May 2020

WORKING PAPER SERIES

2020-EQM-04

Using COVD-19 Mortality to Select Among Hospital Plant Capacity Models: An Empirical Application to the Hubei Province

Kristiaan Kerstens
IÉSEG School of Management and LEM-CNRS (UMR 9221), Lille, France

Zhiyang Shen
Beijing Institute of Technology, Beijing, China

IÉSEG School of Management Lille Catholic University 3, rue de la Digue F-59000 Lille
Tel: 33(0)3 20 54 58 92 www.ieseg.fr

Using COVID-19 Mortality to Select Among Hospital Plant Capacity Models: An Empirical Application to the Hubei Province

Kristiaan Kerstens‡

Zhiyang Shen§

Abstract:

This contribution defines short-run as well as long-run output- and input-oriented plant capacity measures evaluated relative to convex and nonconvex technologies. All these plant capacity concepts are used to measure the use of existing capacities as well as the evolution and build-up of extra hospital capacity in the province of Hubei in China during the outbreak of the COVID-19 epidemic in early 2020. The rather well-established fact from the medical literature that mortality rates increase with high capacity utilization rates is used to select the most plausible among these eight plant capacity concepts. It turns out that the relatively new input-oriented plant capacity concept correlates best with mortality.

Keywords: Data Envelopment Analysis; Free Disposal Hull; efficiency; plant capacity utilisation; mortality.

Declarations of interest: none.

‡ IESEG School of Management, CNRS, Univ. Lille, UMR 9221-LEM, 3 rue de la Digue, FR-59000 Lille, France.

§ Beijing Institute of Technology, 5 ZhongGuanCunNan Street, 100081 Beijing, China. Tel: +86-15210673725, Email: zhiyang86@163.com. Corresponding author.

30 May 2020

1. Introduction

The notion of output-oriented plant capacity has been informally defined into the economic literature by Johansen (1968, p. 362) as "... the maximum amount that can be produced per unit of time with existing plant and equipment, provided that the availability of variable factors of production is not restricted." Färe, Grosskopf and Kokkelenberg (1989) and Färe, Grosskopf and Valdmanis (1989) are the seminal articles providing a formal definition of the output-oriented plant capacity notion within a nonparametric frontier framework. These authors determine a measure of plant capacity utilisation from data on observed inputs and outputs using a couple of output-oriented efficiency measures. This has led to a series of empirical applications mainly in the sectors of fisheries (e.g., Felthoven (2002), Tingley and Pascoe (2005), among others) and health care (e.g., Magnussen and Rivers Mobley (1999), Karagiannis (2015), among others). Occasionally other sectors have been analysed: one study focuses on banking (Sahoo and Tone (2009)) and another describes a macroeconomic application on trade barriers (Badau (2015)). There have also been instances of some methodological refinements. For example, this plant capacity notion has been integrated in a decomposition of the Malmquist productivity index (e.g., De Borger and Kerstens (2000)). But all in all, it is safe to state that for more than two decades no major methodological innovations have occurred related to this plant capacity concept.

Then two major innovations happen one shortly after another. First, Cesaroni, Kerstens and Van de Woestyne (2017) define a new input-oriented concept of plant capacity utilisation based on a pair of input-oriented efficiency measures using the same nonparametric frontier framework. Second, Cesaroni, Kerstens and Van de Woestyne (2019) define new long-run output- and input-oriented plant capacity concepts that allow for changes in all input dimensions simultaneously rather than changes in the variable inputs solely. The plant capacity concepts focusing on changes in the variable inputs alone are then re-interpreted as short-run concepts. The combination of both innovations leads to four different plant capacity concepts that are now available to the empirical practitioner: on the one hand, output-oriented versus input-oriented plant capacity concepts, and on the other hand, short-run versus long-run plant capacity notions.

Furthermore, Kerstens, Sadeghi and Van de Woestyne (2019a) argue and empirically illustrate that the traditional output-oriented plant capacity utilization may be unrealistic since the amounts of variable inputs needed to reach the maximum capacity outputs may not be available at either the firm or industry levels. In response to this so-called attainability issue already indicated by Johansen (1968), Kerstens, Sadeghi and Van de Woestyne (2019a) define a new attainable output-oriented plant capacity utilization that puts a bound on the availability of variable inputs. Of course, the main problem is to define realistic bounds on this availability of variable inputs. Note that the whole issue of attainability also transposes to the long-run plant capacity concepts.

In view of these methodological doubts on the long-standing output-oriented plant capacity utilization notion, a first research question of this contribution is whether input-oriented plant capacity notions perform better or worse than output-oriented plant capacity concepts, and whether short-run plant capacity concepts perform better or worse than long-run plant capacity notions.

It is rather well known that the axiom of convexity has a potentially large impact on empirical analyses based on technologies. For instance, Walden and Tomberlin (2010) are probably the first study to empirically illustrate the effect of convexity on the output-oriented plant capacity notion. In a similar vein, Cesaroni, Kerstens and Van de Woestyne (2017) empirically compare output- and input-oriented plant capacity concepts and indicate the major influence of convexity on both concepts in practice. Finally, Kerstens, Sadeghi and Van de

Woestyne (2019a) also empirically illustrate the impact of convexity on both the traditional as well as the attainable output-oriented plant capacity notions.

However, most researchers tend to ignore the potential impact of convexity on the cost function. This is related to a property of the cost function in the outputs that is often ignored. Indeed, the cost function is nondecreasing and convex in the outputs when the technology is convex (see Jacobsen (1970) or Shephard (1970)): otherwise, the cost function is nonconvex in the outputs. Most empirical studies fail to put this property to a test. Kerstens, Sadeghi and Van de Woestyne (2019b) empirically compare the four different plant capacity concepts (output-oriented versus input-oriented, and short-run versus long-run) with a series of cost-based capacity utilization measures. Two key conclusions emerge. First, input-oriented plant capacity notions lend themselves overall more naturally to comparisons with cost-based capacity notions than output-oriented plant capacity concepts. Second, convexity makes a difference for both technical and economic capacity notions.

Thus, a second research question of this contribution is to further document the impact of convexity or nonconvexity on the empirical fit of the four different plant capacity concepts.

The empirical testing ground for our two main research questions is provided by the outbreak of the COVID-19 pandemic in the Chinese province of Hubei in late 2019 and early 2020. Faced with an unknown virus, the Chinese authorities faced a huge logistic challenge to efficiently use and improve to the extent possible the existing hospital capacity in the Hubei province to be able to adequately treat a surging number of patients. It is well-known from the medical literature that hospital capacity strain is associated with increased mortality and worsened health outcomes (see, e.g., the survey of Eriksson et al. (2017)). We use this known relation from the medical literature to shed light on our research questions from the economic literature as to which short-run plant capacity notions provide a better fit with the empirical data for this pandemic.

The Chinese authorities did not only face the challenge to optimally exploit existing hospital capacities, they also had to find ways to create new extra capacities using temporary makeshift hospitals to face the unknown surging demand for treatment. This build-up of new capacity requires an alternative modelling strategy: we are inclined to think that the long-run plant capacity concepts are particularly suitable to capture this creation of new hospital capacity. The COVID-19 pandemic offers a unique testing ground to see whether these long-run plant capacity concepts hold any water.

This contribution is structured as follows. In Section 2, we start with a literature review of plant capacity notions in the medical sector and we explore briefly the medical literature on the relation between capacity utilization and mortality. The next section starts with a definition of the technology and the efficiency measures needed for the definition of the four plant capacity notions at the center of our interest. Thereafter, detailed definitions of the output-oriented and input-oriented short-run and long-run plant capacity notions are offered. We end with a discussion of nonparametric frontier specifications to estimate the different plant capacity concepts. Section 4 discusses the data from the Hubei province in detail, since the quality of the data conditions our inferences. The next section provides empirical results. A final section concludes.

2. Hospital Plant Capacity and Mortality: A Brief and Candid Literature Review

2.1. Plant Capacity in Hospitals: Economic Literature

To our knowledge, there is a rather limited number of studies devoted to the analysis of plant capacity notions in the hospital sector. In chronological order, we start with the seminal article of Färe, Grosskopf and Valdmanis (1989) analyzing hospitals in Michigan. Magnussen and Rivers Mobley (1999) compare Norwegian and Californian hospitals, while Kerr et al. (1999) analyse Northern Irish acute hospitals. Valdmanis, Kumanarayake and Lertiendumrong

(2004) focus on plant capacity in Thai public hospitals. Valdmanis, Bernet, and Moises (2010) compute state-wide hospital capacity in Florida. Karagiannis (2015) analyses Greek public hospitals, and Valdmanis, DeNicola and Bernet (2015) report on Florida's public health departments. Finally, Arfa et al (2017) report findings for public hospitals in Tunesia. These eight studies are somewhat further analysed for our purposes below.

There are also some methodological variations available in the literature. For instance, Kang and Kim (2015) develop a cost-based frontier capacity notion for regional public hospitals in South Korea. Furthermore, Arfa et al (2017) develop a dual approach to the traditional output-oriented plant capacity notion that includes information on relative shadow prices of certain inputs. Finally, Valdmanis, DeNicola and Bernet (2015) also list bootstrapped plant capacity results to avoid bias from single point estimates.

One major critical methodological issue in the eight studies listed above is the choice of returns to scale assumption when defining the frontier technology. Though it is nowhere implied in the informal definition of Johansen (1968), the two seminal articles of Färe, Grosskopf and Kokkelenberg (1989) and Färe, Grosskopf and Valdmanis (1989) impose constant returns to scale on the technology. This example is followed by the following three studies that solely report plant capacity under constant returns to scale: Kerr et al. (1999), Valdmanis, Kumanarayake and Lertiendumrong (2004), and Valdmanis, Bernet, and Moises (2010).

However, constant returns to scale presupposes that the hospital sector is in long run zero profit competitive equilibrium. This is an unlikely assumption for any sector in general (see Scarf (1994)).[1] Furthermore, there is overwhelming evidence that there are increasing

[1] Scarf (1994, pp. 114–115) relentlessly criticizes the possibility of a constant returns to scale technology as follows: "Both linear programming and the Walrasian model of equilibrium make the fundamental assumption that the production possibility set displays constant or decreasing returns to scale; that there are no economies associated with production at a high scale. I find this an absurd assumption, contradicted by the most casual of observations. Taken literally, the assumption of constant returns to scale in production implies that if technical knowledge were universally available we could all trade only in factors of production, and assemble in our own

returns to scale and economies of scale in the hospital sector at large (see the survey of Giancotti, Guglielmo and Mauro (2017)). This explains the phenomenon of merging hospitals and policies aimed at expanding larger hospitals and restructuring/closing smaller hospitals. Therefore, in our analysis we consistently impose flexible or variable returns to scale on the frontier specifications of the technology in line with the informal definition of Johansen (1968), and with the remaining four studies reported above.

All of the above eight studies employ the short-run output-oriented plant capacity notion. Furthermore, all eight studies maintain the axiom of convexity on technology. Thus, we are the first study to analyse the short-run input-oriented plant capacity notions as well as both long-run plant capacity concepts in the hospital sector. Furthermore, we are the first study testing for the impact of convexity on plant capacity measurement in the hospital sector.

2.2. Hospital Capacity and Mortality: Economic and Medical Literature

There is a huge literature applying efficiency and productivity analysis using frontier technologies on hospitals and other medical care facilities (see, e.g., the surveys in Hollingsworth (2003), Pelone et al. (2015), Rosko and Mutter (2011)). While some studies control for quality of care and mortality, we find little conclusive evidence related to the relation between efficiency and productivity and their components on the one hand, and quality of care and mortality on the other hand.

In the wider economic and operations management literature we find again little clear-cut evidence on the relation between healthcare operational decisions and mortality (see Singh, Scholtes and Terwiesch (2019) for a recent survey). One study finding some evidence using

---

backyards all of the manufactured goods whose services we would like to consume. If I want an automobile at a specified future date, I would purchase steel, glass, rubber, electrical wiring and tools, hire labor of a variety of skills on a part — time basis, and simply make the automobile myself. I would grow my own food, cut and sew my own clothing, build my own computer chips and assemble and disassemble my own international communication system whenever I need to make a telephone call, without any loss of efficiency. Notwithstanding the analysis offered by Adam Smith more than two centuries ago, I would manufacture pins as I needed them."

department level bed occupancy rates is found in Kuntz, Mennicken and Scholtes (2015) who document the existence at the hospital level of a highly nonlinear effect of occupancy on mortality. These authors determine tipping points after which mortality increases rapidly when occupancy levels are further increased.

In the medical literature, there seems to be somewhat more substantial evidence that mortality is strongly correlated with high capacity utilization and high occupancy rates. This is the case at the level of individual diseases (e.g., Ross et al. (2010)), at the level of departments (e.g., Iapichino et al. (2004) for intensive care units), and at the hospital level (e.g., Madsen, Ladelund and Linneberg (2014)). Despite the heterogeneity in measures of capacity strain applied to intensive care units (ICU) and in non-ICU settings, the systematic review of Eriksson et al. (2017) finds that hospital capacity strain in highly developed countries is associated with increased patient mortality in 9 of 12 studies in ICU settings and in 18 of 30 studies overall. It also reports other worsened health outcomes. Overall, we find sufficiently robust medical evidence to expect a positive relation between capacity utilization and mortality.

Thus, we use ex post the real data from the COVID-19 pandemic as it has developed in the Hubei province in China in early 2020 to test for the relation between mortality and the measures of plant capacity utilization levels of eight models in total: both short-run and long-run output-oriented and input-oriented plant capacity notions are employed under both convexity and nonconvexity.

We briefly report on two somewhat related approaches in the literature. First, the recent contribution of Moghadas et al. (2020) use an epidemiological model to simulate the COVID-19 outbreak in the United States and how it can gravely challenge the ICU capacity, thereby exacerbating case fatality rates. Policies encouraging self-isolation may delay the epidemic peak, giving a time window to mobilize an expansion of hospital capacity. Our approach is not ex ante, but proposes an ex post analysis of the compatibility of mortality with the frontier-based plant capacity utilization measures.

Second, within the frontier literature, Valdmanis, Bernet, and Moises (2010) compute short-run output-oriented plant capacity at the hospital level for the state of Florida based on the whole population as part of an emergency preparedness plan. Starting from a scenario involving patient evacuations from Miami due to a major hurricane event, they assess whether hospitals in the proximity to the affected market can absorb the excess patient flow. This scenario analysis is not based on the data of any real emergency and therefore does not provide a valid test for the models.

We can now turn to a detailed discussion of the methodological framework employed in our contribution to measure the different models of plant capacity utilization.

3. Methodology

3.1. Definition of Production Technology

In this section, we introduce some basic notations and define the hospital production technology. The axiomatic production theory introduced by Debreu (1951), Koopmans (1951), Farrell (1957), and Shephard (1970) considers homogenous observed units to determine the shape of the production possibility set while maintaining some minimal set of production assumptions.

Assume a multiple-input, multiple-output production technology under which DMUs consume N types of inputs (*x*) to produce M types of outputs (*y*). The production possibility set or production technology T is given by:

$$T = \left\{ (x, y) \in \mathbb{R}_{+}^{N+M} : x \text{ } can \text{ } produce \text{ } y \right\} \qquad (1)$$

It is necessary to impose the following regularity conditions on the input and output data (see Färe, Grosskopf and Lovell (1994: p. 44-45)): (i) each producer uses nonnegative amounts of each input to produce nonnegative amounts of each output; (ii) there is an aggregate production of positive amounts of every output, and an aggregate utilisation of positive amounts of every input; and (iii) each producer employs a positive amount of at least one input to produce a positive amount of at least one output.

The production technology can also be represented by an output set $P(x)$ which indicates all possible output combinations that can be produced by at most a given level of inputs:

$$P(x)=\left\{y\in\mathbb{R}_{+}^{M}:(x,y)\in T\right\} \quad (2)$$

Alternatively, this same technology can also be represented by an input set $L(y)$ which denotes all possible input combinations that can produce at least a given level of outputs. The input correspondence is therefore formally defined as follows:

$$L(y)=\left\{x\in\mathbb{R}_{+}^{N}:(x,y)\in T\right\} \quad (3)$$

In particular, the technology also satisfies several widely adopted economic assumptions. These general axioms are usually imposed on the production possibility set (Shephard, 1970) as follows:

$$\begin{aligned}
&A_1: (0,0)\in T \text{ and if } (0,y)\in T \text{ then } y=0.\\
&A_2: T \text{ is closed.}\\
&A_3: \text{For each input } x\in\mathbb{R}_{+}^{N},\ T \text{ is bounded.} \qquad\qquad (4)\\
&A_4: \text{If } (x,y)\in T, \text{ then } (\tilde{x},\tilde{y})\in T \text{ for all } (-\tilde{x},\tilde{y})\le(-x,y).\\
&A_5: T \text{ is convex.}
\end{aligned}$$

Assumption A_1 implies that inactivity is feasible and, conversely, that there is no free lunch (i.e., outputs cannot be generated without inputs). Assumption A_2 states that unlimited quantities of outputs cannot be produced from finite quantities of inputs, while assumption A_3 implies that production plans located on the efficient frontier belong to the technology. Assumption A_4 implies free (strong) disposability of inputs and outputs: given outputs can be produced from more inputs than necessary, or given inputs can produce less outputs than currently. Assumption A_5 requires a convex production technology. More detailed discussions are available in, e.g., Hackman (2008).

We sometimes adopt the assumption that the technology is convex. However, we explicitly test for the validity of this assumption. Thus, not all of these axioms are simultaneously maintained in the empirical analysis.[2] Furthermore, note that we do not add a specific returns to scale assumption: this amounts to a flexible or variable returns to scale hypothesis.

In the short-run inputs can be grouped into fixed and variable parts: $x = (x^f, x^v)$ with $N = N_v + N_f$. The fixed part indicates the inputs cannot be varied in a short period and it is denoted by $x^f \in \mathbb{R}_+^{N_f}$, while the variable part can vary in relation with the quantity of outputs produced and it is denoted by $x^v \in \mathbb{R}_+^{N_v}$.

Following up on Färe, Grosskopf and Valdmanis (1989: p. 127), we define a short run technology $T^f = \{(x^f,y)$: there exists some x^v such that (x^f, x^v) can produce at least $y\}$ and the corresponding input set $L^f(y) = \{x^f : (x^f,y) \in T^f\}$ and output set $P^f(x^f) = \{y : (x^f,y) \in T^f\}$. This distinction between fixed and variable inputs leads to a sharpening of the conditions on the input and output data. Färe, Grosskopf and Kokkelenberg (1989: p. 659-660) state: each fixed input is used by some producer, and each producer uses some fixed input. We also need: each variable input is used by some producer, and each producer uses some variable input. Furthermore, the output set $P = \{y : \exists x : (x, y) \in T\}$ denotes the set of all possible outputs regardless of the needed inputs. Finally, $L(0) = \{x : (x,0) \in T\}$ is the input set compatible with a

[2] E.g., the nonparametric convex strongly disposable technology with variable returns to scale does not satisfy inaction: see also infra.

zero output level. See Cesaroni, Kerstens and Van de Woestyne (2019) for more details on these special technology definitions.

3.2. Distance Functions and Efficiency Measures

Distance functions provide an equivalent representation of production technologies and position observations with respect to the boundary of production possibilities sets. When an observation is on the boundary of technology, then it is technically efficient. However, if an observation is positioned below the boundary of technology, then it is technically inefficient and its performance can be improved.

To improve the technical efficiency of a production activity, there are two traditional ways of measuring: one can maximize outputs for given inputs, or one can minimize the inputs for given outputs. Maximizing output efficiency leads to a revenue interpretation, while minimizing input efficiency yields a cost interpretation (see, e.g., Hackman (2008)). Distance functions are related to efficiency measures. In the remainder of this contribution, we focus on output- and input-oriented efficiency measures.

Following Shephard (1970), the radial output efficiency measure is formulated as:

$$DF_{output}(x,y) = \max\{ \theta \in \mathbb{R}_+ : \theta y \in P(x)\} \qquad (5)$$

where θ is the technical efficiency measure. It indicates the maximum proportional expansion of outputs that can be achieved at a given level of inputs. This score is larger than or equal to unity ($DF_{output}(x,y) \geq 1$): an efficient DMU is located on the production frontier ($DF_{output}(x,y) = 1$); and an inefficient unit is situated in the interior of the production possibility set ($DF_{output}(x,y) > 1$).

Similarly, the radial input efficiency measure can be defined as:

$$DF_{input}(x,y) = \min\{ \lambda \in \mathbb{R}_+ : \lambda x \in L(y)\} \qquad (6)$$

where λ indicates the possible proportional decrease in inputs for a given level of outputs. This ratio is situated between zero and unity (0 < $DF_{input}(x,y)$ ≤ 1): the best practice is situated on the frontier ($DF_{input}(x,y)$ = 1); and an inefficient unit is found below the boundary of the input set (0 < $DF_{input}(x,y)$ < 1).

Denoting the radial output efficiency measure of the output set $P^f(x^f)$ by $DF_{output}^{f}\left(x^f, y\right)$, this efficiency measure can be defined as $DF_{output}^{f}\left(x^f, y\right) = \max\left\{\theta : \theta \geq 0, \theta y \in P^f(x^f)\right\}$. Next, we can denote $DF_{output}\left(y\right) = \max\left\{\theta : \theta \geq 0, \theta y \in P\right\}$. This new efficiency measure $DF_{output}\left(y\right)$ does not depend on a particular input vector x in contrast to the traditional radial output efficiency measure (5). Hence, this new measure is allowed to choose the level of inputs needed for maximizing θ.

In addition, we need the following particular definitions. First, we need a sub-vector input efficiency measure $DF_{input}^{SR}(x^f, x^v, y) = \min\left\{\lambda : \lambda \geq 0, (x^f, \lambda x^v) \in L(y)\right\}$ that only aims to reduce the variable inputs. Second, we need a similar sub-vector input efficiency measure $DF_{input}^{SR}(x^f, x^v, 0) = \min\left\{\lambda : \lambda \geq 0, (x^f, \lambda x^v) \in L(0)\right\}$ reducing variable inputs only but evaluated relative to this input set with a zero output level.

3.3. Short-Run Plant Capacity Utilization

Naturally, one can define the short-run output-oriented plant capacity utilization ($PCU_{output}^{SR}(x, x^f, y)$) by a ratio of output efficiency measures between a normal production technology ($DF_{output}(x,y)$) and an identical technology that has no constraints on use of variable inputs ($DF_{output}^{SR}(x^f, y)$) as in Färe, Grosskopf and Kokkelenberg (1989) and in Färe, Grosskopf and Valdmanis (1989):

$$PCU_{output}^{SR}(x, x^f, y) = \frac{DF_{output}(x, y)}{DF_{output}^{SR}(x^f, y)} \quad (7)$$

where $DF_{output}(x, y)$ and $DF_{output}^{SR}(x^f, y)$ are output efficiency measures relative to technologies including respectively excluding the variable inputs. Based on the approach introduced by Färe, Grosskopf and Kokkelenberg (1989), a short-run output-oriented decomposition can be obtained:

$$DF_{output}(x, y) = DF_{output}^{SR}(x^f, y)\ PCU_{output}^{SR}(x, x^f, y) \quad (8)$$

where $DF_{output}(x, y)$ can be decomposed into a biased plant capacity measure ($DF_{output}^{SR}(x^f, y)$) and an unbiased measure ($PCU_{output}^{SR}(x, x^f, y)$) according to the terminology introduced by Färe, Grosskopf and Kokkelenberg (1989), depending on whether one ignores inefficiency or adjusts for inefficiency (see also Cesaroni, Kerstens and Van de Woestyne (2019)). The unbiased measure is the ratio between the maximum possible quantity of outputs produced by a given level of inputs ($DF_{output}(x, y)$) and the maximum amount of outputs produced by a given level of fixed inputs but with any amount of variable inputs within the observed empirical range of the data ($DF_{output}^{SR}(x^f, y)$). Since $1 \leq DF_{output}(x, y) \leq DF_{output}^{SR}(x^f, y)$, thus $0 < PCU_{output}^{SR}(x, x^f, y) \leq 1$.

Following Cesaroni, Kerstens and Van de Woestyne (2017), one can define a short-run input-oriented plant capacity utilization ($PCU_{input}^{SR}(x, x^f, y)$) by a ratio of input efficiency measures evaluated relative to a production technology targeting to only reduce variable inputs ($DF_{input}^{SR}(x^f, x^v, y)$) and an identical technology with a level of null outputs ($DF_{input}^{SR}(x^f, x^v, 0)$):

$$PCU_{input}^{SR}(x, x^f, y) = \frac{DF_{input}^{SR}(x^f, x^v, y)}{DF_{input}^{SR}(x^f, x^v, 0)} \quad (9)$$

where $DF^{SR}_{input}(x^f, x^v, y)$ and $DF^{SR}_{input}(x^f, x^v, 0)$ are input efficiency measures aimed at reducing variable inputs for a given level of outputs or null outputs, respectively. Following Cesaroni, Kerstens and Van de Woestyne (2019), a short-run input-oriented decomposition can be given:

$$DF^{SR}_{input}(x^f, x^v, y) = DF^{SR}_{input}(x^f, x^v, 0)\ PCU^{SR}_{input}(x, x^f, y) \qquad (10)$$

where $DF^{SR}_{input}(x^f, x^v, 0)$ is a biased measure and $PCU^{SR}_{input}(x, x^f, y)$ is an unbiased measure of input-oriented plant capacity utilization. Similarly, this unbiased measure is the ratio between the minimum use of variable inputs for producing a given level of outputs ($DF^{SR}_{input}(x^f, x^v, y)$) and the minimum quantity of variable inputs for initiating the production process ($DF^{SR}_{input}(x^f, x^v, 0)$). Since $0 < DF^{SR}_{input}(x^f, x^v, 0) \leq DF^{SR}_{input}(x^f, x^v, y) \leq 1$, hence $PCU^{SR}_{input}(x, x^f, y) \geq 1$.

The combination of short-run output- and input-oriented and biased and unbiased plant capacity utilization yields four measures in total. These four measures of short-run plant capacity utilization are summarized in Table 1.

Table 1 Measurements of Short-Run Plant Capacity Utilization

	Measure	Notation	Interval
Output-oriented	Biased	$DF^{SR}_{output}(x^f, y)$	[1, +∞)
	Unbiased	$PCU^{SR}_{output}(x, x^f, y)$	(0,1]
Input-oriented	Biased	$DF^{SR}_{input}(x^f, x^v, 0)$	(0,1]
	Unbiased	$PCU^{SR}_{input}(x, x^f, y)$	[1, +∞)

3.4. Long-Run Plant Capacity Utilization

In the long-run, all inputs can be regarded as variable inputs as decision making units have sufficient time to adjust input utilizations. Thus, fixed and variable inputs need no longer to be treated differently.

Cesaroni, Kerstens and Van de Woestyne (2019) introduce a measure of long-run output-oriented plant capacity utilization given as:

$$PCU_{output}^{LR}(x,y) = \frac{DF_{output}(x,y)}{DF_{output}^{LR}(y)} \qquad (11)$$

where $DF_{output}(x,y)$ and $DF_{output}^{LR}(y)$ are output efficiency measures relative to a standard production technology and one without any constraints on the availability of inputs. Note that the numerator in (9) and (11) is identical. Cesaroni, Kerstens and Van de Woestyne (2019) propose a decomposition of this long-run output-oriented measure as follows:

$$DF_{output}(x,y) = DF_{output}^{LR}(y)\ PCU_{output}^{LR}(x,y) \quad (12)$$

where $DF_{output}^{LR}(y)$ and $PCU_{output}^{LR}(x,y)$ are biased and unbiased output-oriented measures of long-run plant capacity utilization, respectively. This unbiased measure is the ratio between the maximum possible quantity of outputs produced by a given level of inputs ($DF_{output}(x,y)$) and the maximum amount of outputs produced by any amount of inputs within the observed empirical range of the data ($DF_{output}^{LR}(y)$). Since $1 \leq DF_{output}(x,y) \leq DF_{output}^{LR}(y)$, $PCU_{output}^{LR}(x,y)$ is situated between 0 and unity.

Similarly, the long-run input-oriented measure of plant capacity utilization is defined by Cesaroni, Kerstens and Van de Woestyne (2019) as follows:

$$PCU_{input}^{LR}(x,y) = \frac{DF_{input}^{LR}(x,y)}{DF_{input}^{LR}(x,0)} \qquad (13)$$

where $D_{input}^{LR}(x,y)$ and $D_{input}^{LR}(x,0)$ are input efficiency measures estimated with a given level of outputs or at the level of null outputs, respectively. The decomposition of this long-run input-oriented measure is given as:

$$DF_{input}^{LR}(x,y) = DF_{input}^{LR}(x,0)\; PCU_{input}^{LR}(x,y) \quad (14)$$

where $DF_{input}^{LR}(x,0)$ and $PCU_{input}^{LR}(x,y)$ are biased and unbiased input-oriented measures of long-run plant capacity utilization, respectively. This unbiased measure is the ratio between the minimum possible use of inputs for a given level of outputs ($DF_{input}^{LR}(x,y)$) and the minimum usage of inputs to initiate the production process. Since $0 < DF_{input}^{LR}(x,0) \le DF_{input}^{LR}(x,y) \le 1$, $PCU_{input}^{LR}(x,y)$ is larger than unity.

The combination of long-run output- and input-oriented and biased and unbiased plant capacity utilization yields four measures in total. Four measures of long-run plant capacity utilization are summarized in Table 2.

Table 2 Measurements of Long-Run Plant Capacity Utilization

	Measure	Notation	Interval
Output-oriented	Biased	$DF_{output}^{LR}(y)$	[1, +∞)
	Unbiased	$PCU_{output}^{LR}(x,y)$	(0,1]
Input-oriented	Biased	$DF_{input}^{LR}(x,0)$	(0,1]
	Unbiased	$PCU_{input}^{LR}(x,y)$	[1, +∞)

For a graphical illustration of all short-run and long-run plant capacity notions, one can eventually consult Figures 1 to 4 in Cesaroni, Kerstens and Van de Woestyne (2019).

3.5. Nonparametric Frontier Estimation

We compute these plant capacity concepts using deterministic nonparametric frontier technologies.[3] The input-output vectors denoted by (x_k, y_k) are used to construct the empirical technology ($k = 1, \ldots, K$) under mainly the assumptions of strong input and output disposability, convexity, and flexible or variable returns to scale (see Hackman (2008)). The corresponding piece-wise linear frontier technology is then defined as:

$$T_{VRS}^{Convex} = \left\{ (x, y) : \sum_{k=1}^{K} z_k x_k \leq x, \sum_{k=1}^{K} z_k y_k \geq y, \sum_{k=1}^{K} z_k = 1, z \geq 0 \right\} \quad (15)$$

where z is the activity vector with non-negative elements. The convexity constraint ensures that linear combinations of the observed production plans are feasible. By relaxing the latter convexity assumption, one ends up with a nonconvex production frontier:

$$T_{VRS}^{Nonconvex} = \left\{ (x, y) : \sum_{k=1}^{K} z_k x_k \leq x, \sum_{k=1}^{K} z_k y_k \geq y, \sum_{k=1}^{K} z_k = 1, z \in \{0,1\} \right\} \quad (16)$$

where z is the activity vector with binary integer elements. We refer to Cesaroni, Kerstens and Van de Woestyne (2019) for all details on the underlying programming problems for computing the plant capacity measures in Tables 1 and 2 relative to technologies (15)-(16).

3.6. Conclusions

As already briefly alluded to in the introduction, the short-run plant capacity notions are used to assess the efficient use of existing hospital capacity in the Hubei province and their correlation with mortality is tested. The long-run plant capacity concepts are used to assess the build-up of new hospital capacity. We now turn to a discussion of the data we have available to implement these different plant capacity models.

[3] In fact, these plant capacity notions are difficult to estimate using traditional parametric specifications.

4. Data and Model Specifications

To analyze the plant capacity utilization for hospitals, we select the Hubei province in China as our sample. The Hubei province is the first region in China affected by the COVID-19 epidemic outbreak.[4] Several types of hospitals treat different patients according to their symptoms. At the individual level, each hospital has potentially some diversity in terms of staff and in terms of patients: production technologies are slightly heterogeneous. By defining the hospital production technology at the city level, we are better positioned to ensure the assumption of a homogenous production technology is valid. In Hubei province, 17 main cities are considered into our investigation: Wuhan, Huanggang, Xiaogan, Jingmen, Xianning, Jingzhou, Suizhou, Xiangyang, Shiyan, Ezhou, Huangshi, Yichang, Enshi, Xiantao, Tianmen, Qianjiang, and Shennongjia. We collect data from three main sources: the reader can consult the Appendix for all details regarding the sample. The sample covers eight weeks in the year 2020 from 19 January to 15 March during the COVID-19 epidemic.

Following Hollingsworth (2003), or Pelone et al. (2015), or Rosko and Mutter (2011), we define the hospital production technology at the city level by just two types of inputs and a single output. The two inputs in the hospitals are personnel and beds available for all patients. Personnel contains main medical staffs, such as licensed doctors, registered nurses, pharmacists and other technical staff. Beds are usually considered as a kind of capital stock for the hospital operations. Our single output is the number of COVID-19 patients. We also have information on the number of cured COVID-19 patients and the number of deaths from COVID-19. Averages for the two inputs and the single output as well as the cured and death patients are displayed in Table 3 for each of the eight weeks.

[4] The exact location of the outbreak remains controversial. The only certainty is that Hubei province is the area of the first large-scale transmission of the COVID-19 virus in China.

As population sizes and densities are relatively smaller in Tianmen, Qianjiang, and Shennongjia we combine their data to avoid the problem of zero output in the beginning of the observation period. A zero output violates the conditions on input and output matrices spelled out above. This reduces the number of cities analyzed from the original 17 to 15.

The single output is the number of patients who are infected by COVID-19 or which have similar symptoms. Note that diagnosing patients in these early weeks of the epidemic may have been difficult and likely some errors in classification have occurred. According to medical rules in China, all infected persons have to be inpatient. With no vaccine and no established curative treatment, the patient ends after a certain hospitalization period either as cured or as dead. The mortality rate is the ratio of COVID-19 deaths to the total number of COVID-19 patients.

Table 3. Average Inputs and Outputs (in 1000 persons, January-March 2020)

	Week	Personnel	Beds	Patients	Cured	Death
1	Jan 19 - Jan 25	710.47	371.72	0.96	0.04	0.05
2	Jan 26 - Feb 1	710.47	371.72	8.60	0.14	0.24
3	Feb 2 - Feb 8	710.47	394.82	24.79	1.26	0.49
4	Feb 9 - Feb 15	739.91	394.82	49.02	4.18	0.82
5	Feb 16 - Feb 22	739.91	394.82	45.99	9.67	0.75
6	Feb 23 - Feb 29	752.90	394.82	32.96	15.89	0.41
7	Mar 1 - Mar 7	752.90	394.82	19.71	13.82	0.23
8	Mar 8 - Mar 15	752.90	394.82	9.61	10.08	0.11

Note that it is well-known that the case definition adopted by the Chinese authorities has been initially narrow and it has gradually broadened to allow detection of more cases as knowledge increased (see Tsang et al. (2020)). However, to the best of our knowledge the data used in our study stick to the same case definition throughout the observation period.

Note furthermore that while personnel and beds are available for all patients, we have only information on COVID-19 patients in our single output. Thus, we have no information on the other patients under treatment during these eight weeks. Thus, we must assume that the proportion of COVID-19 patients to other patients is about the same for all cities observed at any given time period. Otherwise, our estimates of the various plant capacity notions may contain a bias. The aggregation at the city level mitigates any eventual deviations from this hypothesis at the underlying hospital level. It should also be borne in mind that all hospitals have in fact been obliged to follow very similar strategies in case of this medical emergency: separating COVID-19 and other patients, creating different logistic chains, canceling non-urgent interventions, dismissing patients to free up capacity, etc. (see Cao et al. (2020), Gagliano et al. (2020), among others).

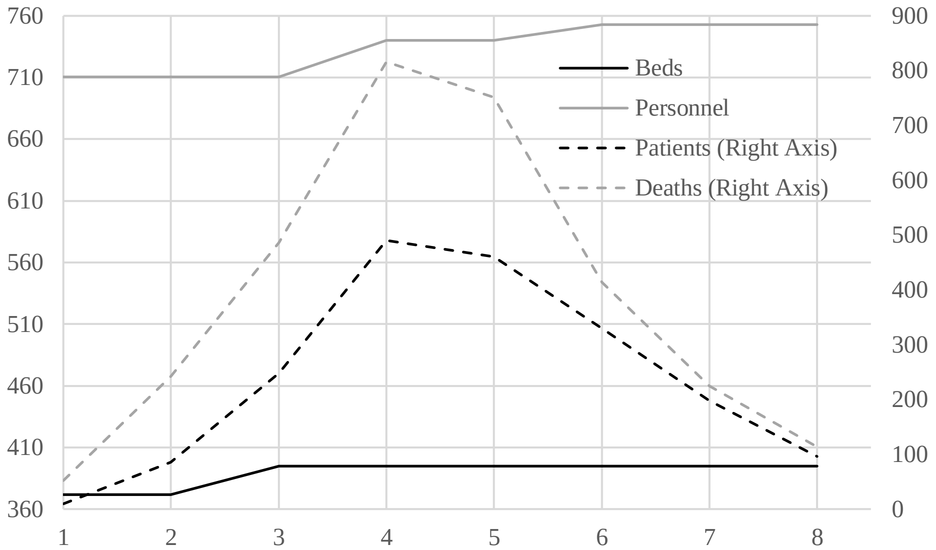
During an epidemic, it is important to exploit existing hospital capacities as good as possible, and if this capacity is insufficient to cope with peak demand, then it is crucial to build up extra capacity as soon as one possibly can. The short-run plant capacity concepts are suitable to model the exploitation of existing capacity. The long-run plant capacity notions are needed to capture the extension of existing capacity and the build-up of new capacity. New capacity is often based on makeshift (shelter) hospitals or the temporary conversion of existing buildings: see, e.g., Zhou et al. (2020) describing the conversion of schools and convention centers into hospitals. Hospital administrators and policy makers may face difficult choices to manage short-run and long-run decisions to optimize hospitals operations of existing and new capacities.

All of the eight studies employing the short-run output-oriented plant capacity notion specify fixed and variable inputs in a variety of ways. In our study, we decide on the choice of fixed and variable input pragmatically by looking at the evolution of both inputs over time. As shown in Figure 1, the number of beds remains constant in the initial two weeks only and starts to increase from the third week onwards as makeshift hospitals are put into use (e.g., Fire-God Mountain hospital and Thunder God Mountain hospital in Wuhan). The number of medical staff remains constant in the initial first three weeks and starts moving up from week 4 onwards. Thus, in our sample the beds are more variable than the medical staff. Furthermore, from week

4 onwards both inputs become variable and change in numbers: this is then clearly the long-run period. Therefore, the first three weeks are considered as the short-run period whereby beds are a variable input, and personnel is a fixed input.

Note that medical staff is often regarded as a fixed input since professional qualifications or certifications are often a prerequisite to be able to work. Since it is difficult to supplement medical staff in the short-run, the Chinese central government has been forced to transfer medical personnel from other provinces to Hubei to increase the supply in the fourth and sixth weeks. While we have information on different personnel qualification in the first three weeks (see above), we cannot differentiate the medical staff reinforcements. Therefore, we have to use aggregate personnel as a single fixed input. The number of bed expansions and personnel reinforcements are taken from XinhuaNet: this is an official media department of the Chinese central government. The detailed description of the data is available in the Appendix.

Figure 1. Evolution of Inputs, Output and Deaths over Time



Note: Unit of Beds and Personnel is in 1000 persons; units of patients and deaths are in 100 persons and person, respectively.

Finally, Figure 1 also shows that both numbers of COVID-19 patients and deaths increase rapidly to reach the turning point at week 4. Thereafter, a slow decline in patients and deaths can be observed while personnel keeps increasing and reaches a peak in week 6. This indicates that the situation of the epidemic has in fact improved before the long-run capacity has achieved the maximum level.

Finally, we have to specify the a priori relations between convex (C) and nonconvex (NC) plant capacity notions. Kerstens, Sadeghi and Van de Woestyne (2019: p. 704) specify in their Propositions 3.1 and 3.2 the relations between all biased and unbiased plant capacity concepts, respectively. For the biased plant capacity concepts, the C output-oriented ones are always larger than or equal to the NC ones, while the C input-oriented ones are always smaller than or equal to the NC ones. For the unbiased plant capacity concepts, the C output-oriented and input-oriented ones can be smaller, equal or larger than the NC ones: thus, there is no ranking possible.

Given the fact that we have only one output and one variable input in our sample, we must specify two more relations:

Proposition 1:

(a) Under a single variable input, $DF_{input}^{SR}(x^f, x^v, 0)$ is identical under C and NC.

(b) Under a single output, $DF_{output}^{LR}(y)$ is identical under C and NC.

Proof: Trivial: it suffices to look at the empirical results.

5. Empirical Results

In the short-run, 45 observations (15 cities over 3 weeks) are included in an intertemporal frontier estimation of short-run plant capacities. The descriptive statistics for these short-run plant capacity measures are listed in the first two parts of Table 4. The technical efficiency scores can be decomposed into biased and unbiased plant capacity measures following expressions (8), (10), (12), and (14). Technical inefficiency is very substantial, even under NC. For biased output-oriented measures of short-run plant capacity utilization, the average values are 164.30 and 67.55 under C and NC technologies, respectively. One can notice that the result of the biased input-oriented short-run plant capacity measures is 0.36 on average: following Proposition 1, it is identical for C and NC technologies. The average values of unbiased output-oriented (input-oriented) short-run plant capacity measures are 0.91 (1.30) and 0.80 (2.14) under C and NC technologies, respectively. These numbers are more modest because technical inefficiency has been eliminated.

Table 4 Descriptive for decomposition of plant capacity utilization

Technology	Convex	Nonconvex	Convex	Nonconvex	Convex	Nonconvex
Short-run Output-oriented	$DF_{output}(x,y)$		$DF_{output}^{SR}(x^f,y)$		$PCU_{output}^{SR}(x,x^f,y)$	
Mean	153.42	56.75	164.30	67.55	0.91	0.80
St. Dev.	662.85	210.81	665.78	221.93	0.20	0.27
Max	4451.82	1391.50	4451.82	1391.50	1.00	1.00
Min	1.00	1.00	1.00	1.00	0.30	0.28
Short-run Input-oriented	$DF_{input}^{SR}(x^f,x^v,y)$		$DF_{input}^{SR}(x^f,x^v,0)$		$PCU_{input}^{SR}(x,x^f,y)$	
Mean	0.39	0.47	0.36	0.36	1.30	2.14
St. Dev.	0.28	0.32	0.28	0.28	1.14	3.91
Max	1.00	1.00	1.00	1.00	8.40	21.12
Min	0.06	0.08	0.05	0.05	1.00	1.00
Long-run Output-oriented	$DF_{output}(x,y)$		$DF_{output}^{LR}(y)$		$PCU_{output}^{LR}(x,y)$	
Mean	92.60	36.65	1118.17	1118.17	0.17	0.09
St. Dev.	420.33	136.85	3984.48	3984.48	0.22	0.20
Max	4451.82	1391.50	35979.00	35979.00	1.00	1.00
Min	1.00	1.00	1.00	1.00	0.01	0.01
Long-run Input-oriented	$DF_{input}^{LR}(x,y)$		$DF_{input}^{LR}(x,0)$		$PCU_{input}^{LR}(x,y)$	
Mean	0.54	0.61	0.50	0.50	1.60	2.13
St. Dev.	0.30	0.32	0.31	0.31	2.62	3.82
Max	1.00	1.00	1.00	1.00	17.83	18.00
Min	0.07	0.08	0.05	0.06	1.00	1.00

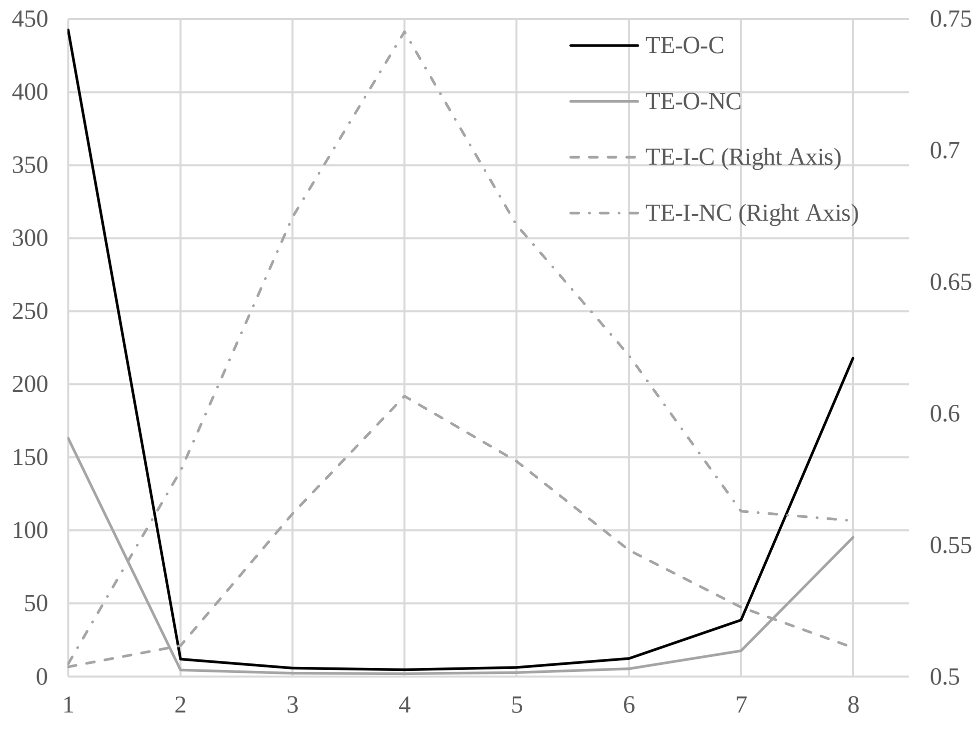
For the long-run plant capacity measures, the results are computed over the whole sample of 120 observation (15 cities over 8 weeks) using an intertemporal frontier. Descriptive statistics are listed in the last two parts of Table 4. Technical inefficiency is even more substantial now. Note that the results of biased long-run output-oriented plant capacity measures are identical under C and NC technologies: this follows from Proposition 1. Moreover, the results of biased long-run input-oriented plant capacity measures under C and NC technologies are approximately equal: this is a coincidence. The average values of unbiased output-oriented (input-oriented) long-run plant capacity measures are 0.17 (1.60) and 0.09 (2.13) under C and NC technologies, respectively: these values are smaller compared to the biased ones since technical inefficiency has been removed.

Overall, these descriptive statistics teach us that C and NC results differ substantially (as also reported in Walden and Tomberlin (2010), Cesaroni, Kerstens and Van de Woestyne (2017), and Kerstens, Sadeghi and Van de Woestyne (2019a)). Otherwise, there is little one can say regarding the pertinence of input-oriented vs. output-oriented and short-run vs. long-run plant capacity concepts: these seem to measure somewhat different realities. Since technical inefficiency is substantial, there is no point from now onwards to analyze biased plant capacity measures, because these are not cleaned from technical inefficiency.

Having discussed basic descriptive statistics, we now turn to the evolution of some of the above discussed elements over the course of the 8 weeks of the pandemic. The evolution of long run technical efficiency measures over time is presented in Figure 2 at the aggregate level of the province of Hubei. One can clearly observe that the output-oriented efficiency measures trace a U-shaped curve, while the input-oriented measures display an inverted U-shaped evolution. This means that output-oriented technical inefficiencies are decreasing at the beginning to remain close to the unit efficiency level between weeks 2-6, and then inefficiency increases again from week 7 onwards. The input-oriented technical efficiencies reveal an inverted U-shaped curve, but otherwise present a similar trend.

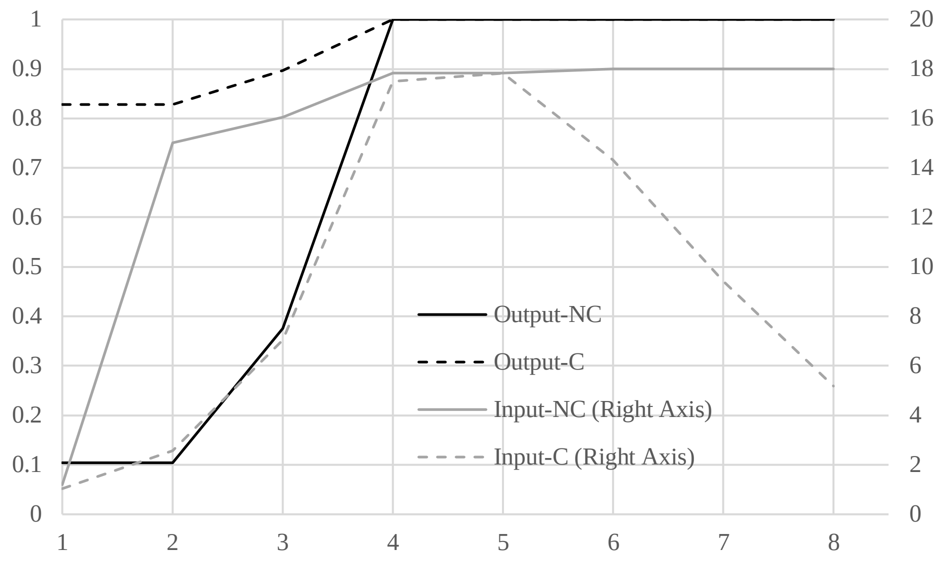
At the city level, we select Wuhan to investigate in some detail the evolution of long-run plant capacity over time. It is the city that has been most severely affected by COVID-19 in China. In Figure 3, one can observe that all unbiased long-run plant capacity measures are increasing in the beginning, and then keep constant from the fourth week onwards. Thus, full plant capacity utilization coincides with the peak in COVID-19 patients in week 4. But, only the long-run input-oriented plant capacity measure under a convex technology picks up the fact that the peak in patients precedes the final personnel reinforcements in week 6 and that capacity utilization thus in fact starts declining.

Figure 2 Evolution of Technical Efficiency Measures at the Aggregate Province



Note: TE-O denotes output-oriented long-run technical efficiency measure which is computed by $DF_{output}(x, y)$. TE-I denotes input-oriented long-run technical efficiency measure which is computed by $DF_{input}(x, y)$.

Figure 3 Unbiased Long-Run PCU Measures for Wuhan



To examine the impact of short-run and long-run measures of plant capacity utilization on the mortality rate, regression analysis is conducted. Firstly, the effect of short-run capacity measures is tested in a simple model with ordinary least square (OLS) since only three weeks of observations are available. The dependent and independent variables are the mortality rate due to COVID-19 on the one hand and technical efficiency or some plant capacity measure and a constant intercept on the other hand. Both variables are considered in logarithmic format. Note that three weeks sample normally contain 45 observations, but some cities with zero mortality rate at the beginning of the observation period are ignored: this results in 34 observations. The regression results are presented in the first two parts of Table 5.

As to technical inefficiency, we only establish a negative effect for the convex short-run input-oriented technical efficiency measure: the higher the technical efficiency, the lower the mortality. A significant positive sign between the mortality rate and the convex and nonconvex short-run unbiased input-oriented measure of plant capacity utilization is observed: higher plant capacity utilization increases mortality. This validates the conjecture from the medical literature. Moreover, the value of the R-square under a NC technology is marginally higher than that for the C approach.

Table 5 Relation between Mortality Rate, Technical Efficiency and Plant Capacity Utilization

Technology	Convex	Nonconvex	Convex	Nonconvex
Indep. Var.	$DF_{output}(x,y)$		$PCU^{SR}_{output}(x,x^f,y)$	
Observations	34	34	34	34
R^2	0.007	0.003	0.029	0.064
Coefficient	0.037	-0.025	0.410	-0.447
Constant	-4.684***	-4.538***	-4.539***	-4.712***
Indep. Var.	$DF^{SR}_{input}(x^f,x^v,y)$		$PCU^{SR}_{input}(x,x^f,y)$	
Observations	34	34	34	34
R^2	0.100	0.008	0.123	0.130
Coefficient	-0.375*	-0.102	1.089***	0.433**
Constant	-5.027***	-4.687***	-4.702***	-4.707***
Indep. Var.	$DF_{output}(x,y)$		$PCU^{LR}_{output}(x,y)$	
Observations	92	92	92	92
R^2	0.008	0.008	0.007	0.000
Coefficient	0.034	0.034	0.329	0.009
Constant	-4.594***	-4.567***	-3.706***	-4.475***
Indep. Var.	$DF^{LR}_{input}(x,y)$		$PCU^{LR}_{input}(x,y)$	
Observations	92	92	92	92
R^2	0.003	0.066	0.002	0.061
Coefficient	-0.109	-0.434**	-0.090	-0.407**
Constant	-4.589***	-4.798***	-4.493***	-4.417***

Note: mortality rate is independent variable; ***, **, and * denote 1%, 5% and 10% significance levels respectively.

In the long-run analysis, the effect of capacity measures on mortality rates is tested in a fixed effect panel model on the sample of 120 observation: again, cities with zero mortality rate at the start of the period are ignored, resulting in 92 observations. Regression results are reported in the third and fourth parts of Table 5. Again, a negative relation is detected between mortality rate and the NC input-oriented long-run technical efficiency measure: increasing technical efficiency lowers mortality. Contrary to the short-run result, now a significant negative effect is observed between the mortality rate and the NC long-run input-oriented measure of plant capacity utilization: a higher plant capacity utilization decreases mortality. This is probably related to the fact that the peak of the epidemic precedes the finalization of the new capacity build-up. This requires further exploration and ideally corroboration.

Overall, we can deduce the following conclusions from this regression analysis. First, input-oriented technical efficiency correlates with low mortality. This is in line with some of the findings on cost efficiency and mortality reported in Rosko and Mutter (2011). Second, higher short-run input-oriented plant capacity utilization rates seem to increase mortality, just as indicated in the medical literature. Third, such a positive relation is not found for the long-run input-oriented plant capacity utilization notion: this requires further research.

6. Conclusions

This contribution has started out by summarizing all known existing studies on the measurement of plant capacity in the hospital sector. Then, we have explored the evidence in the economic and medical literatures on the relation between capacity utilization and mortality. In the methodological sections, we have defined in great detail the short-run as well as the long-run output- and input-oriented plant capacity measures. These four plant capacity notions are evaluated relative to convex and nonconvex technologies: this yields eight different models.

All these plant capacity concepts are used to measure the evolution and build-up of hospital capacity in the province of Hubei in China during the outbreak of the COVID-19 epidemic in eight weeks during early 2020. After describing the limited data, all eight different models are computed for this limited sample. The fact that mortality rates increase with high capacity utilization rates is used to select the most plausible among these eight plant capacity concepts.

The empirical analysis has led to the following main conclusions. First, the descriptive statistics of technical efficiency and plant capacity measures reveal that C and NC results differ substantially (in line with earlier studies). Second, the regression analysis results show that input-oriented technical efficiency correlates with low mortality. Third, high levels of short-run input-oriented plant capacity utilization increase mortality, corroborating earlier findings in the medical literature. Overall, the relatively recent input-oriented plant capacity notions seem to challenge the much older output-oriented plant capacity concepts. Given the earlier doubts raised about the attainability of the traditional output-oriented plant capacity notions, this should lead the applied researcher to reflect more carefully about the proper choice of plant capacity concept.

Obviously, our study has a series of important limitations that need to be kept in mind and that may shape the agenda for future research. First, the sample is rather small: especially the three weeks available for computing the short-run concepts are very limited. Thus, testing of these same plant capacity notions on more substantial samples is being called for. Second, the data are imperfect in that we do not have the information on COVID-19 beds and COVID-19 personnel solely. Also, the absence of information on personnel categories of the reinforcements is most regrettable. Thus, more detailed studies are certainly necessary to corroborate the preliminary findings that we have come up with.

References:

Arfa, C., H. Leleu, M. Goaïed, C. van Mosseveld (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.

Badau, F. (2015) Ranking Trade Resistance Variables using Data Envelopment Analysis, *European Journal of Operational Research*, 247(3), 978-986.

Cao, Y., Q. Li, J. Chen, X. Guo, C. Miao, H. Yang, Z. Chen, C. Li, L. Li (2020) Hospital Emergency Management Plan During the COVID-19 Epidemic, *Academic Emergency Medicine*, 27(4), 309-311.

Cesaroni, G., K. Kerstens, I. Van de Woestyne (2017) A New Input-Oriented Plant Capacity Notion: Definition and Empirical Comparison, *Pacific Economic Review*, 22(4), 720-739.

Cesaroni, G., K. Kerstens, I. Van De Woestyne (2019) Short-and Long-Run Plant Capacity Notions: Definitions and Comparison, *European Journal of Operational Research*, 275(1), 387-397.

De Borger, B., K. Kerstens (2000) The Malmquist Productivity Index and Plant Capacity Utilization, *Scandinavian Journal of Economics*, 102(2), 303–310.

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

Eriksson, C.O., R.C. Stoner, K.B. Eden, C.D. Newgard, J.-M. Guise (2017) The Association Between Hospital Capacity Strain and Inpatient Outcomes in Highly Developed Countries: A Systematic Review, *Journal of General Internal Medicine*, 32(6), 686696.

Färe, R., S. Grosskopf, E. Kokkelenberg (1989) Measuring Plant Capacity, Utilization and Technical Change: A Nonparametric Approach, *International Economic Review*, 30(3), 655-666.

Färe, R., S. Grosskopf, V. Valdmanis (1989) Capacity, Competition and Efficiency in Hospitals: A Nonparametric Approach, *Journal of Productivity Analysis*, 1(2), 123-138.

Farrell, M. (1957) The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society*, 120A(3), 253-281.

Felthoven, R.G. (2002) Effects of the American Fisheries Act on Capacity, Utilization and Technical Efficiency, *Marine Resource Economics*, 17(3), 181-205.

Gagliano, A., P.G. Villani, F.M. Co', A. Manelli, S. Paglia, P.A.G. Bisagni, G.M. Perotti, E. Storti, M. Lombardo (2020) COVID-19 Epidemic in the Middle Province of Northern Italy: Impact, Logistics, and Strategy in the First Line Hospital, *Disaster Medicine and Public Health Preparedness*, forthcoming.

Giancotti, M., A. Guglielmo, M. Mauro (2017) Efficiency and Optimal Size of Hospitals: Results of A Systematic Search, *PLoS One*, 12(3), e0174533.

Hackman, S.T. (2008) *Production Economics: Integrating the Microeconomic and Engineering Perspectives*, Berlin, Springer.

Hollingsworth, B. (2003) Non-Parametric and Parametric Applications Measuring Efficiency in Health Care, *Health Care Management Science*, 6(4), 203-218.

Iapichino, G., L. Gattinoni, D. Radrizzani, B. Simini, G. Bertolini, L. Ferla, G. Mistraletti, F. Porta, D.R. Miranda (2004) Volume of Activity and Occupancy Rate in Intensive Care Units. Association with Mortality, *Intensive Care Medicine*, 30(2), 290–297.

Jacobsen, S. (1970) Production Correspondences, *Econometrica*, 38(5), 754–771.

Johansen, L. (1968) Production Functions and the Concept of Capacity, Namur, Recherches Récentes sur la Fonction de Production (Collection “Economie Mathématique et Econometrie”, n° 2) [reprinted in F.R. Førsund (ed.) (1987) *Collected Works of Leif Johansen, Volume 1*, Amsterdam, North Holland, 359-382].

Kang, S.-M., M.-H. Kim (2015) Short-Run Cost Minimization and Capacity Utilization of Regional Public Hospitals in South Korea, *Modern Economy*, 6(1), 21-29.

Karagiannis, R. (2015) A System-of-Equations Two-Stage DEA Approach for Explaining Capacity Utilization and Technical Efficiency, *Annals of Operations Research*, 227(1), 25-43.

Kerr, C.A., J.C. Glass, G.M. McCallion, D.G. McKillop (1999) Best-Practice Measures of Resource Utilization for Hospitals: A Useful Complement in Performance Assessment, *Public Administration*, 77(3) 639–650.

Kerstens, K., J. Sadeghi, I. Van de Woestyne (2019a) Plant Capacity and Attainability: Exploration and Remedies, *Operations Research*, 67(4), 1135–1149.

Kerstens, K., J. Sadeghi, I. Van de Woestyne (2019b) Convex and Nonconvex Input-Oriented Technical and Economic Capacity Measures: An Empirical Comparison, *European Journal of Operational Research*, 276(2), p. 699-709.

Koopmans, T.C. (1951) Analysis of Production as an Efficient Combination of Activities, in: T.C. Koopmans (ed) *Activity Analysis of Production and Allocation*, New Haven, Yale University Press, 33-97.

Kuntz, L., R. Mennicken, S. Scholtes (2015) Stress on the Ward: Evidence of Safety Tipping Points in Hospitals, *Management Science*, 61(4), 754-771.

Madsen, F., S. Ladelund, A. Linneberg (2014) High Levels of Bed Occupancy Associated with Increased Inpatient and Thirty-Day Hospital Mortality in Denmark, *Health Affairs*, 33(7), 1236-1244.

Magnussen, J., L. Rivers Mobley (1999) The Impact of Market Environment on Excess Capacity and the Cost of an Empty Hospital Bed, *International Journal of the Economics of Business*, 6(3), 383-398.

Moghadas, S.M., A. Shoukat, M.C. Fitzpatrick, C.R. Wells, P. Sah, A. Pandey, J.D. Sachs, Z. Wang, L.A. Meyers, B.H. Singer, A.P. Galvani (2020) Projecting Hospital Utilization During the COVID-19 Outbreaks in the United States, *Proceedings of the National Academy of Sciences*, 117(16), 9122-9126.

Pelone, F., D.S. Kringos, A. Romaniello, M. Archibugi, C. Salsiri, W. Ricciardi (2015) Primary Care Efficiency Measurement Using Data Envelopment Analysis: A Systematic Review, *Journal of Medical Systems*, 39(1), 156.

Rosko, M.D., R.L. Mutter (2011) What Have We Learned From the Application of Stochastic Frontier Analysis to U.S. Hospitals?, *Medical Care Research and Review*, 68(1S), 75S-100S.

Ross, J.S., S.-L.T. Normand, Y. Wang, D.T. Ko, J. Chen, E.E. Drye, P.S. Keenan, J.H. Lichtman, H. Bueno, G.C. Schreiner, H.M. Krumholz (2010) Hospital Volume and 30-Day Mortality for Three Common Medical Conditions, *New England Journal of Medicine*, 362(12), 1110-1118.

Sahoo, B.K., K. Tone (2009) Decomposing Capacity Utilization in Data Envelopment Analysis: An Application to Banks in India, *European Journal of Operational Research*, 195(2), 575-594.

Scarf, H. (1994) The Allocation of Resources in the Presence of Indivisibilities, *Journal of Economic Perspectives*, 8(4), 111–128.

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

Singh, D.K.C., S. Scholtes, C. Terwiesch (2019) Empirical Research in Healthcare Operations: Past Research, Present Understanding, and Future Opportunities, *Manufacturing & Service Operations Management*, 22(1), 73-83.

Tingley, D., S. Pascoe (2005) Factors Affecting Capacity Utilisation in English Channel Fisheries, *Journal of Agricultural Economics*, 56(2), 287-305.

Tsang, T.K., P. Wu, Y. Lin, E.H.Y. Lau, G.M. Leung, B.J. Cowling (2020) Effect of Changing Case Definitions for COVID-19 on the Epidemic Curve and Transmission Parameters in Mainland China: A Modelling Study, *The Lancet Public Health*, 5(5), e289-e296.

Valdmanis, V., P. Bernet, J. Moises (2010) Hospital Capacity, Capability, and Emergency Preparedness, *European Journal of Operational Research*, 207(3), 1628-1634.

Valdmanis, V., A. DeNicola, P. Bernet (2015) Public Health Capacity in the Provision of Health Care Services, *Health Care Management Science*, 18(4), 475-482.

Valdmanis, V., L. Kumanarayake, J. Lertiendumrong (2004) Capacity in Thai Public Hospitals and the Production of Care for Poor and Nonpoor Patients, *Health Services Research*, 39(6p2), 2117-2134.

Walden, J.B., D. Tomberlin (2010) Estimating Fishing Vessel Capacity: A Comparison of Nonparametric Frontier Approaches, *Marine Resource Economics*, 25(1), 23-36.

Zhou, F., X. Gao, M. Li, Y. Zhang (2020) Shelter Hospital: Glimmers of Hope in Treating Coronavirus 2019, *Disaster Medicine and Public Health Preparedness*, forthcoming.

Appendix: Data Description (Online Supplement)

The original data of inputs are from Hubei Provincial Bureau of Statistics (2020) where each city releases its Statistical Bulletins of National Economic and Social Development. Two types of inputs are used in our estimation: personnel and beds in public hospitals. Personnel contains main medical staffs, licensed doctors, registered nurses, pharmacists and other technical staff. Beds are aggregate numbers of beds in public hospitals for each city. The initial value of inputs is obtained from the last record (version 2019) in Statistical Bulletins of National Economic and Social Development at the city level.

Both personnel and beds were expanded during the epidemic. Personnel reinforcements were mainly provided by the Chinese central government who sent experienced doctors, nurses, pharmacists and other technical staffs from other provinces to Hubei. The data on these reinforcements is available from XinhuaNet (2020). However, XinhuaNet only reports aggregate number of reinforcements for each city without a detailed personnel classification. Therefore, we have to use the aggregate number of personnel in our estimation. The variation in inputs is shown in Table 1A.

Table 1A. Inputs variation

City	Beds Initial	Beds Expanded	Personnel Initial	Personnel Reinforced
Wuhan	91.6	109.8	217.9	255.2
Huanggang	35.9	37.1	33.5	34.5
Xiaogan	22.8	23.4	46.1	46.7
Jingmen	16.9	16.9	35.0	35.2
Xianning	14.4	14.4	17.2	17.7
Jingzhou	30.3	30.9	59.0	59.5
Suizhou	11.4	11.4	23.5	23.8
Xiangyang	36.5	36.5	65.7	66.3
Shiyan	25.6	25.6	51.8	51.8
Ezhou	6.1	7.3	13.6	14.1
Yellowstone	15.7	17.0	34.4	34.7
Yichang	28.2	28.2	65.1	65.3
Enshi	22.7	22.7	13.8	13.9
Xiantao	5.2	5.2	15.5	15.7
Tianmen, Qianjiang, and Shennongjia	8.4	8.4	18.3	18.6

Furthermore, the output data are patients who are infected by COVID-19 or which have similar symptoms. The mortality rate of COVID-19 is computed by the ratio of death patients on total patients. The data on patients and death patients are both from the daily reports of the Health Commission of Hubei Province (2020).

References:

Health Commission of Hubei Province. (2020) Accessed on April 30, 2020: http://wjw.hubei.gov.cn/fbjd/xxgkml/sjfb/ (in Chinese).

Hubei Provincial Bureau of Statistics. (2020) Statistical Bulletins of National Economic and Social Development, accessed on 30 April 2020: http://tjj.hubei.gov.cn/tjsj/tjgb/ndtjgb/sztjgb/ (in Chinese)

XinhuaNet (2020) Series of reports on supporting the front line of Hubei prevention and control, Accessed on 30 April: http://www.xinhuanet.com/politics/ (in Chinese)