Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP05/2020

Robust efficiency analysis of public hospitals in Queensland, Australia

Bao Hoang Nguyen and Valentin Zelenyuk

Date: April 2020

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Robust efficiency analysis of public hospitals in Queensland, Australia

Bao Hoang Nguyen[*] Valentin Zelenyuk[†]

April 21, 2020

Abstract

In this study, we utilize various approaches for efficiency analysis to explore the state of efficiency of public hospitals in Queensland, Australia in the year 2016/17. Besides the traditional nonparametric approaches like DEA and FDH, we also use a more recent and very promising robust approach–order-α quantile frontier estimators (Aragon, Daouia, & Thomas-Agnan, 2005). Upon obtaining the individual estimates from various approaches, we also analyse performance on a more aggregate level – the level of Local Hospital Networks by using an aggregate efficiency measure constructed from the estimated individual efficiency scores. Our analysis suggests that the relatively low efficiency of some Local Hospital Networks in Queensland can be partially explained by the fact that the majority of their hospitals are small and located in remote areas.

Keywords: Hospital efficiency, Aggregate efficiency, DEA, FDH, Alpha frontier.
JEL Codes: C24, C61, I11, I18.

[*]School of Economics, University of Queensland, Brisbane, Qld 4072, Australia
[†]School of Economics and Centre for Efficiency and Productivity Analysis, University of Queensland, Brisbane, Qld 4072, Australia

1 Introduction

In Australia, the provision of free public hospital services is the responsibility of the state and territory governments. The management of public hospitals in states and territories is usually geographically based. Since the National Health Reform Agreement in 2012 (Council of Australian Governments, 2011), the governance of public hospitals in Australia has become more decentralised with the establishment of Local Hospital Networks. The Local Hospital Network is an independent statutory body established by each Australian state/territory government. Local Hospital Networks directly operate a group of public hospitals and are directly responsible for their performance.

In Queensland, Local Hospital Networks are known as Hospital and Health Services (HHSs). There are 16 HHSs in the state, of these 15 HHSs are geographically based, the remaining one is a specialist statewide HHS dedicated to caring for children and young people. Each HHS is independently and locally controlled by a Hospital and Health Board and operated by a Health Service Chief Executive. HHSs relate to Queensland Department of Health (Queensland Health) through a service agreement in which Queensland Health acts as a system manager who has responsibility for purchasing healthcare services to cover the healthcare needs of citizens as well as monitoring the performance of HHSs, while each HHS acts as a provider whose function is to deliver healthcare services to its local community.

Although state and territory governments are responsible for delivering public hospital services, funding for public hospitals is provided by both federal and state/territory governments based on taxes collected from all states/territories across Australia. In the year 2016/17, 50% of expenditure on public hospital services in Queensland came from the state government, while 40% of the expenditure was provided by the Australian government (Australian Institute of Health and Welfare [AIHW], 2018). Public hospitals are funded either via Activity Based Funding or a Block Funding model. In Queensland, 36 hospitals (predominantly large and urban hospitals) are funded by Activity Based Funding.[1] Meanwhile, 87 hospitals (mainly small and rural hospitals) are funded by Block Funding.

Public hospitals in Queensland are widely dispersed geographically with a relatively high proportion in regional and remote areas, which in part reflects the share of the state's population living outside the major cities and the obligation of the state government to provide equitable access to public hospital services for all residents. Public hospitals in

[1] Under Activity Based Funding, hospitals are reimbursed based on the number and the complexity of patient care episodes they provide. Hospitals receive a fixed rate for each episode, and the value of the fixed rate is determined by the DRG to which the episode belongs.

the state are also diverse in terms of size: 91 out of 123 hospitals have 50 beds or fewer, yet 19 out of 123 hospitals have more than 200 beds and account for 75% of Queensland's total hospital beds (AIHW, 2019).

As public hospitals are the key institution in the acute healthcare sector where the majority of healthcare expenditure occurs, improving hospital efficiency has been viewed as a fundamentally important means to contain healthcare costs in Australia.[2] In the study published in 2010, the Productivity Commission (PC, 2010) pointed out that the average inefficiency level of Australian hospitals is around 10% and they would decrease operating expenditures by about 7% if the inefficiency was eliminated. Given that the efficiency of public hospitals is an important issue of public concern and has now become the main responsibility of HHSs, it is crucial to analyse hospital performance at HHS level. These analyses will provide useful information about the relative performance of HHSs and possibly identify sources of efficiency differentials, which are imperatively needed for any plan to promote hospital efficiency.

This study will provide such an analysis by exploring the state of efficiency of public hospitals at the level of HHSs in Queensland, Australia in the year 2016/17. To analyse performance on the aggregate level, we utilize an aggregate efficiency measure constructed from individual efficiency scores estimated using various approaches. Besides the traditional nonparametric approaches like DEA and FDH, we also use a more recent and very promising robust approach–order-α quantile frontier estimators (Aragon, Daouia, & Thomas-Agnan, 2005). The order-α quantile frontier estimators appear to be more appealing than the conventional nonparametric approaches because they are more robust with respect to extreme values and/or outliers in a finite sample and do not suffer from the well-known curse of dimensionality (Simar & Wilson, 2013).

Based on the robust estimates of aggregate efficiency, we use k-mean clustering technique (an unsupervised machine learning algorithm) to classify HHSs in Queensland into three groups, namely relatively low, medium, and high efficiency. Moreover, our analysis also suggests that the relatively low efficiency of some HHSs in Queensland can be partially explained by the fact that the majority of their hospitals are small and located in remote areas.

Our paper is organized as follows. Section 2 presents theoretical frameworks for efficiency measures and their nonparametric estimators. Section 3 provides a description of the data sources and variables used in this study. Section 4 discusses the results, and

[2]In the fiscal year 2016/17, Australia spent $181 billion on healthcare (more than $7,400 per person and 10% of its GDP), about a 57% increase since 2006/07 (after adjusting for inflation). This turns out to be an average annual growth rate of 4.67% over the decade: around 2% higher than average growth of GDP (AIHW, 2018).

Section 5 provides concluding remarks.

2 Methodology

2.1 Theoretical concepts

Let us consider a production process in which a production unit uses a set of N inputs, denoted as $x = (x_1, ..., x_N)' \in \Re_+^N$, to produce a univariate output, denoted as $y \in \Re_+$.[3] According to the production theory (Shephard, 1953, 1970), the production technology can be characterized by a technology set defined as

$$\Psi = \left\{(x, y) \in \Re_+^N \times \Re_+ : x \text{ can produce } y\right\}. \tag{1}$$

Some regularity conditions are usually assumed for the technology set, among those the three most common assumptions are as follows [4]

A1. Ψ is closed.

A2. The output sets (defined in (2) below) are bounded, $\forall x \in \Re_+^N$.

A3. All inputs and outputs are strongly disposable, i.e, $(x_0, y_0) \in \Psi \Rightarrow (x, y) \in \Psi, \forall x \geqq x_0, y \leqq y_0$.

The production technology can also be described mathematically in terms of its sections: input requirement set and output attainable set. In this paper, we measure efficiency in output direction, thus our discussion here focuses on the output attainable set. It is defined as

$$P(x) = \{y \in \Re_+ : (x, y) \in \Psi\}, x \in \Re_+^N. \tag{2}$$

When efficiency is a concern, the boundary of the technology set is of interest. For the case of univariate output, the upper bound of the output attainable set (the production frontier) is also referred to as production function and defined as

$$\partial P(x) = \max_y \{y \mid y \in P(x)\}. \tag{3}$$

[3] For the cases of multiple-output, one can either follow the multivariate conditional quantile approach proposed by Daouia and Simar (2007) or utilize aggregation techniques to aggregate outputs. In this study, we adopt Daraio and Simar's (2007) approach (the approach based on Principal Component Analysis) to aggregate hospital outputs into a single output measure. An alternative approach would be to use a price-based aggregation approach (Zelenyuk, 2020).

[4] Other standard regularity conditions are "No Free Lunch" and "Producing Nothing is Possible" (see more details in Sickles & Zelenyuk, 2019)

The Farrell-type output oriented technical efficiency for the production unit is then defined as a radial distance from a point in output space representing the production unit toward the boundary and is defined mathematically as

$$\lambda(x, y) = \sup_{\lambda} \{\lambda > 0 \mid \lambda y \in P(x)\} = \sup_{\lambda} \{\lambda > 0 \mid (x, \lambda y) \in \Psi\}. \tag{4}$$

One might find it more convenient to look at the reciprocal of the output oriented efficiency (also known as the Shephard distance function) since it gives an efficiency measure between 0 and 1, where 1 stands for 100% efficiency.

Now let us look at a more aggregate level, consider a group, say group ℓ, of n_ℓ production units with input-output allocation being $\mathcal{X}^{\ell}_{n_\ell} = \left\{\left(X^{\ell}_i, Y^{\ell}_i\right) \mid i = 1, \ldots, n_\ell\right\}$. One can measure the efficiency of group ℓ by using the aggregate efficiency measure proposed by Färe and Zelenyuk (2003), extended by Simar and Zelenyuk (2007) and further elaborated on in Simar and Zelenyuk (2018). The main advantage of this measure is that it uses meaningful weights derived from the economic optimization principle to aggregate individual efficiency scores in order to construct a group measure (see detail in Färe & Zelenyuk, 2003). In the case of univariate output, the aggregate efficiency for group ℓ is the weighted average of individual efficiency scores, where weights are output shares of individual production units in the group and is defined as

$$\overline{TE}^{\ell}_{n_\ell} = \sum_{i=1}^{n_\ell} \lambda\left(X^{\ell}_i, Y^{\ell}_i\right) \times S^{\ell}_i, \quad S^{\ell}_i = \frac{Y^{\ell}_i}{\sum_{i=1}^{n_\ell} Y^{\ell}_i}. \tag{5}$$

2.2 Nonparametric estimators

2.2.1 DEA and FDH

In practice, Ψ is unknown and thus needs to be estimated from a sample of production units, $\mathcal{X}_n = \{(X_i, Y_i) \mid i = 1, \ldots, n\}$. There have been two widely-used approaches to estimate the production frontiers in the literature, usually referred to as the 'deterministic frontier models' and the 'stochastic frontier models'. The deterministic frontier models assume all observed production units belong to the technology set with probability one, whereas the stochastic frontier models allow some observations to be outside of the technology set by including two-sided random noise. The traditional stochastic frontier approach (SFA) requires parametric restrictions on the shape of the production frontier and on the data generating process to estimate the frontier and to identify the inefficiency term from the random noise component.[5] Recently, semiparametric and non-

[5]The traditional stochastic frontier approach was proposed independently by Aigner, Lovell, and Schmidt (1977) and Meeusen and van Den Broeck (1977).

parametric estimators have been developed for stochastic frontier models to mitigate the parameterization of the approach (see more details in Parmeter & Zelenyuk, 2019).

The deterministic frontier models appear to be more appealing because they are usually handled via nonparametric estimators and rely on less restrictive assumptions. The most flexible deterministic frontier model is the Free Disposal Hull (FDH) estimator introduced by Deprins, Simar, and Tulkens (1984), which requires only the strong disposability assumption on the technology set. If, in addition, one imposes the convexity assumption on the technology set, one can use the Data Envelopment Analysis (DEA) estimator, which was initiated by Farrell (1957) and popularized by Charnes, Cooper, and Rhodes (1978). For DEA models, one can further impose constant returns to scale (CRS) or variable returns to scale on the technology set to obtain CRS-DEA or VRS-DEA estimators (Färe, Grosskopf, & Logan, 1983; Banker, Charnes, & Cooper, 1984). The three estimators can be formulated respectively as follows

$$\widehat{\Psi}_{FDH} \equiv \left\{ (x,y) : y \leq \sum_{i=1}^{n} \zeta_i Y_i, x \geq \sum_{i=1}^{n} \zeta_i X_i, \sum_{i=1}^{n} \zeta_i = 1, \zeta_i \in \{0,1\}, i = 1, \ldots, n \right\}, \quad (6)$$

$$\widehat{\Psi}_{CRS-DEA} \equiv \left\{ (x,y) : y \leq \sum_{i=1}^{n} \zeta_i Y_i, x \geq \sum_{i=1}^{n} \zeta_i X_i, \zeta_i \geq 0, i = 1, \ldots, n \right\}, \quad (7)$$

$$\widehat{\Psi}_{VRS-DEA} \equiv \left\{ (x,y) : y \leq \sum_{i=1}^{n} \zeta_i Y_i, x \geq \sum_{i=1}^{n} \zeta_i X_i, \sum_{i=1}^{n} \zeta_i = 1, \zeta_i \geq 0, i = 1, \ldots, n \right\}. \quad (8)$$

The FDH/DEA estimators of technical efficiency are then obtained by plugging $\widehat{\Psi}_{FDH}$ or $\widehat{\Psi}_{CRS-DEA}$ or $\widehat{\Psi}_{VRS-DEA}$ in (4). The asymptotic properties of FDH/DEA estimators have been well-established in the literature (e.g., see Kneip, Park, & Simar, 1998; Park, Simar, & Weiner, 2000; Kneip, Simar, & Wilson, 2008; Park, Jeong, & Simar, 2010). In summary, under appropriate assumptions, the estimators are consistent (converging to the true values when sample sizes go to infinity) and have limiting distributions. Convergence rates depend on the type of estimators and the dimension of input-output space (the number of inputs, p, plus the number of outputs, q). To be more specific, the convergence rates for FDH, CRS-DEA, VRS-DEA estimators are n^{κ} , where $\kappa = 1/(p+q)$, $2/(p+q)$, or $2/(p+q+1)$, respectively (e.g., see more discussion in Simar & Wilson, 2015; Sickles & Zelenyuk, 2019).

2.2.2 Partial frontiers

The deterministic frontier models, however, are particularly sensitive to extreme values and/or outliers because by construction, they fully envelop all observed data. Various techniques have been proposed to deal with the disadvantage. One approach is to identify and possibly delete any outliers in the data, but the approach, to some extent, depends on how the researcher defines an 'outlier' (Simar & Wilson, 2015). As an alternative, one can also use the stochastic versions of DEA and FDH, where data is prewhitened from the noise and outliers using nonparametric SFA in the first stage and DEA/FDH is applied to estimate efficiency in the second stage (e.g., see Simar, 2007; Simar & Zelenyuk, 2011).

Another approach is to use robust partial frontier estimators. There are mainly two types of partial frontiers, which are: (i) order-m frontiers introduced by Cazals, Florens, and Simar (2002) and (ii) order-α quantile frontiers introduced by Aragon, Daouia, and Thomas-Agnan (2005) and extended by Daouia and Simar (2007). The idea of partial frontier estimators is to estimate something "close" to but not the boundary of the technology set (Simar & Wilson, 2013). For example, in output orientation, order-m frontiers are defined as the expected maximum obtainable outputs among m production units drawn randomly from the population of those using at most a given level of inputs. Meanwhile, order-α quantile frontiers represent the expected maximum output levels that are exceeded by $100\,(1-\alpha)\,\%$ of production units using less than or equal to a given level of inputs.

The nonparametric estimators of these frontiers turn out to be more appealing than the conventional deterministic frontier models because they do not suffer from the well-known curse of dimensionality and achieve the standard parametric root-n ($\sqrt{n}$) rate of convergence (Cazals, Florens, & Simar, 2002; Aragon, Daouia, & Thomas-Agnan, 2005; Daouia & Simar, 2007). Moreover, both the estimators are also consistent estimators of the full frontier and share asymptotic properties with FDH estimators but are more robust with respect to extreme values and/or outliers in finite sample than the conventional FDH or DEA estimators (Simar & Wilson, 2013).

Among the two above-mentioned partial frontier approaches, the order-α quantile frontier estimators are argued to have better robustness properties than the order-m frontier estimators. For example, Aragon, Daouia, and Thomas-Agnan (2005) compared the two estimators with various simulated data sets, and reached the same conclusion with all the data sets that the order-m frontier estimators are less resistant to outliers than the order-α quantile frontier estimators. Daouia and Simar (2007) examined the robustness properties of the two estimators from the theoretical points of view using the concept of influence function, and came up with the same conclusion. Thus, we will use the order-α quantile frontier estimators in our analysis and focus our discussion on these estimators.

Let us define the technology set Ψ as the support of the joint distribution of a random variable (X, Y), which generates the random sample $\mathcal{X}_n$. Here, we focus on the interior of the set, $\Psi^* = \{(x, y) \in \Psi | F_X(x) > 0\}$, where $F_X(\cdot)$ represents the marginal distribution of X. As in Cazals, Florens, and Simar (2002), the production function defined in (3) can be rewritten in a probabilistic representation as

$$\partial P(x) = \sup_y \left\{ y \;\middle|\; F_{Y|X}(y|x) < 1 \right\}, \tag{9}$$

where $F_{Y|X}(y|x)$ is the conditional distribution of Y given $X \leq x$, i.e.,

$$F_{Y|X}(y|x) = \frac{F_{XY}(x, y)}{F_X(x)}, \tag{10}$$

where $F_{XY}(x, y) = \text{Prob}(X \leq x, Y \leq y)$ is the joint distribution of (X, Y).

Equivalently, $\partial P(x)$ can be formulated as the order one quantile of the distribution of Y given $X \leq x$ as

$$q_1(x) = \inf_y \left\{ y \geq 0 \;\middle|\; F_{Y|X}(y|x) = 1 \right\}. \tag{11}$$

One can interpret $q_1(x)$ as the minimum output level not exceeded by any production unit using at most x inputs. Based on the formulation, Aragon, Daouia, and Thomas-Agnan (2005) introduced a concept of order-α quantile frontiers as the quantile functions of order α, $\alpha \in [0, 1]$, of the distribution of Y given that X is less than or equal to a given level of inputs and defined as

$$q_\alpha(x) = \inf_y \left\{ y \geq 0 \;\middle|\; F_{Y|X}(y|x) \geq \alpha \right\}. \tag{12}$$

The order-α quantile frontier, $q_\alpha(x)$, represents the output threshold exceeded by $100(1-\alpha)\,\%$ of production units using at most x inputs. The efficiency measure with respect to the frontier is referred to as the order-α quantile efficiency and defined as

$$\lambda_\alpha(x, y) = \inf_\lambda \left\{ \lambda \;\middle|\; F_{Y|X}(\lambda y|x) \geq \alpha \right\}. \tag{13}$$

The order-α quantile efficiency represents the radial distance from a point in output space representing the production unit toward the order-α quantile frontier. The measure $\lambda_\alpha(x, y)$ can have values between 0 and $+\infty$, where $\lambda_\alpha(x, y) < 1$ indicates that the production unit with input-output allocation (x, y) is above the order-α quantile frontier (i.e., super-efficient production unit).

To estimate order-α quantile frontiers and order-α quantile efficiency, we can again apply the plug-in principle by replacing $F_{Y|X}(\cdot|\cdot)$ in (12) and (13), respectively, by its empirical analogue

$$\widehat{q}_{\alpha,n}(x) = \inf_y \left\{ y \geq 0 \;\middle|\; \widehat{F}_{Y|X}(y|x) \geq \alpha \right\} \tag{14}$$

and

$$\widehat{\lambda}_{\alpha,n}(x,y) = \inf_{\lambda} \left\{ \lambda \;\middle|\; \widehat{F}_{Y|X}(\lambda y|x) \geq \alpha \right\}. \tag{15}$$

As an extension of Theorem 4.1 in Aragon, Daouia, and Thomas-Agnan (2005), Daouia and Simar (2007) show that under appropriate assumptions, order-α quantile efficiency estimators have asymptotic normality with the standard parametric root-n ($\sqrt{n}$) rate of convergence. Mathematically, the result is stated in Theorem 3.2 in Daouia and Simar (2007) as follows

Theorem 2.1. *Let $0 < \alpha < 1$ be a fixed order and let $(x,y) \in \Psi$ be a fixed production unit in interior of Ψ. Assume that $G(\lambda) = 1 - F_{Y|X}(\lambda y|x)$ is differentiable at $\lambda_\alpha(x,y)$ with negative derivative $G'(\lambda_\alpha(x,y))$. Then,*

$$\sqrt{n}\left(\widehat{\lambda}_{\alpha,n}(x,y) - \lambda_\alpha(x,y)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_\alpha(x,y)\right) \qquad \text{as } n \to \infty, \tag{16}$$

where $\sigma^2_\alpha(x,y) = \alpha(1-\alpha) / [G'(\lambda_\alpha(x,y))]^2 F_X(x)$.

Moreover, the order-α quantile efficiency estimators converge to the FDH estimator as $\alpha \to 1$

$$\lim_{\alpha \to 1} \widehat{\lambda}_{\alpha,n}(x,y) = \widehat{\lambda}_{FDH}(x,y) \tag{17}$$

More details on this interesting method can be found in Aragon, Daouia, and Thomas-Agnan (2005) and Daouia and Simar (2007), while in the next section we will apply it to analyse the efficiency of public hospitals in Queensland, Australia.

3 Variables and Data

In this study, we compare the technical efficiency of public hospitals across 15 geographically based HHSs in Queensland in the financial year 2016/17.[6] Our sample includes 111 public acute hospitals.[7] The hospital data are sourced from two data collections of Queensland Health, namely Financial and Residential Activity Collection (FRAC) and Monthly Activity Collection (MAC). We obtained the information about hospital staffing and drug, surgical and medical supply expenditures from the FRAC, while the MAC

[6]There are 16 HHSs in Queensland, but only 15 HHSs directly manage and operate public hospitals in defined local geographical areas, the remaining HHS is a specialist statewide HHS dedicated to caring for children and young people from across Queensland.

[7]Public hospitals in Queensland include acute hospitals, mixed sub- and non-acute hospitals, early parenting centres, women's and children's hospitals, and psychiatric hospitals. We only consider public acute hospitals, which account for more than 90% of inpatient cases treated. Our sample does not include hospitals that were just opened in 2017 and hospitals that are not operated by a HHS.

provided us with the data on the number of beds, non-admitted patient activities, and admitted patient episodes of care by diagnosis related groups (DRGs).

Following the common practice in the literature on performance analysis of hospitals [8], to model the production process of hospitals, we use full-time equivalent (FTE) staff as a proxy for labour input, the number of bed (BEDS) as a proxy for capital input, and drug, surgical and medical supply expenditure (DMSEXP) as a proxy for consumable input (see more detailed discussion about the selection and construction of hospital inputs and outputs in Grosskopf, Nguyen, Yong, & Zelenyuk, 2020). Regarding hospital staffing, we adopt Daraio and Simar's (2007) approach (the approach based on Principal Component Analysis) to aggregate six labour categories into a single measure of labour input, called labour factor (FLABOUR).[9] The aggregation approach helps to increase the discriminant power of the nonparametric envelopment estimator, but still covers the information contained in all the labour categories.

Similarly, using Daraio and Simar's (2007) approach, we aggregate two widely-used measures of hospital outputs, namely (i) non-admitted occasions of services and (ii) casemix weighted inpatient episodes, into a single output measure, called output factor (we denote as FOUTPUT). The information about non-admitted occasions of services, which include both outpatient visits and emergency department services, is readily available in our datasets. Meanwhile, the casemix weighted inpatient episode is constructed as the weighted sum of the number of inpatient episodes by DRG, where the weight is the inlier DRG cost weight obtained from the Independent Hospital Pricing Authority.[10]

In addition, we obtain information about hospital peer groups and geographic location from AIHW (2015). Based on hospital peer groups, in our study, hospitals are classified as large hospitals if they are principal referral hospitals, public acute group A hospitals, or public acute group B hospitals, and classified as small hospitals if they are public acute group C hospitals, or public acute group D hospitals. [11] Moreover, hospitals in our sample

[8] See the reviews in O'Neill, Rauner, Heidenberger, and Kraus (2008), Kohl, Schoenfelder, Fügener, and Brunner (2019).

[9] Data on hospital staffing is provided in the form of FTE staff in six major categories including salaried medical officers, nurses, diagnostic and health professionals, other personal care staff, administrative and clerical staff, and domestic and other staff.

[10] Ideally, outputs of hospitals should be measured by the improvement in medical condition of patients. However, it is technically difficult to obtain this measure in practice, thus most of the hospital efficiency studies use quantities of services as an alternative measure of hospital outputs (Hollingsworth, 2008).

[11] Public acute hospitals in Australia are divided into five groups listed in descending order of activity volume and service diversification, as follows: principal referral hospitals, public acute group A hospitals, public acute group B hospitals, public acute group C hospitals, public acute group D hospitals. According to AIHW (2015), hospitals in the first three groups are generally larger than hospitals in the last two groups.

are also categorized into two groups based on their geographic location, namely remote hospitals (located in remote and very remote areas), and non-remote hospitals (located in major cities, inner regional, or outer regional areas). [12]

Table 1 provides information about the proportion of remote and small hospitals as well as total inputs utilised and total outputs provided by all hospitals belonging to each HHS in our sample. We can see that HHS 402, HHS 403, and HHS 436 are the only HHSs where all hospitals are small hospitals.[13] Moreover, almost all of their hospitals are located in remote areas. Meanwhile, the majority of hospitals managed by HHS 408, HHS 431, HHS 487, and HHS 494 are large hospitals and located in non-remote areas.

Table 1: Descriptive Statistics of variables by HHSs

Random ID	Proportion of Remote hosptials	Proportion of Small hospitals	No. of Beds (Total)	DMSEXP (Total) ($ millions)	FLABOUR (Total) FTE	FOUTPUT [14] (Total)
402	1.00	1.00	104	3.33	159.42	0.39
403	0.91	1.00	166	5.09	190.88	0.53
408	0.00	0.00	1167	170.77	3849.69	8.34
418	0.00	0.80	509	44.25	1270.69	3.19
423	0.25	0.88	335	50.44	1033.18	2.74
431	0.00	0.20	2354	387.34	5516.82	16.14
435	1.00	0.83	119	8.76	331.07	1.04
436	1.00	1.00	69	2.95	110.71	0.27
442	0.17	0.83	526	64.35	1211.94	3.17
451	0.00	0.86	843	112.95	2420.74	5.04
468	0.38	0.88	823	97.46	2207.21	5.25
478	0.00	0.70	584	64.53	1406.52	4.01
481	0.11	0.95	701	70.22	1624.25	3.98
487	0.00	0.25	300	90.02	1772.83	4.23
494	0.00	0.20	1937	316.03	5647.32	13.65

From Table 1, we also see that the utilisation of inputs and the provision of services varies significantly across the HHSs in our sample. For instance, HHS 436 has the total

[12]The classification is based on the remoteness area information provided in the Australian hospital peer groups in which the remoteness of a hospital is measured by the physical road distance to its nearest urban center.

[13]Note that the IDs here are not the real ID but randomly generated for each HHS.

[14]Since the unit of measurement of non-admitted occasions of services and casemix weighted inpatient episodes are different, we normalize them by their standard deviations before the aggregation.

number of beds of only 69. Meanwhile, the total number of beds operated by HHS 431 is 2354, being nearly 35 times higher than that of HHS 436. The similar pattern is also observed on the output side, where the highest figure of output factor is around 60 times higher than its lowest figure.

4 Results and Discussions

4.1 Univariate Input-Output Illustration

In this subsection, we aim at providing a graphical illustration of different types of frontier estimators. To do so, we utilize the same technique as discussed in Section 3 to aggregate inputs further into a single variable, we denote as FINPUT, representing all resources utilized by hospitals. In the case of univariate input-output production technology, we can present the estimated production frontiers (i.e., production functions) on a 2-D graph together with data points as shown in Figure 1. As we can see, DEA and FDH estimators envelope all the data points, whereas some data points are above the estimated order-α quantile frontiers (even for a relatively high value of α, say, $\alpha = 0.99$). Moreover, when α increases to 1, the estimated order-α quantile frontiers get closer to the estimated FDH frontier. Actually, as pointed out in Section 2, the FDH frontier is a special case of order-α quantile frontiers when $\alpha = 1$.

4.2 Main Analysis: Multiple inputs case

Before discussing the results, it is worth mentioning here that the results in this study are reported based on the reciprocal of the output oriented efficiency score, which gives an efficiency measure between 0 and 1 for FDH and DEA estimators, and an efficiency measure between 0 and $+\infty$ for order-α quantile frontier estimators. As a result, if a hospital has an efficiency score from the order-α quantile frontier estimators in $(0, 1)$, $\{1\}$, or $(1, +\infty)$, then it is interpreted, respectively, as "below", "on", or "above" the corresponding order-α quantile frontier..

Figure 2 shows how $p(\alpha)$, the percentage of hospitals being above the estimated order-α quantile frontier, changes when the order α increases. It is remarkable that when the order α increases from 0 to 0.8, $p(\alpha)$ decreases slowly, indicating that the quantile frontiers of orders α in this range are very tight. From the order of around 0.8, the decreasing rate of $p(\alpha)$ increases significantly, showing that the quantile frontiers become more spaced. The values of $p(\alpha)$ are, however, still relatively high for the values of α close to one. For example, $p(\alpha)$ is 51% for $\alpha = 0.95$, 30% for $\alpha = 0.98$ and still 5% for $\alpha = 0.99$. This

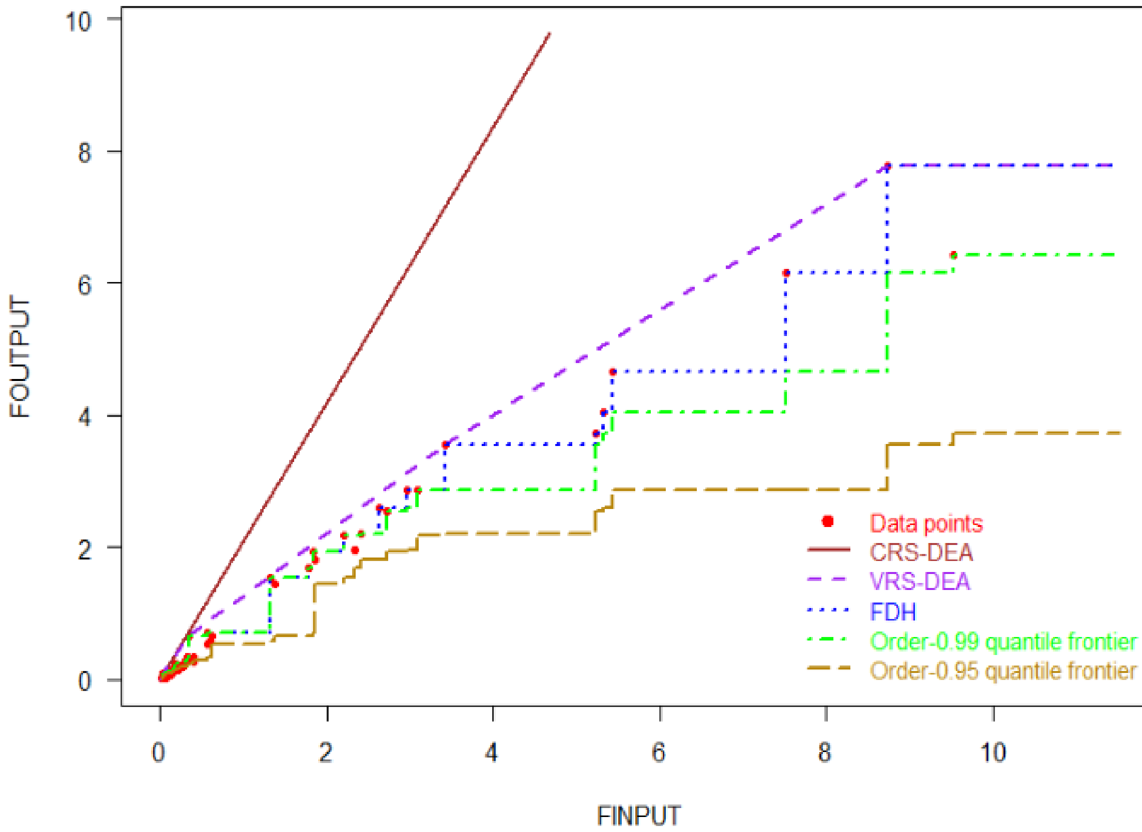

Figure 1: Estimated frontiers for the case of univariate input and output

fact suggests that only quantile frontiers of orders extremely close to one are possibly influenced by extreme values.

In the following analysis, we measure hospital efficiency with respect to the quantile frontiers of order $\alpha = 0.99$. This quantile frontier is less likely to be affected by extreme values/outliers (justified by the value of $p(\alpha)$) and still represents the output threshold exceeded by only 1% of hospitals in population using at most a given level of inputs. In our sample, all of those super-efficient hospitals are large and non-remote hospitals.

We are interested in comparing hospital efficiency across HHSs, thus after obtaining the individual estimates from various estimators (including order-0.99 quantile frontier, FDH, VRS-DEA, and CRS-DEA), we utilize the aggregate efficiency measure discussed in Section 2 to analyse the performance of HHSs. Table 2 reports the estimated aggregate efficiencies and Figure 3 presents histograms of the estimated aggregate efficiencies for different types of estimators.

For both VRS-DEA and CRS-DEA estimators, some HHSs turn out to be very inefficient, especially for CRS-DEA estimators, where 9 out of 15 HHSs are at least 40%

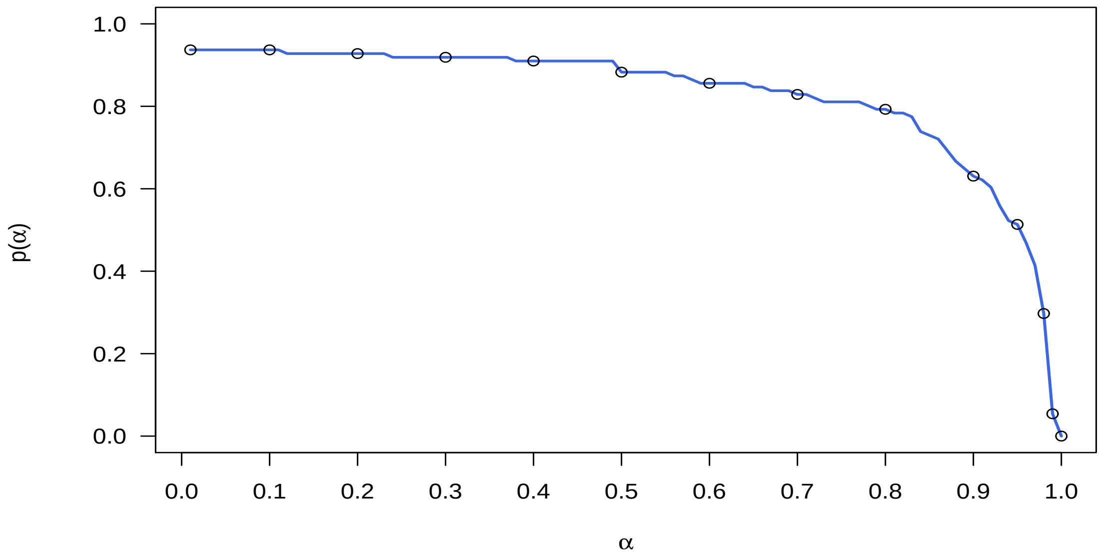

Figure 2: Evolution of the proportion of hospitals being above the estimated order-α quantile frontier

inefficient (see Figure 3). On the other hand, these models give a high variation in efficiency (or have high discriminative power) that might be explained through additional analysis. The prevalence of inefficient HHSs might be attributed to the fact that the frontiers estimated by DEA estimators are particularly sensitive to extreme values. A very few super-efficient production units can possibly shift the whole estimated frontiers outward and substantially change the distribution of the estimated efficiency scores. Identifying and removing these outliers from the sample (and studying them separately) may be useful for further analysis with the CRS-DEA model since it has value in itself. Indeed, provided there are no outliers, CRS-DEA can be considered as the most appropriate benchmark from a social point of view to evaluate the performance of production units in the public sector because it identifies the level of highest utilization of inputs into outputs (or highest average productivity) and the best practice socially-optimal scale.[15] In our sample, five hospitals are on the CRS-DEA frontier, which are hospital 1119, hospital 1031, hospital 1035, hospital 1095 and hospital 1001.[16] Among these five hospitals, four

[15]See more discussion in Grosskopf, Nguyen, Yong, and Zelenyuk (2020) and Nguyen and Zelenyuk (2020)

[16]Note that the IDs here are not the real ID but randomly generated for each hospital.

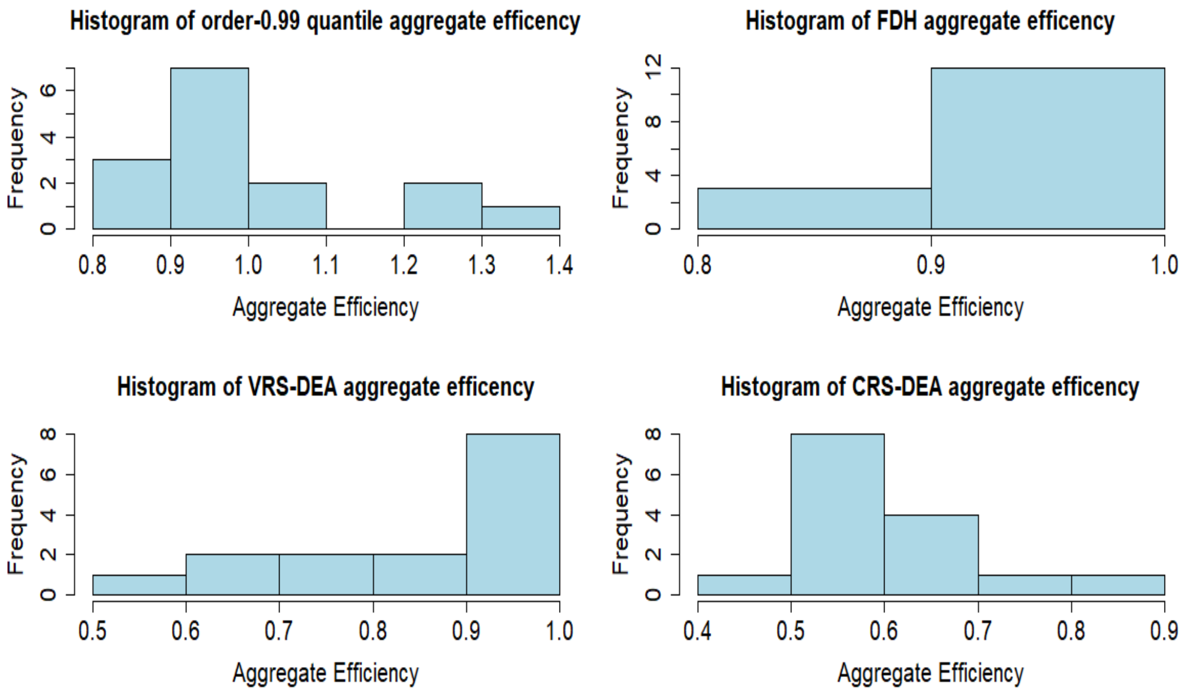

Figure 3: Histograms of aggregate efficiencies for different types of estimators

hospitals are small and located in remote areas. It might be an indicator that large hospitals in our sample are not operating near the socially-optimal scale, and this can be explored in future research.

Compared to DEA estimators, FDH estimators are less sensitive to extreme values. The estimated aggregate efficiencies obtained from FDH estimators are relatively reasonable ranging from 0.8 to 1. However, the FDH model has low discriminative power with many observations attaining high or 100% efficiency scores, of these some appear to be very inefficient when benchmarking using DEA. Moreover, for some HHSs, the evaluation of relative performance seems still to be influenced by the presence of super-efficient production units. For example, looking at Figure 4, where we present the estimated order-0.99 quantile aggregate efficiencies and FDH aggregate efficiencies on a heatmap, we can see that HHS 494 is in the top highest performance HHSs based on order-0.99 quantile aggregate efficiency, but it is in the bottom lowest performance HHSs based on FDH aggregate efficiencies. Due to the limited space, in the following discussion, we focus exclusively on the results obtained from order-0.99 quantile frontier estimators.

Based on order-0.99 quantile aggregate efficiencies, we use k-mean clustering technique to classify HHSs in Queensland into three groups, namely relatively low, medium, and

Table 2: Estimated Aggregate Efficiencies

Random ID	Efficiency Estimators				Clusters
	Order-0.99 quantile frontiers	FDH	VRS-DEA	CRS-DEA	
436	0.82	0.82	0.57	0.53	3
403	0.87	0.87	0.69	0.62	3
402	0.89	0.89	0.68	0.54	3
442	0.93	0.93	0.76	0.59	2
481	0.97	0.97	0.80	0.56	2
478	0.98	0.98	0.88	0.64	2
423	0.98	0.98	0.93	0.63	2
435	0.98	0.98	0.91	0.74	2
418	0.99	0.99	0.93	0.57	2
487	0.99	0.99	0.96	0.84	2
451	1.04	1.00	0.76	0.48	2
494	1.06	0.92	0.91	0.57	2
408	1.24	1.00	0.99	0.50	1
468	1.27	0.99	0.95	0.54	1
431	1.36	1.00	0.96	0.64	1

high efficiency (denoted as clusters 3, 2, and 1 respectively in Table 2).[17] The relatively low efficiency group (in dark yellow colour on the heatmap) includes HHS 402, HHS 403, and HHS 436. The relatively high efficiency group (in light yellow colour on the heatmap) includes HHS 408, HHS 431 and HHS 468. The relatively medium efficiency group is composed of the remaining HHSs.

To further investigate the differences in efficiency of HHSs, we look at characteristics of their hospitals. As discussed in Section 3, HHS 402, HHS 403, and HHS 436 are the only HHSs with all their hospitals being small hospitals. Moreover, almost all of their hospitals are located in remote areas. The boxplots in Figure 5 provide some insights about the relative performance of hospitals according to these characteristics. In our sample, large hospitals and hospitals in non-remote areas are relatively more efficient than small hospitals and hospitals in remote areas, respectively.

The above explanatory analysis suggests that the relatively low efficiency of HHS

[17] K-mean clustering is an unsupervised machine learning algorithm helping cluster data into a predetermined number of clusters so as to minimize the within-cluster sum of squares.

402, HHS 403, and HHS 436 with respect to the order-0.99 quantile frontier can be partially explained by the fact that the majority of their hospitals are small and located in remote areas. Rural hospitals are argued to face many disadvantageous conditions (e.g., shortages of medical staff, high chronic illness rate in the rural population, and stagnation in the rural economy); thus they might not provide health services as efficiently as urban hospitals do (Weisgrau, 1995). Similarly, compared to large hospitals, small hospitals might be less efficient because they usually have a lower level of standardization and specialization, resulting in weaker communication and coordination between hospital facilities (Munson & Zuckerman, 1983).

The evidence about the relative inefficiency in utilizing healthcare resources of small and remote hospitals might have useful policy implications for managers of relevant HHSs as well as Queensland Health. The presence of public hospitals in remote and very remote areas is an important vehicle to ensure equitable access to health services for all residents in Queensland given its geographically dispersed population. However, given the inefficiency of small and remote hospitals, other models of health service delivery, such as Telehealth, perhaps should be given a higher priority to develop as an alternative measure to better meet the healthcare needs of communities in rural areas.

It is worth recalling here that with the robust order-α quantile frontier estimator, hospitals are benchmarked relative to the frontier non-parametrically estimated from its closest peers, without imposing any assumptions of returns to scale, monotonicity or convexity. This flexibility may be viewed both as an advantage in some respects as well as a limitation in other respects. For example, the estimated efficiency scores could be very high, e.g., 100% or higher, for some very large hospitals perhaps because there are not many (or even any) peers to compare them with and reveal their inefficiency. In particular, all such large hospitals could be very large and very inefficient relative to the socially- optimal scale frontier (see more discussion in Nguyen & Zelenyuk, 2020).

In previous studies in the Australian context, large and urban hospitals were also found to be more efficient than small and rural hospitals (Paul, 2002; PC, 2010). However, as in the current paper, the constant returns to scale assumption is not imposed in all these studies, and thus hospitals are not benchmarked with respect to the socially-optimal scale frontier. In future research, it might be useful to explore hospital efficiency with respect to the socially-optimal scale frontier using a CRS-DEA model, since the scale efficiency might possibly be substantially different between small and large hospitals and might influence their relative efficiency.

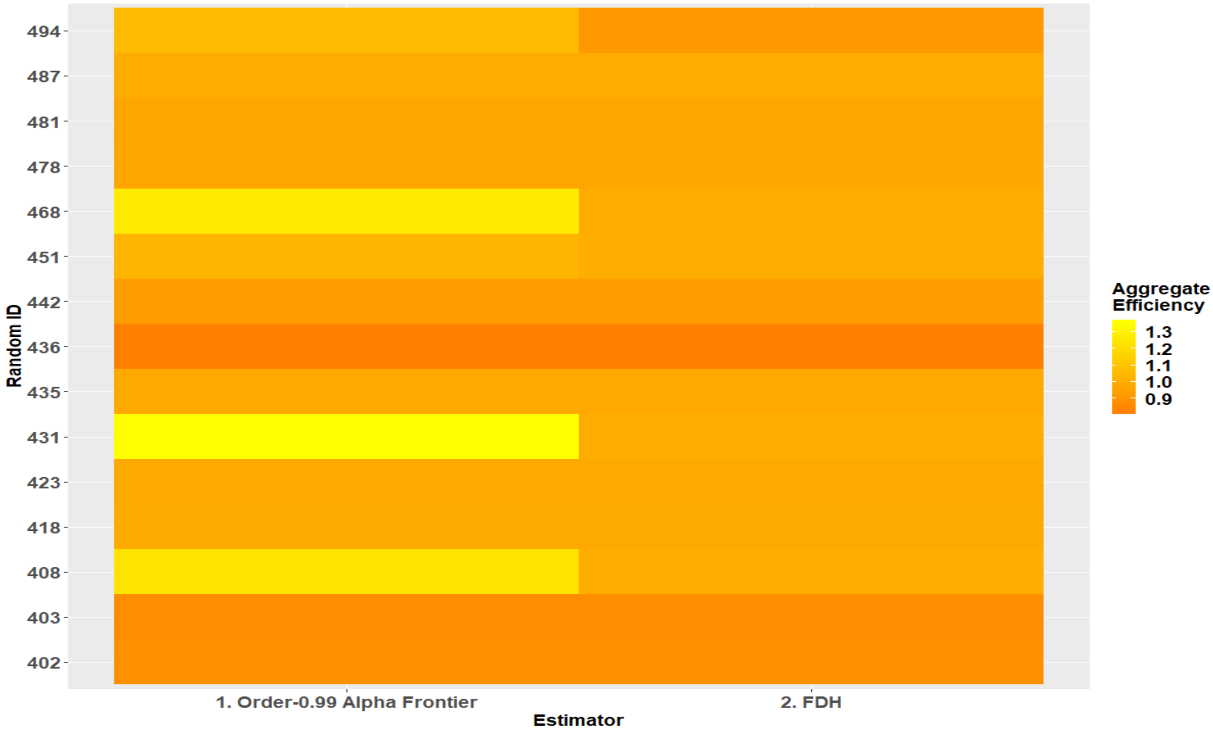

Figure 4: Aggregate Efficiencies of Queensland Local Hospital Networks (HHSs)

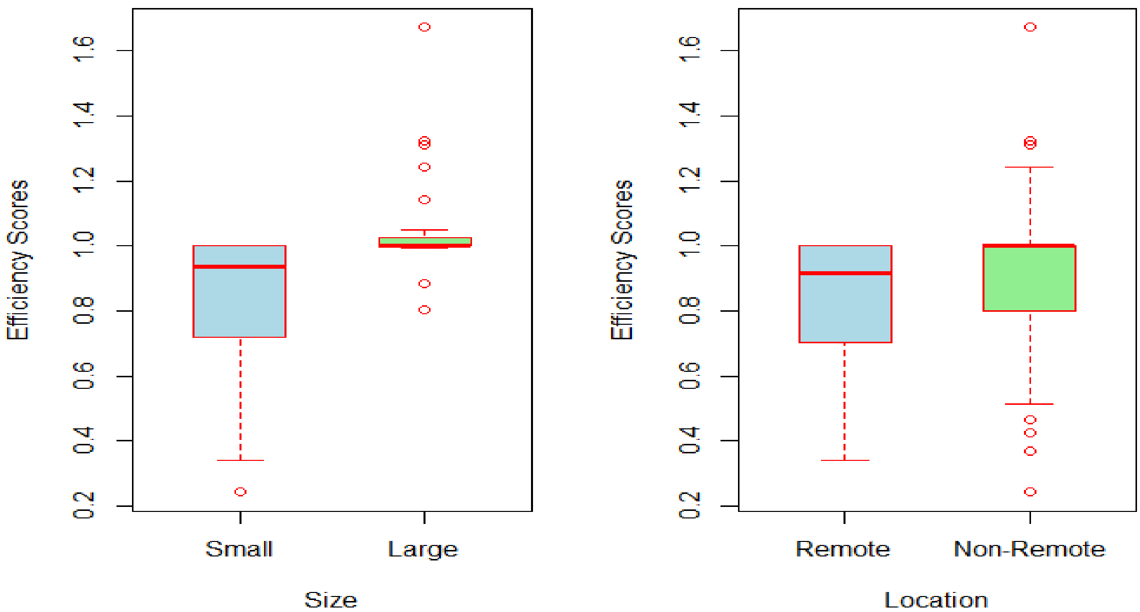

Figure 5: Boxplots of individual hospital order-0.99 quantile efficiencies by size and location

5 Conclusion Remarks

In this study, we explored the state of the efficiency of public hospitals at the level of Hospital and Health Services – independent statutory bodies who directly operate a group of public hospitals in a defined geographical area, in Queensland, Australia. To analyse their performance on the aggregate level, we utilize an aggregate efficiency measure constructed from individual efficiency scores which were estimated using various approaches. Besides the traditional nonparametric approaches like DEA and FDH, we also use a more recent and very promising robust approach–order-α quantile frontier estimators (Aragon, Daouia, & Thomas-Agnan, 2005). Our analysis suggests that efficiency scores of some Local Hospital Networks in Queensland are relatively low, which can be partially explained by the fact that the majority of their hospitals are small and located in remote areas.

Care is, however, needed when interpreting the results. High efficiency scores of large hospitals with respect to the order-α quantile frontier do not necessarily mean that they are efficient from a social point of view. These hospitals might utilise too many resources to deliver what can be otherwise done by smaller size hospitals that operate at a socially-optimal scale (see more discussion with intuitive examples in Nguyen & Zelenyuk, 2020). Indeed, operating at a socially-optimal scale is of vital importance for the healthcare systems, particularly in urgent circumstances, like pandemics. It allows hospitals to flexibly expand their operations to efficiently deliver the necessary healthcare services to society.

Moreover, the relatively low aggregate efficiency scores of some HHSs do not necessarily mean that they are not as efficient as other HHSs in operating public hospitals. There might possibly be other factors beyond the control of managers that are negatively affecting the performance of their hospitals. Remoteness and size are just two among many factors that are necessary to take into account. Moreover, although the above explanatory analysis is an important step to identify sources of efficiency differentials, more analysis that will account for other confounding factors is needed (e.g., see Grosskopf, Nguyen, Yong, and Zelenyuk (2020) for such an analysis in the context of Queensland, Australia).

Similar to many other studies in the literature, due to data availability, this study does not take into account quality dimension when estimating hospital efficiency and comparing the performance of HHSs. This might be unfair for those who have to utilize more resources to maintain the high quality of services. Therefore, a natural recommendation is to gather more data to incorporate the output quality indicator(s) in the analysis. Another fruitful direction of research would be to develop and apply statistical tests based on Central Limit Theorems for aggregate efficiency recently developed by Simar and Zelenyuk (2018) to statistically compare the performance of hospitals at HHS level.[18]

Acknowledgments

The authors acknowledge the support from their institution. We also acknowledge the financial support from the Australian Research Council (from the ARC Future Fellowship grant FT170100401). We thank Dan O'Halloran for his fruitful comments. We also thank David Du, Hong Ngoc Nguyen, Zhichao Wang and Evelyn Smart for their feedback from proofreading. We acknowledge and thank Queensland Health for providing part of the data that we used in this study. These individuals and organizations are not responsible for the views expressed in this paper.

[18]A work in progress in this direction is currently being done by Nguyen and Zelenyuk (2020).

References

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and astimation of stochastic frontier production function models. *Journal of Econometrics*, *64*(6), 1263–1297.

Aragon, Y., Daouia, A., & Thomas-Agnan, C. (2005). Nonparametric frontier estimation: A conditional quantile-based approach. *Econometric Theory*, *21*(2), 358–389.

Australian Institute of Health and Welfare. (2015). *Australian hospital peer groups* (tech. rep. No. 66). Australian Institute of Health and Welfare. Canberra, ACT.

Australian Institute of Health and Welfare. (2018). *Health expenditure Australia 2016–17* (tech. rep. No. 64). Australian Institute of Health and Welfare. Canberra, ACT.

Australian Institute of Health and Welfare. (2019). *Hospital resources 2017–18: Australian hospital statistics* (tech. rep. No. 233). Australian Institute of Health and Welfare. Canberra, ACT.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, *30*(9), 1078–1092.

Cazals, C., Florens, J.-P., & Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of econometrics*, *106*(1), 1–25.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.

Council of Australian Governments. (2011). *National health reform agreement.*

Daouia, A., & Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, *140*(2), 375–400.

Daraio, C., & Simar, L. (2007). Economies of scale, scope and experience in the italian motor-vehicle sector. In C. Daraio & L. Simar (Eds.), *Advanced robust and non-parametric methods in efficiency analysis: Methodology and applications* (pp. 135–165). Springer Science & Business Media.

Deprins, D., Simar, L., & Tulkens, H. (1984). Measuring labor efficiency in post offices. In M. G. Marchand, P. Pestieau, & H. Tulkens (Eds.), *The performance of public enterprises: Concepts and measurements* (pp. 243–267). Amsterdam, North-Holland.

Färe, R., Grosskopf, S., & Logan, J. (1983). The relative efficiency of illinois electric utilities. *Resources and Energy*, *5*(4), 349–367.

Färe, R., & Zelenyuk, V. (2003). On aggregate Farrell efficiencies. *European Journal of Operational Research*, *146*(3), 615–620.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, *120*(3), 253–290.

Grosskopf, S., Nguyen, B. H., Yong, J., & Zelenyuk, V. (2020). Healthcare structural reform and the performance of public hospitals: The case of Queensland, Australia [In Progress].

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. *Health Economics*, *17*(10), 1107–1128.

Kneip, A., Park, B. U., & Simar, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory*, *14*, 783–793.

Kneip, A., Simar, L., & Wilson, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory*, *24*(6), 1663–1697.

Kohl, S., Schoenfelder, J., Fügener, A., & Brunner, J. O. (2019). The use of data envelopment analysis (DEA) in healthcare with a focus on hospitals. *Health care management science*, *22*(2), 245–286.

Meeusen, W., & van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, *18*(2), 435–444.

Munson, F. C., & Zuckerman, H. S. (1983). The managerial role. In S. M. Shortell & A. D. Kaluzny (Eds.), *Health care management: A text in organization theory and behavior* (pp. 48–58). John Wiley & Sons.

Nguyen, B. H., & Zelenyuk, V. (2020). Aggregate efficiency of industry and its groups: The case of Queensland public hospitals [In Progress].

O'Neill, L., Rauner, M., Heidenberger, K., & Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio-Economic Planning Sciences*, *42*(3), 158–189.

Park, B. U., Jeong, S.-O., & Simar, L. (2010). Asymptotic distribution of conical-hull estimators of directional edges. *The Annals of Statistics*, *38*(3), 1320–1340.

Park, B. U., Simar, L., & Weiner, C. (2000). FDH efficiency scores from a stochastic point of view. *Econometric Theory*, *16*, 855–877.

Parmeter, C. F., & Zelenyuk, V. (2019). Combining the virtues of stochastic frontier and data envelopment analysis. *Operations Research*, *67*(6), 1628–1658.

Paul, C. J. M. (2002). Productive structure and efficiency of public hospitals. In K. J. Fox (Ed.), *Efficiency in the public sector* (pp. 219–248). Springer Science & Business Media.

Productivity Commission. (2010). *Public and private hospital: Multivariate analysis* (tech. rep. Supplement to Research Report). Productivity Commission. Canberra, ACT66.

Shephard, R. W. (1953). *Cost and production functions*. Princeton University Press.

Shephard, R. W. (1970). *Theory of cost and production functions*. Princeton University Press.

Sickles, R., & Zelenyuk, V. (2019). *Measurement of productivity and efficiency*. Cambridge University Press.

Simar, L. (2007). How to improve the performances of DEA/FDH estimators in the presence of noise? *Journal of Productivity Analysis*, *28*(3), 183–201.

Simar, L., & Wilson, P. W. (2013). Estimation and inference in nonparametric frontier models: Recent developments and perspectives. *Foundations and Trends® in Econometrics*, *5*(3–4), 183–337.

Simar, L., & Wilson, P. W. (2015). Statistical approaches for non-parametric frontier models: A guided tour. *International Statistical Review*, *83*(1), 77–110.

Simar, L., & Zelenyuk, V. (2007). Statistical inference for aggregates of Farrell-type efficiencies. *Journal of Applied Econometrics*, *22*(7), 1367–1394.

Simar, L., & Zelenyuk, V. (2011). Stochastic FDH/DEA estimators for frontier analysis. *Journal of Productivity Analysis*, *36*(1), 1–20.

Simar, L., & Zelenyuk, V. (2018). Central limit theorems for aggregate efficiency. *Operations Research*, *66*(1), 137–149.

Weisgrau, S. (1995). Issues in rural health: Access, hospitals, and reform. *Health care financing review*, *17*(1), 1–14.

Zelenyuk, V. (2020). Aggregation of inputs and outputs prior to data envelopment analysis under big data. *European Journal of Operational Research*, *282*(1), 172–187.