 0,\lambda x\in P(0,0)\}$. Notice that $PCU_i^{LR}(x,y,z)\geq 1$, since $0<DF_i^{LR}(x,0,0)\leq DF_i^{LR}(x,y,z)\leq 1$. $DF_i^{LR}(x,y,z)$ is the minimum input under a given output, $DF_i^{LR}(x,0,0)$ is the minimum of all inputs at the start of production. The scale of comparison between this two is greater than one unit. $PCU_i^{LR}(x,y,z)$ explains that the number of all inputs corresponding to the initial production must be increased proportionally to generate the current output. Similarly, LR input-oriented measure can be decomposed into:

$$DF_i^{LR}(x,y,z)=DF_i^{LR}(x,0,0)\cdot PCU_i^{LR}(x,y,z). \quad (14)$$

$DF_i^{LR}(x,y,z)$ is divided into two plant capacity measures, where $DF_i^{LR}(x,0,0)$ is biased, and $PCU_i^{LR}(x,y,z)$ is unbiased. To summarize the above, we obtain the LR plant capacity measures as shown in Table 2: *Z* is the death (rate) of the undesirable output.

Table 2 Long-run plant capacity measures

Without *Z*	Measure	Notation	Scope
Output-oriented	Biased	$DF_o^{LR}(y)$	$[1,+\infty)$
	Unbiased	$PCU_o^{LR}(x,y)$	$(0,1]$
Input-oriented	Biased	$DF_i^{LR}(x,0)$	$(0,1]$
	Unbiased	$PCU_i^{LR}(x,y)$	$[1,+\infty)$

With *Z*	Measure	Notation	Scope

Output-oriented	Biased	$DF_o^{LR}(y,z)$	$[1,+\infty)$
	Unbiased	$PCU_o^{LR}(x,y,z)$	$(0,1]$
Input-oriented	Biased	$DF_i^{LR}(x,0,0)$	$(0,1]$
	Unbiased	$PCU_i^{LR}(x,y,z)$	$[1,+\infty)$

Overall, the short-run and long-run PCU measures with undesirable outputs have been introduced. The detailed graphical illustration of PCU measure is available from Cesaroni et al. (2019). The initial PCUs without undesirable outputs can be measured by distance functions under a nonparametric approach. For PCU measures with undesirable outputs, we apply directional distance functions following Shen et al. (2022), who introduce the weak disposability model to incorporate undesirable outputs. Alternatively, we introduce a by-production approach to model hospital production technology.

4. Nonparametric technologies

In view of the practicality of the study, we use nonparametric frontier technology to define the plant capacity. Therefore, the input-output vector (x_i, y_i, z_i) is expressed as $(i=1,2,...,I)$ by introducing the undesirable output, where the input and output are processed (Murty et al., 2012; Baležentis et al., 2021). We define this as:

$$\begin{aligned} T_{VRS} = \{(x,y,z): x \geq \sum_{i=1}^{I}\sigma_i^1 x_i, y \leq \sum_{i=1}^{I}\sigma_i^1 y_i, \sum_{i=1}^{I}\sigma_i^1 = 1, \sigma_i^1 \geq 0, \\ \sum_{i=1}^{I}\sigma_i^2 y_i = \sum_{i=1}^{I}\sigma_i^1 y_i, z \geq \sum_{i=1}^{I}\sigma_i^2 z_i, \sum_{i=1}^{I}\sigma_i^2 = 1, \sigma_i^2 \geq 0, i=1,2,...,I.\} \end{aligned}, \quad (15)$$

where σ_i^1 and σ_i^2 are the activity variables for two sub-technologies. We adopt the variable returns to scale (VRS) for the hospital production technology. Without desirable outputs, traditional hospital production technology with normal inputs and outputs is introduced. With desirable outputs, the mortality rate is considered while normal inputs may not cause death directly in hospitals. Alternatively, surgical risk is strongly associated with medical mortality, and an increase in the number of surgical procedures, inpatients, and outpatients may lead to greater medical risk for causing a higher mortality. Therefore, the latter is no more a production technology with inputs and outputs, and it can be a benchmark model. To keep the same optimal level of desirable outputs between sub-technologies ($\sum_{i=1}^{I}\sigma_i^2 y_i = \sum_{i=1}^{I}\sigma_i^1 y_i$), the benchmark model is to seek a low level of morality rate among hospitals (see details in Baležentis et al., 2021).

4.1 Plant capacity utilization without undesirable outputs

4.1.1 Short-run plant capacity utilization

For a clearer explanation, we add linear programs (LPs) to calculate SR plant capacity measure. It should be noted medical institutions do not introduce undesirable outputs. From the observed value (x_0, y_0), we obtain $DDF_o^{SR}(x, y)$ as follows:

$$
\begin{aligned}
& DDF_o^{SR}(x,y) \;=\; \max_{\mu,\sigma} \mu \\
s.t.\; & \sum_{i=1}^{I}\sigma_i\, x_i^n \le x_{i'}^n, n = 1,2,...,N \\
& \sum_{i=1}^{I}\sigma_i\, y_i^m \ge y_{i'}^m + \mu g_y, m = 1,2,...,M \\
& \sigma_i \;\ge 0, \qquad\qquad . \qquad\qquad (16) \\
& \sum_{i=1}^{I}\sigma_i \;=1, \\
& i = 1,2,...,I
\end{aligned}
$$

Following Chambers et al. (1996), we calculate $DDF_o^{SR}(x^f, y)$ from (x_0, y_0):

$$DDF_o^{SR}(x^f, y) = \max_{\mu,\sigma} \mu$$

$$\begin{aligned}
s.t.\ & \sum_{i=1}^{I} \sigma_i x_i^{N_f} \le x_{i'}^{N_f}, N = 1,2,\dots,N_f \\
& \sum_{i=1}^{I} \sigma_i y_i^m \ge y_{i'}^m + \mu g_y, m = 1,2,\dots,M \\
& \sigma_i \ge 0, \\
& \sum_{i=1}^{I} \sigma_i = 1, \\
& i = 1,2,\dots,I
\end{aligned} \quad . \qquad (17)$$

Each variable input dimension introduces an optional LP with a scalar, and there is no input constraint on the variable input. LP (16) is equivalent to transforming the input of each variable into a decision variable. Next, we discuss Equation (9) without undesirable output. From the observed value (x_0, y_0), we obtain $DDF_i^{SR}(x^f, x^v, y)$ as follows:

$$DDF_i^{SR}(x^f, x^v\ y) = \max_{\gamma,\sigma} \gamma$$

$$\begin{aligned}
s.t.\ & \sum_{i=1}^{I} \sigma_i x_i^{N_f} \le x_{i'}^{N_f}, N = 1,2,\dots,N_f \\
& \sum_{i=1}^{I} \sigma_i x_i^{N_v} \le x_{i'}^{N_v} - \gamma g_{x^v}, N = 1,2,\dots,N_v \\
& \sum_{i=1}^{I} \sigma_i y_i^m \ge y_{i'}^m, m = 1,2,\dots,M \\
& \sigma_i \ge 0, \\
& \sum_{i=1}^{I} \sigma_i = 1, \\
& i = 1,2,\dots,I
\end{aligned} \quad . \qquad (18)$$

Then, we calculate $DDF_i^{SR}(x^f, x^v, 0)$ from (x_0, y_0):

$$
\begin{aligned}
& DDF_i^{SR}(x^f, x^v 0) = \max_{\gamma,\sigma} \gamma \\
s.t. \ & \sum_{i=1}^{I} \sigma_i x_i^{N_f} \le x_{i'}^{N_f}, N = 1,2,\dots,N_f \\
& \sum_{i=1}^{I} \sigma_i x_i^{N_v} \le x_{i'}^{N_v} - \gamma g_{x^v}, N = 1,2,\dots,N_v \\
& \sum_{i=1}^{I} \sigma_i y_i^{m} \ge 0, m = 1,2,\dots,M \\
& \sigma_i \ge 0, \\
& \sum_{i=1}^{I} \sigma_i = 1, \\
& i = 1,2,\dots,I
\end{aligned} \quad . \tag{19}
$$

In Equation (19), the output in the condition is set to zero. If each input is fixed at the minimum level of all observations, the right side of each DMU and the results of $DDF_i^{SR}(x^f, x^v, 0)$ will be the same. LPs (17) and (19) have similarities: The former is the constraint of variable inputs, and the latter is the output constraint.

4.1.2 Long-run plant capacity utilization

According to Formulas (11) and (13), we only need to calculate three efficiency measures. First, when calculating the output-oriented situation, it should be noted that $DDF_o^{LR}(x, y)$ has been calculated in Equation (16), thus we only need to calculate $DDF_o^{LR}(y)$ from the observed value (x_0, y_0):

$$
\begin{aligned}
& DDF_o^{LR}(y) = \max_{\mu,\sigma,x} \mu \\
s.t.\ & \sum_{i=1}^{I} \sigma_i x_i^n \le x^n, n = 1,2,...,N \\
& \sum_{i=1}^{I} \sigma_i y_i^m \ge y_{i'}^m + \mu g_y, m = 1,2,...,M \\
& \sigma_i \ge 0, \\
& \sum_{i=1}^{I} \sigma_i = 1, \\
& x^n \ge 0, \\
& i = 1,2,...,I
\end{aligned} \quad . \qquad (20)
$$

LP (20) is the long-run representation of LP (17), and x is no longer a real unit value. Next, we turn to the input-oriented situation and compute $DDF_i^{LR}(x,y)$ as follows:

$$
\begin{aligned}
& DDF_i^{LR}(x,y) = \max_{\gamma,\sigma} \gamma \\
s.t.\ & \sum_{i=1}^{I} \sigma_i x_i^n \le x_{i'}^n - \gamma g_x, n = 1,2,...,N \\
& \sum_{i=1}^{I} \sigma_i y_i^m \ge y_{i'}^m, m = 1,2,...,M \\
& \sigma_i \ge 0, \\
& \sum_{i=1}^{I} \sigma_i = 1, \\
& i = 1,2,...,I
\end{aligned} \quad . \qquad (21)
$$

Here, the input vector no longer distinguishes between fixed and variable inputs since policymakers have enough time to adjust and utilize the required inputs in the long run. Finally, calculating $DDF_i^{LR}(x,0)$:

$$
\begin{aligned}
& DDF_i^{LR}(x,0) \;=\; \max_{\gamma,\sigma} \gamma \\
s.t. \;& \sum_{i=1}^{I} \sigma_i \, x_i^n \le x_{i'}^n - \gamma g_x, n = 1, 2, ..., N \\
& \sum_{i=1}^{I} \sigma_i \, y_i^m \ge 0, m = 1, 2, ..., M \\
& \sigma_i \;\ge 0, \\
& \sum_{i=1}^{I} \sigma_i \;= 1, \\
& i = 1, 2, ..., I
\end{aligned} \qquad . \qquad (22)
$$

In Equation (22), the output in the condition is set to zero, that is, the input-oriented output is zero. If each input is fixed at the minimum level of all observations, the right side of each DMU and the results of $DDF_i^{LR}(x,0)$ are the same. Among them, LPs (20) and (22) have similarities. The former completely comprises input constraints and the latter output constraints.

4.2 Plant capacity utilization with undesirable outputs

4.2.1 Short-run plant capacity utilization

Here, we introduce the undesirable output and define the capacity measurement of each plant. Following Baležentis et al. (2021), we adopt a refined by-production model to incorporate mortality rate into production technology. From the observed values (x_0, y_0, z_0), we obtain $DDF_o^{SR}(x,y,z)$ as follows:

$$DDF_o^{SR}(x,y,z) = \max_{\mu,\sigma} \mu$$

$$\begin{aligned}
s.t. \ & \sum_{i=1}^{I} \sigma_i^1 x_i^n \le x_{i'}^n, n=1,2,...,N \\
& \sum_{i=1}^{I} \sigma_i^1 y_i^m \ge y_{i'}^m + \mu g_y, m=1,2,...,M \\
& \sigma_i^1 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^1 = 1, \\
& \sum_{i=1}^{I} \sigma_i^2 y_i^m = \sum_{i=1}^{I} \sigma_i^1 y_i^m, m=1,2,...,M \\
& \sum_{i=1}^{I} \sigma_i^2 z_i^l \le z_{i'}^l - \mu g_z, l=1,2,...,L \\
& \sigma_i^2 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^2 = 1, \\
& i=1,2,...,I
\end{aligned} \quad (23)$$

.

According to the above section, $DDF_o^{SR}(x^f, y, z)$ can be calculated as follows:

$$DDF_o^{SR}(x^f,y,z) = \max_{\mu,\sigma} \mu$$

$$\begin{aligned}
s.t. \ & \sum_{i=1}^{I} \sigma_i^1 x_i^{N_f} \le x_{i'}^{N_f}, N=1,2,...,N_f \\
& \sum_{i=1}^{I} \sigma_i^1 y_i^m \ge y_{i'}^m + \mu g_y, m=1,2,...,M \\
& \sigma_i^1 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^1 = 1, \\
& \sum_{i=1}^{I} \sigma_i^2 y_i^m = \sum_{i=1}^{I} \sigma_i^1 y_i^m, m=1,2,...,M \\
& \sum_{i=1}^{I} \sigma_i^2 z_i^l \le z_{i'}^l - \mu g_z, l=1,2,...,L \\
& \sigma_i^2 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^2 = 1, \\
& i=1,2,...,I
\end{aligned} \quad (24)$$

.

Next, we discuss Equation (11) with the undesirable output. From the observed values (x_0, y_0, z_0), we obtain $DDF_i^{SR}(x^f, x^v, y, z)$ as follows:

$$
\begin{aligned}
& DDF_i^{SR}(x^f, x^v, y, z) = \max_{\gamma,\sigma} \gamma \\
s.t.\ & \sum_{i=1}^{I} \sigma_i^1 x_i^{N_f} \leq x_{i'}^{N_f}, N = 1,2,...,N_f \\
& \sum_{i=1}^{I} \sigma_i^1 x_i^{N_v} \leq x_{i'}^{N_v} - \gamma g_{x^v}, N = 1,2,...,N_v \\
& \sum_{i=1}^{I} \sigma_i^1 y_i^m \geq y_{i'}^m, m = 1,2,...,M \\
& \sigma_i^1 \geq 0, \\
& \sum_{i=1}^{I} \sigma_i^1 = 1, \\
& \sum_{i=1}^{I} \sigma_i^2 y_i^m = \sum_{i=1}^{I} \sigma_i^1 y_i^m, m = 1,2,...,M \\
& \sum_{i=1}^{I} \sigma_i^2 z_i^l \leq z_{i'}^l, l = 1,2,...,L \\
& \sigma_i^2 \geq 0, \\
& \sum_{i=1}^{I} \sigma_i^2 = 1, \\
& i = 1,2,...,I
\end{aligned} \tag{25}
$$

Then, we can obtain $DDF_i^{SR}(x^f, x^v, 0, 0)$:

$$
\begin{aligned}
& DDF_i^{SR}(x^f, x^v, 0, 0) = \max_{\gamma,\sigma} \gamma \\
s.t. \quad & \sum_{i=1}^{I} \sigma_i^1 x_i^{N_f} \le x_{i'}^{N_f}, N = 1, 2, ..., N_f \\
& \sum_{i=1}^{I} \sigma_i^1 x_i^{N_v} \le x_{i'}^{N_v} - \gamma g_{x^v}, N = 1, 2, ..., N_v \\
& \sum_{i=1}^{I} \sigma_i^1 y_i^m \ge 0, m = 1, 2, ..., M \\
& \sigma_i^1 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^1 = 1, \\
& \sum_{i=1}^{I} \sigma_i^2 y_i^m = \sum_{i=1}^{I} \sigma_i^1 y_i^m, m = 1, 2, ..., M \\
& \sum_{i=1}^{I} \sigma_i^2 z_i^l \ge 0, l = 1, 2, ..., L \\
& \sigma_i^2 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^2 = 1, \\
& i = 1, 2, ..., I
\end{aligned} \tag{26}
$$

We observe that LPs (24) and (26) have similarities. The former is the constraint of variable inputs, and the latter is the output constraint.

4.2.2 Long-run plant capacity utilization

As in the previous section, to obtain LR plant capacity measures, we only need to calculate three efficiency measures. First, we calculate the output-oriented situation, and it should be noted that $DDF_o^{LR}(x, y, z)$ is calculated in Equation (23). Thus, we only need to calculate $DDF_o^{LR}(y, z)$:

$$
\begin{aligned}
& DDF_o^{LR}(y,z) = \max_{\mu,\sigma,x} \mu \\
s.t.\ & \sum_{i=1}^{I} \sigma_i^1 x_i^n \le x^n, n = 1,2,...,N \\
& \sum_{i=1}^{I} \sigma_i^1 y_i^m \ge y_{i'}^m + \mu g_y, m = 1,2,...,M \\
& \sigma_i^1 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^1 = 1, \\
& \sum_{i=1}^{I} \sigma_i^2 y_i^m = \sum_{i=1}^{I} \sigma_i^1 y_i^m, m = 1,2,...,M \\
& \sum_{i=1}^{I} \sigma_i^2 z_i^l \le z_{i'}^l - \mu g_z, l = 1,2...,L \\
& \sigma_i^2 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^2 = 1, \\
& x^n \ge 0, \\
& i = 1,2,...,I
\end{aligned} \quad . \tag{27}
$$

LP (27) is the long-run representation of LP (24). Here, x is no longer a real unit value. Next, regarding the input-oriented situation, we calculate $DDF_i^{LR}(x,y,z)$:

$$
\begin{aligned}
& DDF_i^{LR}(x,y,z) = \max_{\gamma,\sigma} \gamma \\
s.t.\ & \sum_{i=1}^{I} \sigma_i^1 x_i^n \le x_{i'}^n - \gamma g_x, n = 1,2,...,N \\
& \sum_{i=1}^{I} \sigma_i^1 y_i^m \ge y_{i'}^m, m = 1,2,...,M \\
& \sigma_i^1 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^1 = 1, \\
& \sum_{i=1}^{I} \sigma_i^2 y_i^m = \sum_{i=1}^{I} \sigma_i^1 y_i^m, m = 1,2,...,M \\
& \sum_{i=1}^{I} \sigma_i^2 z_i^l \le z_{i'}^l, l = 1,2,...,L \\
& \sigma_i^2 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^2 = 1, \\
& i = 1,2,...,I
\end{aligned} \quad . \tag{28}
$$

Similarly, the input vector no longer distinguishes between fixed and variable inputs. Finally, according to the given observation value (x_0, y_0, z_0), we calculate $DDF_i^{LR}(x,0,0)$ as follows:

$$
\begin{aligned}
& DDF_i^{LR}(x,0,0) = \max_{\gamma,\sigma} \gamma \\
& s.t. \ \sum_{i=1}^{I} \sigma_i^1 x_i^n \le x_{i'}^n - \gamma g_x, n = 1,2,...,N \\
& \sum_{i=1}^{I} \sigma_i^1 y_i^m \ge 0, m = 1,2,...,M \\
& \sigma_i^1 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^1 = 1, \\
& \sum_{i=1}^{I} \sigma_i^2 y_i^m = \sum_{i=1}^{I} \sigma_i^1 y_i^m, m = 1,2,...,M \\
& \sum_{i=1}^{I} \sigma_i^2 z_i^l \ge 0, l = 1,2,...,L \\
& \sigma_i^2 \ge 0, \\
& \sum_{i=1}^{I} \sigma_i^2 = 1, \\
& i = 1,2,...,I
\end{aligned}
\tag{29}
$$

In Equation (29), the output in the condition is set to zero, that is, the input-oriented output is zero. LPs (27) and (29) have similarities: the former is entirely input constraints and the latter is entirely output constraints.

Finally, we used the SR and LR plant capacity concepts to evaluate the effective utilization of medical institution capacity in 31 provinces in China over 11 years and examined their correlation with the death rate. Then, we calculated the long-run plant capacity for public hospitals. We will now incorporate real-life data into these plant capacity models.

5. Data collection

To analyze plant capacity utilization in medical institutions, we chose 31 provinces in China to comprise our samples. China is the country with the largest population, so national health is important to policymakers. Different hospitals offer distinct treatments according to the diverse conditions of patients. In other words, hospitals differ in terms of medical services, employees, patients, and production technologies. We used existing data for medical institutions in 31 provinces from 2009 to 2019 from the China Health Statistical Yearbook, China Statistical Yearbook, and China Statistical Yearbook of Traditional Chinese Medicine, with a total sample size of 3410. Through feature selection in machine learning, we used PCA to select eight features. We considered two inputs and two outputs. The input "medical institution" was divided into the number of personnel and beds. Practitioners, registered nurses, and pharmacists were fixed inputs since they would not change significantly in the short run. Other medical personnel and beds were considered as variable inputs. The outputs constituted the number of operations and hospitalizations. We also gathered data on deaths in the health institutions: the number of deaths comprised the sum of emergency deaths, deaths of discharged patients, and deaths in observation rooms. Emergency deaths refer to the number of deaths that occurred in emergency rooms divided by the number of emergency patients, and deaths in observation rooms refer to the number of deaths that occurred in observation rooms divided by the number of observation room cases.

6. Data description and calculation

We defined the production technology of the medical and health institutions according to two types of inputs and two types of outputs. Because of the large scale of the data, they were incomplete, repeated, and messy. Thus, we first needed to preprocess the data before building the model. We set 17 indicators: number of licensed physicians, nurses, relevant technical personnel, other staff, beds, operations, inpatients, outpatients, discharges, total number of deaths, number of medical and health institutions, total assets, total income, total expenditures, the total health costs, total health costs per capita, number of insured individuals, and number of individuals with maternity insurance at the end of the year; these were designated as A-1 to A-17. We used PCA to select features, presented in Table 3.

Table 3 Descriptive analysis of Principal Component Analysis

	1	2	3	4	5	6	7	8
A-1	.987	-.046	-.048	.035	-.014	-.009	.025	.029
A-2	.969	-.203	.008	-.030	.052	-.024	-.067	-.013
A-3	.970	.062	.091	-.047	.047	-.083	.011	-.076
A-4	.944	.081	-.117	.037	.091	-.106	-.013	-.242
A-5	.951	-.139	-.222	-.011	.089	.051	-.045	.058
A-6	.938	-.192	.172	-.121	.016	-.107	-.047	.043
A-7	.942	-.129	-.233	-.076	.143	.053	-.037	.064
A-8	.942	-.127	-.231	-.076	.142	.053	-.038	.064
A-9	.695	.135	-.304	.552	-.287	-.041	-.127	.023
A-10	.791	.133	-.497	.099	-.031	.046	.309	.008
A-11	.611	.748	.200	.033	.016	.037	.012	.037
A-12	.562	.758	.311	.036	.043	.060	-.014	.012
A-13	.480	.810	.306	-.022	.017	.095	-.016	-.004
A-14	.887	-.379	.216	.041	.030	-.050	-.019	.054
A-15	.043	-.572	.625	.474	.190	.108	.080	-.017
A-16	.765	-.429	.129	-.213	-.270	.297	-.026	-.084
A-17	.753	-.165	.519	-.150	-.244	-.175	.114	.051

Li et al. (2010) explained how to correctly apply SPSS software for PCA, proposing that PCA depends on the original variables, and the selection of the original indicators is important. If the original variables are independent, the dimension reduction may fail. The more relevant the data is, the better the effect will be. In PCA, the principal component generally contains at least 90% of the information of each original variable. Therefore, in Table 3, we can extract eight main indicators with coefficients greater than 0.9 through PCA, namely the number of practitioners, nurses, related technical personnel, other staff, beds, operations, hospitalizations, and discharges. Next, we conducted the analysis according to these eight indicators and the undesirable output of mortality.

As mentioned above, the fixed inputs are practitioners, registered nurses, pharmacists, and the variable inputs are other medical personnel and beds. We also introduced the undesirable output of the death rate. Table 4 shows the descriptive statistics of national medical and health institutions from 2009 to 2019. It should be noted that the number of operations, inpatients, and outpatients constitute units of 10 million and the number of hospitalizations units of 100,000.

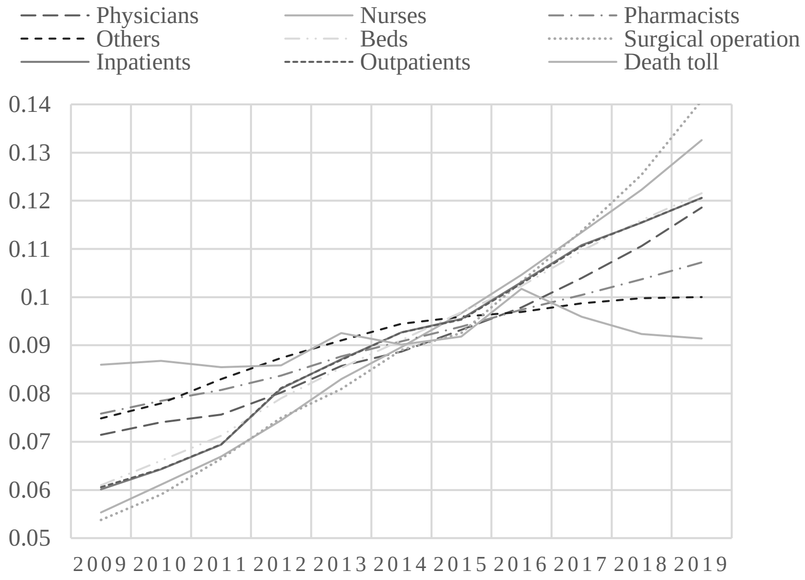
Table 4 Descriptive statistics of national medical and health institutions from 2009 to 2019

		Mean	St.Dev.	Min	Max
Fixed Input	Physicians	29.64	4.84	23.29	38.67
	Nurses	30.48	8.16	18.55	44.45
	Pharmacists	4.10	0.46	3.42	4.83
Variable Input	Other Staffs	12.26	1.15	10.09	13.49
	Beds	65.86	14.08	44.17	8.81
Output	Surgical operation	4.47	1.30	2.65	6.93
	Inpatients	20.05	4.35	13.23	26.60
	Outpatients	19.98	4.31	13.31	25.60
Undesirable Output	Death toll	41.99	2.18	39.49	46.99

This study used two inputs in the interest of pragmatism by observing the evolution of five inputs over time. Among these, the ratio of the number of inpatients to the beds is greater than 1, and the growth rates of inpatients and beds are different. Therefore, the bed is considered a variable input during this period. Regarding staff, since professional qualifications or certification are prerequisite for practitioners, registered nurses and pharmacists are usually regarded as fixed inputs and do not change significantly. We divided other medical personnel into variable inputs.

As shown in Figure 1, the number of licensed doctors, registered nurses, and pharmacists (fixed inputs) increased annually, as did the number of other medical personnel and beds (variable inputs). In addition, the number of surgical operations and inpatients (outputs) also increased annually. At the same time, the number of medical deaths increased gradually before 2016 and decreased gradually after 2016. These data indicate that medical mortality improved before the long-run reaches its maximum level.

Fig 1 The evolution of inputs, outputs, and deaths from 2009 to 2019



When output increases, input-oriented capacity utilization will also increase, and this increase in output requires more input. However, compared with the output-oriented plant capacity measurement, the input-oriented plant capacity measurement does not reach the maximum. Therefore, the input-oriented capacity measurement can determine the minimum variable input that is compatible with zero output (the minimum input needs to start producing non-zero output).

It is noteworthy that introducing undesirable output in the short run has little effect on PCU. Therefore, we primarily considered LR input-oriented or output-oriented medical and health institutions' plant capacity utilization. Table 5 provides descriptive statistics of LR output-oriented efficiency results and PCU. Following Equation (11), technical efficiency is divided into two different measures, biased and unbiased. Among them, the average value of $DF_o^{LR}(y)$ is 6.13 without introducing the death rate; for $PCU_o^{LR}(x, y)$, the average is 0.47. After introducing the death rate, the average value of $DF_o^{LR}(y,z)$ is 1.32; for $PCU_o^{LR}(x, y, z)$, the average is 0.86. Thus, the introduction of the death rate has a significant impact on output-oriented PCU in the long run.

Table 6 provides descriptive statistics of LR input-oriented efficiency results and plant capacity utilization. Following Equation (13), technical efficiency is also divided into two different measures, biased and unbiased. Among them, we can observe that without the introduction of the death rate, the average value of $DF_i^{LR}(x,0)$ is 0.18; for $PCU_i^{LR}(x,y)$, the average is 9.54. After introducing the death rate, the average value of $DF_i^{LR}(x,0,0)$ is 0.18; for $PCU_i^{LR}(x,y,z)$, the average is 9.55. This shows that the introduction of the death rate has no significant effect on input-oriented PCU in the long run.

Table 5 Descriptive statistics of LR output-oriented plant capacity utilization

Without Z	$DF_o^{LR}(x,y)$	$DF_o^{LR}(y)$	$PCU_o^{LR}(x,y)$
Average	1.12	6.13	0.47
St.Dev.	0.17	11.78	0.28
Min	1.00	1.00	0.02
Max	1.61	65.88	1.00

With Z	$DF_o^{LR}(x,y,z)$	$DF_o^{LR}(y,z)$	$PCU_o^{LR}(x,y,z)$
Average	1.11	1.32	0.86
St.Dev.	0.17	0.22	0.11
Min	1.00	1.00	0.65
Max	1.54	1.62	1.00

Table 6 Descriptive statistics of LR input-oriented plant capacity utilization

Without Z	$DF_i^{LR}(x,y)$	$DF_i^{LR}(x,0)$	$PCU_i^{LR}(x,y)$
Average	0.91	0.18	9.54
St.Dev.	0.10	0.21	6.21
Min	0.64	0.04	1.00
Max	1.00	1.00	24.56

With Z	$DF_i^{LR}(x,y,z)$	$DF_i^{LR}(x,0,0)$	$PCU_i^{LR}(x,y,z)$
Average	0.91	0.18	9.55
St.Dev.	0.10	0.21	6.21
Min	0.64	0.04	1.00
Max	1.00	1.00	24.56

Overall, these descriptive statistics show that there are significant differences in the results of long-run plant capacity utilization of medical and health institutions when the undesirable output is introduced. They show that input-oriented measures of plant capacity are better, as when LR input-oriented PCU increases over time, the death rate will also increase.

This study compared the LR output-oriented biased and unbiased capacity utilization measurements with and without the death rate. Figures 2 and 3 show that when the mortality rate is introduced, the biased capacity utilization measurement will decrease, while the unbiased capacity utilization measurement will increase.

Fig 2 Compare the biased output-oriented measures of LR plant capacity utilization with and without undesirable output Z

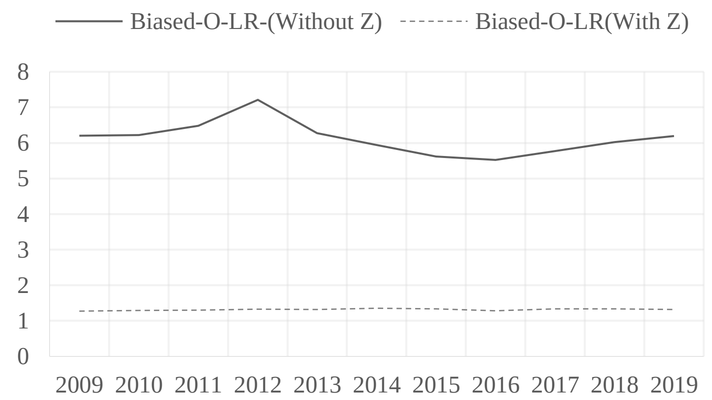
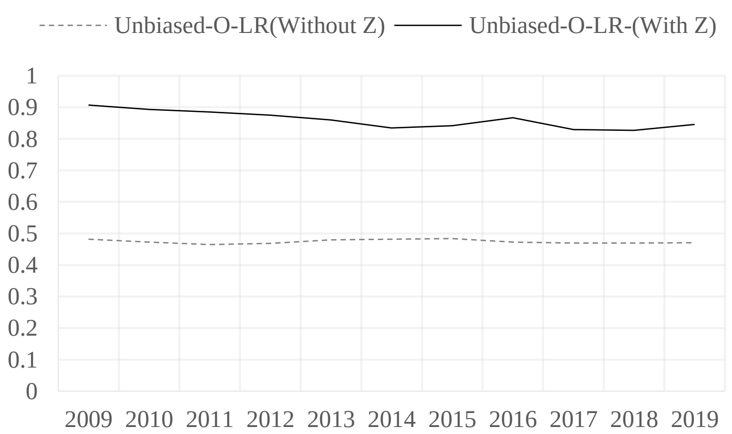
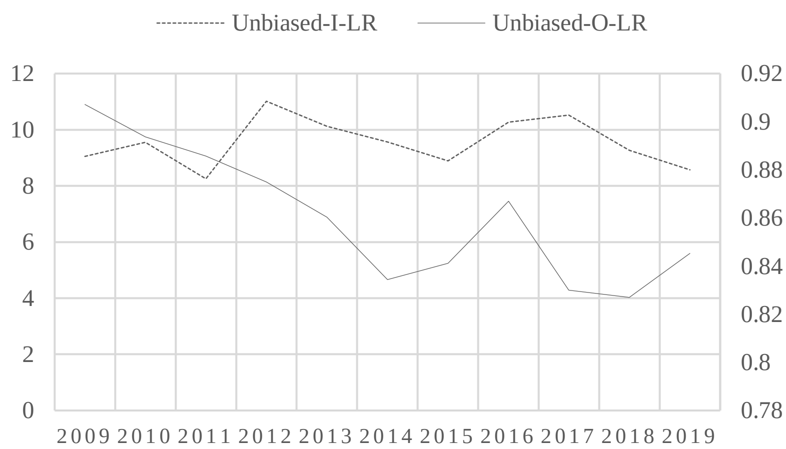


Fig 3 Compare the unbiased output-oriented LR plant capacity measures with and without undesirable output Z



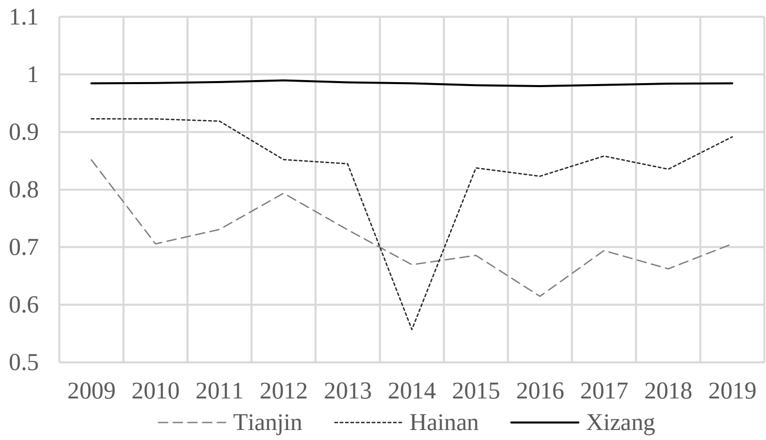
Next, the study focused on the change of LR PCU from 2009 to 2019 after introducing the undesirable output. Figure 4 shows the long-run PCU with the undesirable output under the overall order method from 2009 to 2019. If the dotted line is approximated as a smooth curve, the output-oriented efficiency measurement resembles a U-shaped curve, with the lowest efficiency around 2014. The sudden change in 2016 is due to medical institutions' mortality rate dropping suddenly in 2016, which begins to increase after 2016 with the continuous expansion of the number of beds and other inputs. The input-oriented efficiency measurement is similar to a U-shaped curve.

Fig 4 The LR description of PCU with undesirable outputs Z



Finally, the 31 healthcare institutions' capacity utilization is compared according to the difference between output-oriented PCU with and without introducing mortality. If the difference tends toward 0, the LR output-oriented PCU after introducing the death rate remains essentially unchanged; if the difference tends toward 1, the LR output-oriented PCU changes significantly after introducing the death rate. After introducing the death rate, the output-oriented PCU of five provinces—Hebei, Shandong, Henan, Guangdong, and Sichuan—remained basically unchanged; however, after introducing the death rate, the output-oriented PCU of 26 provinces increased by different degrees. Of these, the output-oriented PCU in Tianjin, Hainan, and Tibet increased sharply. We selected the three provinces with the most output-oriented PCU change after the introduction of the death rate to present in Figure 6.

Fig 6 Unbiased output-oriented LR plant capacity measures in three Chinese provinces with undesirable output *Z*



7. Conclusion

This study measured and summarized the plant capacity of all medical institutions in 31 provinces of China over 11 years. On the basis of the research results, this study provided corresponding policy recommendations. For the Chinese government to make full use of the existing hospital capacity to ensure the health of the people. Challenges are being faced to enable Chinese authorities to use and improve existing hospital capacity more effectively to adequately treat patients.

First, we examined a large amount of economic and medical literature to provide favorable evidence for the connection between plant capacity utilization and the death rate. Next, we introduced the output-oriented and input-oriented plant capacity measures and defined the plant capacity measures in detail from the LR and SR perspectives, this approach breaks through the traditional plant capacity measure. We also introduced an undesirable output and tested the results in four different models that either included or did not include the undesirable output. We know that SR plant capacity measures use fixed inputs, and LR plant capacity measures use all input variables, which can be changed; that is, the variable and fixed inputs are no longer distinguished in long run. Thus, the maximum plant capacity output in the output-oriented case and the minimum input in the input-oriented case with non-zero output can be obtained. After describing all the available data, we obtained eight results from the four different models.

This study made two main contributions. First, we introduced the death rate into the performance evaluation of healthcare institutions in terms of capacity utilization. Second, we calculated the LR plant capacity utilization of public hospitals. We can draw three main conclusions from our results. First, the input-oriented plant capacity measure is superior to the output-oriented plant capacity measure, and high-level LR input-oriented plant capacity utilization is associated with an increased mortality rate, which is consistent with previous research. Second, in the long run, the introduction of the death rate has a significant impact on output-oriented plant capacity utilization. Third, when introducing the death rate into the performance evaluation of healthcare institutions, differences will occur depending on the use of biased and unbiased plant capacity measures: the former will decrease and the latter will increase.

Based on our findings, this paper made the following recommendations:

(1) Further transform the operation mode of medical and health institutions and introduce high-quality medical and health personnel. To actively enhance the adjustment and upgrading of hospital structure, promote the construction of medical infrastructure and improve the medical experience of patients. At the same time, the government needs to expand and adjust the consumption structure, and heighten the utilization rate of hospital capacity through demand-side reform.

(2) Encourage colleges and universities in provinces to add medical specialties. The Bureau of education ought to expand the scope of enrollment, increase the number of exchange and study places with medical schools in other provinces and cities. Then, setting up a reward mechanism, encouraging medical students to take high-quality medical talents as the goal. Additionally, hospitals should perfect the targeted internship

mechanism between medical students and public hospitals, enhance the sense of belonging and mission of medical staff to the hospital, and improve the capacity utilization of hospitals in the province with sustainable medical fresh blood investment.

(3) Support grassroots on-the-job doctors to improve their educational level and receive extended training. Hospitals should strengthen the training of grassroots medical and health personnel with general practitioners as the focus, further optimize the talent structure, and comprehensively improve the professional ethics, professional level and service ability of medical personnel.

This study did have some limitations that may have affected the results. The data used were incomplete, as the China Health Statistics Yearbook only provides data for healthcare institutions in 31 provinces up to 2019. Meantime, because the analysis covered 11 years of data, the analysis of short-term capacity utilization of healthcare institutions was not adequately specific compared to the long-term capacity utilization of healthcare institutions. Additionally, the undesirable output was introduced into the input-output model for the first time, and its logic needed to be improved. Therefore, more detailed studies are needed to confirm the present findings.

References

Arfa, C., Leleu, H., Goaïed, M., van Mosseveld, C., 2017. Measuring the capacity utilization of public district hospitals in Tunisia: Using dual data envelopment analysis approach. International Journal of Health Policy and Management, 6(1), 9–18.

Azad, E., Ketabi, S., Soltani, I., Bagherzade, M., 2012. Analysis of efficiency and resource allocation at different wards in Shariati hospital, Isfahan, Iran, using data envelopment analysis. Health Information Management, 8(7), 938-947.

Balezentis, T., Blancard, S., Shen, Z., Streimikiene, D., 2021. Analysis of environmental total factor productivity evolution in European agricultural sector. Decision Sciences, 52 (2), 483–511.

Boussemart J.-P., Ferrier G., Leleu H., Shen Z., (2020). An expanded decomposition of the luenberger productivity indicator with an application to the Chinese healthcare sector. Omega, 91, 102010.

Briec, W., Cavaignac, L., Kerstens, K., 2020. Input efficiency measures: A generalized, encompassing formulation. Operations Research, 68(6), 1836-1849.

Burdett, R., Kozan, E., 2016. A multi-criteria approach for hospital capacity analysis. European Journal of Operational Research, 255(2), 505–521.

Cesaroni, G., Kerstens, K., Van De Woestyne, I., 2017. A new input-oriented plant capacity notion: Definition and empirical comparison. Pacific Economic Review, 22(4), 720–739.

Cesaroni, G., Kerstens, K., Van De Woestyne, I., 2019. Short- and long-run plant capacity notions: Definitions and comparison. European Journal of Operational Research, 275(1), 387–397.

Cummins, D., Weiss, M., 2000. Analyzing firm performance in the insurance industry using frontier efficiency and productivity methods. In G. Dionne (Ed.), Handbook of insurance, 767–829. Boston: Kluwer.

Eriksson, C.O., Stoner, R.C., Eden, K.B., Newgard, C.D., Guise, J.M., 2017. The association between hospital capacity strain and inpatient outcomes in highly developed countries: A systematic review. Journal of General Internal Medicine, 32(6), 686–696.

Färe, R., 1984. The existence of plant capacity. International Economic Review, 25(1), 209–213.

Färe, R., Grosskopf, S., Kokkelenberg, E.C., 1989a. Measuring plant capacity, utilization and technical change: A nonparametric approach. International Economic Review, 30(3), 655–666.

Färe, R., Grosskopf, S., Lovell, C.A.K., 1994. Production frontiers. Cambridge: Cambridge University Press.

Färe, R., Grosskopf, S., Valdmanis, V., 1989b. Capacity, competition and efficiency in hospitals: A nonparametric approach. Journal of Productivity Analysis, 1(2), 123–138.

Fukuyama, H., Hashimoto, A., Weber, W.L., 2020. Environmental efficiency, energy efficiency and aggregate well-being of Japanese prefectures. Journal of Cleaner Production, 271, 122810.

Fukuyama, H., Liu, H., Song, Y., Yang, G., 2021. Measuring the capacity utilization of the 48 largest iron and steel enterprises in China. European Journal of Operational Research, 288(2), 648–665.

Fukuyama, H., Song, Y., Ren, X., Yang, G., 2022. Using a novel DEA-based model to investigate capacity utilization of Chinese firms. Omega, 106, 102534.

Ghaderi, H., Godarzi, A., Gohary, M., 2006. Determining of technical efficiency in Iran university of medical science's hospital by using DEA method: 2000-2004. Journal of Health Administration, 9(26):31-38.

Hollingsworth, B., 2003. Non-parametric and parametric applications measuring efficiency in health care. Health Care Management Science, 6(4), 203–218.

Hongyan, C., Shuai, X., Lifeng, Z., Roy, E., Welsch, B.L.P., Horn, J., 2018. The key techniques and future vision of feature selection in machine learning. Journal of Beijing University of Posts and Telecommunications. (CSCO) 41(1), 1–12. (in Chinese)

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), 417–441.

Jamasb, T., Pollitt, M., 2000. Benchmarking and regulation: international electricity experience. Utilities policy, 9(3), 107-130.

Johansen, L., 1968. Production functions and the concept of capacity. Recherches Recentes sur la Fonction de Production, Collection Economie mathematique et econometrie, 2, 52.

Kerstens, K., Sadeghi, J., Van de Woestyne, I., 2019. Convex and nonconvex input-oriented technical and economic capacity measures: An empirical comparison. European Journal of Operational Research, 276(2), 699–709.

Kerstens, K., Shen, Z., 2021. Using COVID-19 mortality to select among hospital plant capacity models: An exploratory empirical application to Hubei Province. Technological Forecasting and Social Change, 166, 120535.

Kerstens, K., Van de Woestyne, I., 2021. Cost functions are nonconvex in the outputs when the technology is nonconvex: Convexification is not harmless. Annals of Operations Research, 305(1–2), 81–106.

Kerstens, K., Sadeghi, J., Van de Woestyne, I., 2019. Plant capacity and attainability: Exploration and remedies. Operations Research, 67(4), 1135–1149.

Kuntz, L., Mennicken, R., Scholtes, S., 2015. Stress on the ward: Evidence of safety tipping points in hospitals. Management Science, 61(4), 754-771.

Langley, P., 1994. Selection of relevant features in machine learning. Proceedings of the AAAI Fall Symposium on Relevance, 184: 245-271.

Lapichino, G., Gattinoni, L., Radrizzani, D., Simini, B., Bertolini, G., Ferla, L., Mistraletti, G., Porta, F., Miranda, D.R., 2004. Volume of activity and occupancy rate in intensive care units. Association with mortality. Intensive Care Medicine, 30(2), 290–297.

Leleu, H., Arfa, C., Goaïed, M., J., 2006. Capacity utilization and efficiency of public district hospitals in Tunisian: Nonparametric approach. International Journal of Health Policy and Management, 6(1), 9-18.

Li, X., Chen, Z., 2010. Correctly using SPSS software for principal components analysis. Statistical Research. 27(08), 105–108. (in Chinese)

Mohamadi, E., Takian, A., Manesh, A. O., Majdzadeh, R., et al., 2020. Measuring efficiency of public hospitals in Iran: A comparative study using extended data envelopment analysis, 2012-2016.

Mitropoulos, P., Talias, MA., Mitropoulos, I., 2015. Combining stochastic DEA with Bayesian analysis to obtain statistical properties of the efficiency scores: An application to Greek public hospitals. European Journal of Operational Research, 243(1), 302–311.

Murty, S., Russell, R. R., Levkoff, S. B., 2012. On modeling pollution-generating technologies. Journal of environmental economics and management, 64(1), 117-135.

Pearson, K., 2010. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin philosophical magazine and journal of science, 2(11), 559-572.

Pelone, F., Kringos, D.S., Romaniello, A., Archibugi, M., Salsiri, C., Ricciardi, W., 2015. Primary care efficiency measurement using data envelopment analysis: A systematic review. Journal of Medical Systems, 39(1), 156.

Rosko, M.D., Mutter, R.L., 2008. Stochastic frontier analysis of hospital inefficiency. Medical Care Research and Review, 65(2), 131–166.

Shen Z., Balezentis T., Streimikis J., 2022. Capacity utilization and energy-related GHG emission in the European agriculture: A data envelopment analysis approach. Journal of Environmental Management, 318, 115517.

Shen Z., Valdmanis V., (2020). Identifying the contribution to hospital performance among Chinese regions by an aggregate directional distance function. Health Care Management Science, 23(1), 142–152.

Sherman, H. D., Ladino, G., 1995. Managing bank productivity using data envelopment analysis (DEA). Interfaces, 25(2), 60-73.

Sun, P., Wang B., 2019. Research on PCA dimension reduction with its application in machine learning. Journal of Hunan University of Technology, 33(1), 73–78. (in Chinese)

Valdmanis, V., Bernet, P., Moises, J., 2010. Hospital capacity, capability, and emergency preparedness. European Journal of Operational Research, 207(3), 1628–1634.

Yang, G., Fukuyama, H., 2018. Measuring the Chinese regional production potential using a generalized capacity utilization indicator. Omega, 76, 112-127.