Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP08/2022

Activity based funding reform and the performance of public hospitals: The Case of Queensland, Australia

Bao Hoang Nguyen, Shawna Grosskopf, Jongsay Yong
and Valentin Zelenyuk

Date: November 2022

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

ISSN No. 1932 - 4398

Activity based funding reform and the performance of public hospitals: The Case of Queensland, Australia

Bao Hoang Nguyen* Shawna Grosskopf† Jongsay Yong‡
Valentin Zelenyuk§

Abstract

This study investigates the impact of activity-based funding (ABF) on the performance of hospitals by exploiting a natural experiment that happened in the state of Queensland, Australia. To examine the outcome of the reform, we use both a simple measure of performance (the weighted average length of stay) and more sophisticated ones (the technical efficiency estimated from data envelopment analysis (DEA) models). We try to identify the causal effect of ABF on the technical efficiency of hospitals by incorporating difference-in-differences approach in the popular two-stage DEA framework. We find empirical evidence that ABF improves the technical efficiency of hospitals.

Keywords: Hospital efficiency, Activity based funding, Healthcare reform, DEA, Difference-in-Differences, Truncated regression.

JEL Codes: C24, C61, I11, I18.

Funding information: the ARC Future Fellowship grant FT170100401.

Conflict of Interest Statement: the authors have no conflict of interest to declare.

Acknowledgments: Bao Hoang Nguyen and Valentin Zelenyuk acknowledge the support from the University of Queensland and the financial support from the Australian Research Council (FT170100401). We also thank Evelyn Smart and Zhichao Wang for their valuable feedback. 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.

*Corresponding author. School of Economics and Centre for the Business and Economics of Health, University of Queensland, Brisbane, Qld 4072, Australia. Email address: b.h.nguyen@uq.edu.au.

†Economics, School of Public Policy, Oregon State University, Corvallis, OR 97331, USA

‡Melbourne Institute of Applied Economic and Social Research, The University of Melbourne, Parkville, Australia

§School of Economics and Centre for Efficiency and Productivity Analysis, University of Queensland, Brisbane, Qld 4072, Australia

1 Introduction

Since its first application in the US in 1983, funding models based on the system of diagnosis related groups (DRGs) (hereafter, activity based funding - ABF) have been adopted as a basic means for reimbursing hospitals and have acted as a significant component of healthcare reforms in many developed countries, e.g., Italy (in 1995), Spain (in 1997), Germany and Denmark (in 2000), England and France (in 2004).[1]

In Australia, ABF was first introduced in the state of Victoria in 1993 (Duckett, 1995), but its nationwide application did not start until 2012 as a result of the National Health Reform Agreement (NHRA).[2] The NHRA in general and the application of ABF in particular have been viewed as a fundamentally important means of improving public hospital performance, helping to address cost pressure in the Australian healthcare sector and provide timely services for people in need.

Under ABF, the reimbursement takes into account both the quantity and complexity of hospital services. Specifically, hospitals are paid using a fixed rate for each patient care episode they provide, and the fixed rate for each episode is pre-determined based on the DRG to which the episode belongs. In principle, ABF provides financial incentives for hospitals to decrease unit costs and to simultaneously increase the number of episodes being treated, resulting in efficiency improvement (Böcking et al., 2005). The positive changes in some single indicators of hospital efficiency, such as the length of stay and the number of episodes, have been well-documented in many countries.[3] A similar pattern was also observed in Australia: between financial year (FY) 2012/13 and FY 2016/17, hospitalisations increased by an average of 4.3% in Australian public hospitals and the average length of stay decreased by 1.5 % on average.[4][5]

Although the key motivation for moving toward ABF is to improve hospital efficiency, there have been a limited number of studies explicitly investigating its impact on how efficiently hospitals utilize healthcare resources. As discussed above, most studies chose to examine a partial picture where some indicators of efficiency are considered (Street et al., 2011). For studies focusing on efficiency measures, findings about the impact of ABF on hospital efficiency are rather mixed. More importantly, they tend to focus on associations rather than the causal relationships. The implementation of ABF was found to positively

[1] For more details about the shift to ABF in those countries, see the special issue of the journal Health Care Management Science (Busse et al., 2006).

[2] The agreement can be found at `https://federalfinancialrelations.gov.au/sites/federalfinancialrelations.gov.au/files/2021-08/national-agreement.pdf`.

[3] See the summary in Street et al. (2011).

[4] Data from the Australian Institute of Health and Welfare, available at `https://www.aihw.gov.au/getmedia/d5f4d211-ace3-48b9-9860-c4489ddf2c35/aihw-hse-204.pdf.aspx?inline=true`.

[5] In Australia, a financial year starts from 1^{st} July and ends on 30^{th} June of the next calendar year.

correlate with efficiency improvement in Norway (Biørn et al., 2003; Biørn et al., 2010), Sweden (Gerdtham, Löthgren, Tambour and Rehnberg, 1999; Gerdtham, Rehnberg and Tambour, 1999), and Italy (Cavalieri et al., 2018). Meanwhile, studies in the US (Borden, 1988; Chern and Wan, 2000) and Germany (Herwartz and Strumann, 2014) revealed no positive correlation between ABF and hospital efficiency. In the Australian context, to the best of our knowledge, there has been no study examining the relationship.

The mixed results in the literature are attributed to various factors. One can consider the differences in studying settings including pre-existing funding systems and the ways ABF was implemented in studied countries. For example, Street et al. (2011) argued that moving from global budget (e.g., block funding) to ABF strengthens the incentive for increasing hospital activities (thus improving efficiency), as in Sweden and Norway, whereas the opposite is true for moving from fee-for-service reimbursement to ABF, as in the US.

Methodological caveats are another source of the mixed empirical results. In some studies, efficiency scores are simply compared between hospitals funded by ABF and those funded by other reimbursement methods (e.g., block funding or fee-for-service reimbursement), or between the period before and after the implementation of ABF. In others, regression analysis is employed to control for observable confounding factors, but the risk of unobservable confounders is not taken into account. Although more appealing, few studies employ statistical techniques such as difference-in-differences to control for unobservable confounding factors when examining the impact of ABF (Moreno-Serra and Wagstaff, 2010).

In the literature of hospital efficiency studies, data envelopment analysis is a well-established and useful technique, the majority of studies used DEA to estimate hospital efficiency and perform conventional statistical inferences based on the estimated efficiency scores.[6] However, by construction, while the DEA estimates of efficiency are consistent and have some optimality properties (Korostelev et al., 1995*a*,*b*; Kneip et al., 2008), in finite samples they are biased and correlated, so the conventional statistical inferences are theoretically invalid and need to be adapted.

This study addresses the above-mentioned issues by adapting and further developing analytically rigorous methodologies to examine the impact of ABF on hospital efficiency. To do so, we exploit what can be thought of as a natural experiment in the state of Queensland, Australia, where only a fraction of public hospitals was exposed to the new funding method (and at different points in time). To examine the outcome of the reform, we use both a simple measure of performance—the weighted average length of stay—as

[6]See the systematic reviews about the application of DEA in the healthcare sector, especially for public hospitals, in Hollingsworth (2003, 2008); O'Neill et al. (2008); Kohl et al. (2019).

well as the technical efficiency estimated from DEA models. To identify the causal effect of ABF on the technical efficiency of hospitals, we develop the difference-in-differences method for the popular two-stage DEA framework with the double bootstrapped truncated regression of Simar and Wilson (2007). We adapt the framework to the context of panel data to account for individual heterogeneity. These adaptations are novel and constitute the key methodological contributions of the current study.

We use data from 57 public acute hospitals in Queensland, observed during the period FY 2005/06 to 2016/17, and obtained from two data sets provided by the Queensland Department of Health (Queensland Health), namely Financial and Residential Activity Collection (FRAC) and Monthly Activity Collection (MAC). As the implementation of ABF in Queensland started from July 2011, the study time period is reasonable to examine the impact of the new funding method.[7] We extracted the information about the number of beds and hospital outputs from MAC, while FRAC provided us with the data on hospital inputs.

We find evidence that the implementation of ABF has had a positive impact on hospital efficiency. Specifically, our results suggest that ABF can, on average and ceteris paribus, reduce the inefficiency level of a hospital by about 4.8% to 5.4% depending on whether it is benchmarked to the variable returns to scale or constant returns to scale reference technology.

This paper is organized as follows. Section 2 provides background information about the ABF reform in Australia and its distinctive features as compared to those in other countries. Section 3 discusses measures of hospital efficiency, focusing on the Farrell-type technical efficiency and its DEA estimators. Section 4 outlines the regression analysis to estimate the causal effect of ABF on hospital efficiency. Section 5 describes the data and variables used in this study. Section 6 discusses the main results, and some concluding remarks are given in Section 7.

2 Background

Healthcare expenditure in Australia has increased significantly nationwide and across all states over the past decade. According to the Australian Institute of Health and Welfare, in FY 2016/17, Australia spent $181 billion (about 10% of GDP) on healthcare, about a

[7] Although the nationwide application of ABF started in Australia from July 2012, Queensland Health introduced ABF as a dominant mechanism for funding selected public hospitals one year ahead (see Government of Queensland, 2012, p.5). Also note that our study period ends before a material change in the funding agreement between the state and federal governments in Australia, when the growth in contribution of the federal government to ABF funding was capped at 6.5% from July 2017.

57% increase over FY 2006/07, after adjusting for inflation.[8] This increase was equivalent to an average annual growth rate of 4.7%, around 2% higher than average GDP growth.

Australia has a universal healthcare system with a mix of private and public providers. Private hospitals account for one-third of hospital beds and two-thirds of elective care episodes. All persons eligible for Medicare are entitled to access free services as a public patient in public hospitals in Australia. The provision of public hospital services is the responsibility of each state government, but the funding of public hospitals is shared between federal and state governments.

Queensland is the second largest and third most populous state in Australia. In line with with its population share, Queensland's healthcare expenditure made up about 21% of Australia's healthcare spending. Recurrent health expenditure in Queensland increased 65% between FY 2006/07 and FY 2016/17. Major sources of healthcare funding were from the federal and state governments: their combined share was 69% of total healthcare expenditure in Queensland in FY 2016/17. In turn, spending on healthcare accounted for the largest component of the state's general government sector expenditure (29.5% in FY 2016/17). The largest share of healthcare expenditure in the state was public hospital services, which accounted for 29% of the state's healthcare spending.[9]

To contain the healthcare cost and to provide timely services for people in need, the Australian government has made substantial efforts to promote efficiency in the sector. Of particular interest was the introduction of national reforms in the healthcare system. The most recent one was initiated in August 2011 when all states agreed to enter the NHRA aiming at delivering a nationally unified and locally controlled healthcare system. A key component of the reform is the nationwide application of ABF to fund public hospitals.

The implementation of ABF in Australia shared several common policy objectives with the other countries, but still possessed some distinctive features. As in other countries, the ultimate goals of applying ABF are: (i) to enhance public hospital efficiency and (ii) to improve patient access to healthcare services. In addition, ABF, along with other measures of the Australian national healthcare reform, are believed to serve an additional unique objective: to address the 'blame game' between federal and state governments about their responsibility for the provision of healthcare, especially public hospital services.[10] Under the reform, the federal government accepts the role as the principle funder of public

[8] Data from the Australian Institute of Health and Welfare is available at `https://www.aihw.gov.au/getmedia/e8d37b7d-2b52-4662-a85f-01eb176f6844/aihw-hwe-74.pdf.aspx?inline=true`.

[9] Data from the Queensland government is available at `https://s3.treasury.qld.gov.au/files/Report-on-State-Finances-2016-17.pdf`.

[10] "The states have blamed the federal government for insufficient funding to match the growing population needs, and in turn, the federal government has blamed poor management at the state level" (Hall, 2010).

hospitals (funding 45% of the growth in hospital activity via the ABF arrangement), the state governments act as system managers of public hospitals transforming their main role from funding to purchasing and overall planning of the hospital system (Hall, 2010).[11]

However, the ABF arrangement in Australia is also different compared to that in other countries. Although the proportion that the federal government contributes to the funding pool for public hospitals is calculated based on the national weighted activity unit and the national efficient price, state governments are allowed to purchase public hospital services based on their localization model.[12] Queensland Health, for example, utilizes the Queensland ABF model, which is argued to be "based largely on the National ABF model but includes a number of modifications to reflect Queensland priorities and pricing models that are more suitable." (Queensland Health, 2017).

In addition, the NHRA highlights that ABF should be applied wherever practicable, but still allows states to select either ABF or block funding for small hospitals who meet the "low volume thresholds."[13] In our study sample, 27 out of 57 hospitals (predominantly large and urban hospitals) started adopting ABF from July 2011, 2 other hospitals started adopting ABF from July 2013, and 2 additional hospitals started adopting ABF from July 2014. Meanwhile, 26 mainly rural and small hospitals are still non-ABF funded, i.e., by block funding.

The NHRA also led to a structural change in Queensland's health system due to the implementation of ABF. In July 2012, 16 health and hospital services (HHSs) were established in the state as independent statutory bodies with the responsibility of operating public hospitals in their local areas. Each HHS is a provider and budget holder in the service delivery model, whose function is to provide healthcare services to its local community. Meanwhile, Queensland Health is a purchaser who is responsible for purchasing healthcare services to cover the healthcare needs of citizens. The structural change could conceivably affect hospital efficiency by the changed incentive structure within the system.

[11] From July 2017, the growth in contribution of the federal government to ABF funding was capped at 6.5% (Parliament of Australia, 2018).

[12] A national weighted activity unit is a common unit for measuring health service activity against which the national efficient price is paid.

[13] See more details at `https://www.ihpa.gov.au/sites/default/files/publications/signed__pricing_framework_for_australian_public_hospital_services_2017-.pdf`.

3 Hospital Efficiency: Measurement and Estimation

3.1 Measure of Hospital Efficiency

To examine the impact of ABF on hospital performance, many studies examine changes in some indicators of hospital efficiency such as the quantity of services delivered and the average length of stay (e.g., Guterman et al., 1988; Kjerstad, 2003; Theurl and Winner, 2007; Farrar et al., 2009; Moreno-Serra and Wagstaff, 2010). This study goes beyond examining a single indicator of performance, such as weighted average length of stay, to focus on an overall measure of how efficiently a hospital converts its inputs into outputs – the Farrell-type technical efficiency (Farrell, 1957), which can be either in output or input orientation. The output (input) oriented technical efficiency measures how much a hospital needs to radially increase (decrease) all its outputs (inputs) to reach the level of maximal outputs (minimal inputs), given a level of inputs (outputs) and the same technology.

Some studies of hospital efficiency adopt the input-oriented models because hospital managers and policy-makers have less control over outputs than inputs, and thus it is reasonable to assume the goal of the hospitals is to minimize inputs given a level of outputs (O'Neill et al., 2008). The output-oriented models appear to be less popular than input-oriented models, but they are still used in the literature, especially in studies of public hospitals. Because the level of inputs used in public hospitals is usually fixed and influenced by external factors, some recent studies proposed that it would be more appropriate to assume that public hospitals try to maximize outputs given a level of inputs (e.g., Clement et al., 2008; Hu et al., 2012; Besstremyannaya, 2013; Chowdhury and Zelenyuk, 2016).

Besides the reasons discussed above, we adopt an output-oriented model because it is consistent with the aim of Queensland Health to "increase the level of hospital activity for a given level of inputs through technical efficiency" (Queensland Health, 2017, p.5). Moreover, under the constant returns to scale reference technology, efficiency scores obtained from both model orientations are equivalent.[14]

To define the measure of efficiency mathematically, assume that all hospitals use p inputs to produce q outputs, and denote $x = (x_1, ..., x_p)' \in \Re^p_+$ as a p-dimensional input vector, and $y = (y_1, ..., y_q)' \in \Re^q_+$ as a q-dimensional output vector. The Farrell output-oriented measure of technical efficiency for a particular hospital with input-output

[14]See Färe and Lovell (1978) and Sickles and Zelenyuk (2019, Chapter 3) for an extensive discussion of these and other measures.

allocation of (x, y), can be defined as follows

$$\theta(x, y \mid \Psi) = \sup_{\theta} \{\theta > 0 : (x, \theta y) \in \Psi\}, \quad (3.1)$$

where Ψ is a technology set characterizing the technology of the hospital defined as[15]

$$\Psi \equiv \{(x, y) \in \Re_+^p \times \Re_+^q : x \text{ can produce } y\}. \quad (3.2)$$

3.2 DEA estimators of the technical efficiency

The technology set discussed in Section 3.1 is not observable in practice, and thus the efficiency measure is not observable either. As a result, to empirically obtain the efficiency measure, we need estimate Ψ from a sample of data. Data envelopment analysis has been one of the most popular and powerful methods to estimate the technology and efficiency, especially for hospitals (e.g., see the systematic reviews in Hollingsworth, 2003, 2008; O'Neill et al., 2008; Kohl et al., 2019).

The application of DEA is, however, subject to various model specification assumptions, which require researchers to have rational selections based on a sound understanding of the industry and research purposes. In the hospital efficiency literature, an important assumption usually considered is about the returns to scale of the reference technology, e.g., constant returns to scale (CRS), variable returns to scale (VRS), non-increasing returns to scale (NIRS), and so on. A large number of studies impose the CRS assumption based on an intuitive understanding that patients with common conditions will generally be treated with the same procedure, so it would be reasonable to assume hospitals should be able to serve double the number of patients when doubling all inputs (Chilingerian and Sherman, 2004). Studies using the VRS assumption (e.g., Staat, 2006; Tiemann and Schreyögg, 2009) argue that imperfect competition, budget constraints, enormous capital costs involved in hospitals (e.g, buildings and equipment), as well as regulatory constraints on entry, mergers, and exits mean that hospitals may not be able to operate at their optimal scale at a given period of time, although they may strive to do so. We will consider both CRS and VRS approaches and will make a judgement call on what assumption is more relevant.[16]

[15]The technology set is assumed to satisfy the regularity axioms from the production theory, i.e., (i) Ψ is closed, (ii) the output set is bounded, (iii) it is possible to produce nothing, (iv) it is impossible to produce outputs without any inputs, and (v) all inputs and outputs are strongly disposable (see more detailed discussion in Sickles and Zelenyuk, 2019, Chapter 1).

[16]There are also many other variations of DEA including free disposal hull (FDH), DEA with weak disposability, and so on (see Sickles and Zelenyuk (2019, Chapter 8) for an extensive discussion of various variations of DEA).

To formulate the DEA estimator, consider a longitudinal sample of n hospitals over T periods, denoted as $\mathcal{S}_N = \{(x_{it}, y_{it}) : i = 1, \ldots, n,\ t = 1, \ldots, T_i\}$, where i and t index respectively individual hospital and time period, $N = \sum_{i=1}^{n} T_i$, and $T_i \leq T$. The DEA estimator of the reference technology under the CRS assumption is then defined as

$$\widehat{\Psi}_{CRS} \equiv \left\{ (x, y) : y \leq \sum_{i=1}^{n} \sum_{t=1}^{T_i} \xi_{it} y_{it}, x \geq \sum_{i=1}^{n} \sum_{t=1}^{T_i} \xi_{it} x_{it}, \right.$$
$$\left. \xi_{it} \geq 0, i = 1, \ldots, n, t = 1, \ldots, T_i \right\}. \tag{3.3}$$

To obtain the CRS-DEA estimator of technical efficiency, one can apply the plug-in principle, replacing Ψ in (3.1) by $\widehat{\Psi}_{CRS}$ in (3.3). Specifically, for a hospital k at time t, the CRS-DEA estimate of technical efficiency is given as

$$\widehat{\theta}\left(x_{kt}, y_{kt} \mid \Psi_{CRS}, \mathcal{S}_N\right) \equiv \sup_{\theta} \left\{ \theta > 0 : \theta y_{kt} \leq \sum_{i=1}^{n} \sum_{t=1}^{T_i} \xi_{it} y_{it}, x_{kt} \geq \sum_{i=1}^{n} \sum_{t=1}^{T_i} \xi_{it} x_{it}, \right.$$
$$\left. \xi_{it} \geq 0, i = 1, \ldots, n, t = 1, \ldots, T_i \right\}. \tag{3.4}$$

If one benchmarks hospitals with respect to the VRS reference technology, an additional constraint, $\sum_{i=1}^{n} \sum_{t=1}^{T_i} \xi_{it} = 1$, is added into equations (3.3) and (3.4) to obtain the VRS-DEA estimator:

$$\widehat{\Psi}_{VRS} \equiv \left\{ (x, y) : y \leq \sum_{i=1}^{n} \sum_{t=1}^{T_i} \xi_{it} y_{it}, x \geq \sum_{i=1}^{n} \sum_{t=1}^{T_i} \xi_{it} x_{it}, \right.$$
$$\left. \sum_{i=1}^{n} \sum_{t=1}^{T_i} \xi_{it} = 1, \xi_{it} \geq 0, i = 1, \ldots, n, t = 1, \ldots, T_i \right\}, \tag{3.5}$$

and

$$\widehat{\theta}\left(x_{kt}, y_{kt} \mid \Psi_{VRS}, \mathcal{S}_N\right) \equiv \sup_{\theta} \left\{ \theta > 0 : \theta y_{kt} \leq \sum_{i=1}^{n} \sum_{t=1}^{T_i} \xi_{it} y_{it}, x_{kt} \geq \sum_{i=1}^{n} \sum_{t=1}^{T_i} \xi_{it} x_{it}, \right.$$
$$\left. \sum_{i=1}^{n} \sum_{t=1}^{T_i} \xi_{it} = 1, \xi_{it} \geq 0, i = 1, \ldots, n, t = 1, \ldots, T_i \right\}. \tag{3.6}$$

It is worth noting here that in (3.3) and (3.4) as well as (3.5) and (3.6), data are pooled across periods to estimate a common reference technology. The advantage of this approach is that it makes the estimated efficiency scores comparable because all the observations are benchmarked to a common unconditional observed best practice frontier sometimes referred to as 'grand frontier'. Moreover, it also helps to increase the discriminative power and mitigate the 'curse of dimensionality' in the DEA context. Pooling data across periods, however, ignores technological changes. Thus decomposing the whole distance to the grand frontier into technological change and efficiency change and studying the determinants of these components would be a natural direction for future research (which will require substantially larger sample size).

4 Regression Analysis

As with any policy evaluation using observational data, causal inference is a key challenge in examining the impact of ABF. Although useful, simple comparisons between hospitals exposed to ABF and those that were not, or comparisons before and after the implementation of ABF, can be contaminated by the selection bias or temporal trends in the efficiency measures.

There are different approaches to estimate the causal effect in observational settings. They vary based on the assumption about the relationship between the treatment assignment and potential outcomes. Under the unconfoundedness assumption, which requires the treatment assignment to be independent of potential outcomes (conditional on observed covariates), one can choose among a variety of methods including regression, propensity scores, matching, or a combination. Alternatively, one can use instrumental variables or difference-in-differences methods, which require assumptions to some extent less strict than the unconfoundedness, yet additional information is often needed (e.g., valid instruments or a natural experiment setting).[17] In this study, we try to identify the causal effect by employing the difference-in-differences method.[18]

We exploit what can be thought of as a natural experiment where only a fraction of public hospitals in Queensland was exposed to the new reimbursement method (and at different points in time), while others are still funded by block funding. To utilize the difference-in-differences analysis, we model the relationship between a measure of hospitals' efficiency and its determinants by the following regression model

$$\eta_{it} = \delta D_{it} + W_i'\zeta + Z_{it}'\beta + \sum_{t=2}^{T} \alpha_t \, Year_t + c_i + \varepsilon_{it}, \tag{4.1}$$

where η_{it} is the outcome of interest of hospital i at time t, D_{it} is the policy variable representing funding status of hospital i at time t, taking value 1 if hospital i is ABF funded at time t and 0 otherwise, and W_i and Z_{it} are vectors of control variables consisting of time-invariant and time-varying observables, respectively. Furthermore, $Year_t$ is a period

[17] See a review of methods for policy evaluation in Imbens and Wooldridge (2009).

[18] The method is one of the most popular approaches to estimate causal effects in observational settings, popularized by the seminal work of Card and Krueger (1994) and recently extended in several methodological innovations (see recent reviews in de Chaisemartin and D'Haultfoeuille, 2022; Roth et al., 2022).

specific dummy capturing the aggregate effect that may influence the efficiency measures of all hospitals at period t, and c_i is a hospital specific effect capturing time-invariant unobservables (i.e., unobserved individual heterogeneity), which may affect the outcome and possibly correlate with the policy variable.[19] Lastly, ε_{it} is the idiosyncratic error.

The key assumption for the identification of the causal effect is the parallel trends assumption. In line with de Chaisemartin and d'Haultfoeuille (2020) and Sun and Abraham (2021), we assume that the evolution in the expected untreated potential outcomes are the same across hospital cohorts including all ever-treated cohorts and the never-treated cohort.[20] Specifically, let G_g be a dummy variable taking the value of 1 if a hospital starts adopting ABF at period g and 0 otherwise (and $g = \infty$ for the cohort of hospitals never funded by ABF). Then,

$$E(\eta^0_{it} - \eta^0_{i,t-1} \mid G_g = 1) = E(\eta^0_{it} - \eta^0_{i,t-1}), \qquad \forall g \in \mathcal{G} \cup \{\infty\}, \text{and } t = 2, \dots, T, \tag{4.2}$$

where η^0_{it} is the potential untreated outcome of hospital i at time t, and $\mathcal{G}$ is a set of all adoption time excluding ∞. Under the parallel trends assumption (and homogeneous treatment effects), the coefficient δ in model (4.1) is then the parameter of the causal effect. In this study, we focus on two main outcome variables: the weighted average length of stay and the Farrell-type output oriented technical efficiency.[21] Both outcome variables are in logarithmic form, so the partial effect of the activity based funding on the dependent variable in model (4.1) is interpreted as semi-elasticity of the efficiency measure with respect to the funding status.

A noteworthy feature of the Farrell-type output oriented technical efficiency is that it is always greater than or equal to 1 by construction. As a result, for the model of our main focus, the dependent variable (i.e., $\ln \theta_{it}$) in (4.1) is left truncated at 0. There are different approaches to estimate the panel data truncated model (with the individual heterogeneity). One approach is to utilize the fixed effects method (FE), where the individual heterogeneities are viewed as parameters to be estimated. The main advantage of

[19] The setup here is in principal the two-way fixed effects (TWFE) regression. The treatment coefficient in TWFE setup is an unbiased estimator for the average treatment effect if the the parallel trends assumption holds and the treatment effect does not vary over time and across cohorts (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfoeuille, 2022). Adapting heterogeneity-robust difference-in-differences estimators, e.g., de Chaisemartin and d'Haultfoeuille (2020); Callaway and Sant'Anna (2021); Sun and Abraham (2021), into our context would be a fruitful direction for future research.

[20] An ever-treated cohort of hospitals is a group of hospitals that started adopting ABF at the same point in time, and the never-treated cohort is a group of hospitals that never adopt ABF during the study period.

[21] To account for the difference in resources required and the complexity of different diagnoses and treatments, the average length of stay of each hospital is calculated as the weighted mean of average length of stay in each DRG. The weights used for the calculation are case-mix cost weights of DRGs. (See more discussion about case-mix cost weights in Section 5.)

the FE approach is that it does not impose any restriction on the relationship between the unobserved individual heterogeneity and observed covariates. However, in the context of non-linear models as in our case of panel data truncated model, the FE method is well-known to suffer from the incidental parameters problems. Another approach to estimate model (4.1) is to use the random effects (RE) method. Yet, the RE approach requires the assumption of no correlation between the unobserved individual heterogeneity and observed covariates, which seems to be too strict in our context. In this study, we utilize the correlated random effects method (CRE) (by incorporating the Mundlak terms), which is a compromise between the FE and RE approaches, as discussed in Wooldridge (2019). CRE restricts the correlation between the unobserved individual heterogeneity and observed covariates to the relationship stated in (4.3) below.[22] Specifically, following Mundlak (1978), we model the time-invariant unobservables as

$$\begin{aligned} c_i &= c_0 + \sum_{g \in \mathcal{G}} \lambda_g G_g + \overline{Z_{i\cdot}}'\tau + a_i, \\ a_i &\overset{\text{iid}}{\sim} \mathbb{N}(0, \sigma_a^2), \end{aligned} \tag{4.3}$$

where $\overline{Z_{i\cdot}}$ is the time average of Z_{it}, i.e., $\overline{Z_{i\cdot}} = \dfrac{1}{T_i} \sum_{t=1}^{T_i} Z_{it}$. Note that the appearance of cohort dummies in (4.3) is equivalent to including the time average of the policy variable, D_{it} (see the relevant discussion in Wooldridge, 2021).

Substituting (4.3) to (4.1) (with $\ln \theta_{it}$ as the dependent variable), we have

$$\begin{aligned} \ln \theta_{it} &= c_0 + \delta D_{it} + W_i'\zeta + Z_{it}'\beta + \sum_{t=2}^{T} \alpha_t Year_t + \sum_{g \in \mathcal{G}} \lambda_g G_g + \overline{Z_{i\cdot}}'\tau + a_i + \varepsilon_{it} \\ &= V_{it}'\Gamma + a_i + \varepsilon_{it}, \end{aligned} \tag{4.4}$$

where V_{it} is a vector collecting all covariates (including unity) and Γ is a vector collecting all coefficients, i.e., $V_{it}' = (1, D_{it}, W_i', Z_{it}', Year_t, G_g, \overline{Z_{i\cdot}}')$ and correspondingly $\Gamma' = (c_0, \delta, \zeta', \beta', \alpha_t, \lambda_g, \tau')$. Conditioning on a_i and V_{it}, ε_{it} is assumed to be independent and to follow $\mathbb{N}(0, \sigma_\varepsilon^2)$ left truncated at $-V_{it}'\Gamma - a_i$.[23] Model (4.4) can be estimated using the method of maximum likelihood (MLE).

Since the efficiency level of hospital i at time t, θ_{it}, is not observable in practice, to estimate the model specified in (4.4), we replace θ_{it} by its DEA estimate, $\widehat{\theta}_{it}$. As discussed

[22]Another restriction imposed in our study in the panel data context (regardless of whether FE, RE, or CRE approach is utilized) is that the observed covariates are strictly exogenous conditional on the individual heterogeneity. This restriction rules out the 'feedback' effect, i.e., it prevents covariates in the next period from being affected by a shock to the dependent variable in the current period. (See more discussions in Wooldridge (2010, Chapter 10 & Chapter 13)).

[23]In the case where the dependent variable is the logarithm of the weighted average lengths of stay, we assume $\varepsilon_{it} \overset{\text{iid}}{\sim} \mathbb{N}(0, \sigma_\varepsilon^2)$ and estimate the model using feasible generalized least square estimator (FGLS) with standard errors clustered at individual level.

in Simar and Wilson (2007), $\widehat{\theta}_{it}$ is however biased and serially correlated in a complicated way, and thus conventional statistical inferences are not valid. To overcome this, we extend the double bootstrap procedures of Simar and Wilson (2007) into the context of panel data truncated regression.

Another feature of the truncated regression in (4.4), often overlooked by applied researchers in utilizing the method of Simar and Wilson (2007), is that the coefficients in model (4.4) do not precisely measure the average partial effects of corresponding covariates on the mean value of the dependent variable. The partial effects are different across observations in the sample and overestimated by the corresponding coefficients. To obtain the estimate of the average partial effect of each covariate, we take the average of the individual-specific partial effects evaluating at each data point in the sample and also average out the individual heterogeneity.[24][25]

5 Data and Variables

The annual data on Queensland public hospitals for the period from FY 2005/06 to FY 2016/17 were provided by Queensland Health from the two data collections, namely Financial and Residential Activity Collection (FRAC) and Monthly Activity Collection (MAC). We extracted the information about teaching status, drug, surgical and medical supply expenditures, and hospital staffing from FRAC. Meanwhile, the number of beds, non-admitted patient activities, admitted patient episodes of care, and patient days by DRG were obtained from MAC. In addition, the information about hospital peer groups and geographic location was sourced from AIHW (Australian Institute of Health and Welfare, 2015).

To maintain homogeneity, especially to make ever-treated and never-treated hospital cohorts comparable, only public acute hospitals are included in the sample. Moreover, for health facilities in rural and remote areas, due to the variation in their characteristics and service mix, only district and rural hospitals, which serve populations of more than 2000 people with a comprehensive mix of acute health services including medical, surgical, emergency and maternity, are included. We also exclude hospitals that do not have

[24]The partial effects in the framework of Simar and Wilson (2007) have been recently discussed in the work of Badunenko and Tauchmann (2019), where they introduced a new Stata command to implement the double bootstrapped truncated regression. Note that their computation for partial effects utilizes a post-estimation command in Stata, i.e., the `margins` command, and is only applicable for the case of cross-sectional data. We use a generalized version of their approach following Wooldridge (2010, Chapter 15) but adapted to our context. Our computation is done in Matlab using our user-written codes.

[25]For the activity based funding dummy, we average the partial effects across ever-treated hospitals in the post-adoption periods to recover the average treatment effect on the treated.

complete data for the entire study period. There is, however, an exception. In 2013, the Gold Coast University Hospital was opened to replace the Gold Coast Hospital with all patients from the old hospital being transferred to the new one. Moreover, since FY 2012/13, activities of Robina Hospital, which had previously been integrated into the Gold Coast Hospital, have been separately reported. To maintain the consistency in our analysis, we include in our sample a representative hospital created by combining the data of the three hospitals. Our final dataset contains 684 observations including 57 public acute hospitals over a twelve-year period.

5.1 Inputs

5.1.1 Labour Input

Following the recent literature, especially those undertaken in the Australian context (Productivity Commission, 2010; Chua et al., 2011; Nghiem et al., 2011; O'Donnell and Nguyen, 2013), full-time equivalent (FTE) staff is utilized as a proxy for labour input. The data on FTE staff in Queensland public hospitals is classified into six major categories including salaried medical officers (*meo*), nurses (*nur*), diagnostic and health professionals (*dhp*), other personal care staff (*opcs*), administrative and clerical staff (*acs*), and domestic and other staff (*dos*). Each type of labour plays different roles and contributes differently to the production process of hospitals. While staff in the first four categories are involved more directly in diagnostic and treatment procedures, those in the last two categories are more engaged in supporting activities such as administrative and clerical duties, and the provision of food and cleaning services.

Ideally, all these labour categories should be used in the efficiency analysis as separate inputs, since labour is the most important input in hospital production and the difference in labour composition may result in significant differences in hospital performance. However, when the number of inputs increases, the DEA approach (as other nonparametric approaches) suffers from the well-known 'curse of dimensionality', which not only leads to the bias of efficiency estimates but also lowers the power of statistical tests. To overcome this problem, we adopt Daraio and Simar's (2007) approach to aggregate all labour categories into a variable representing labour input, called 'labour factor' (*flabour*). (The approach is based on principal component analysis, for more details see the Online Supporting Information.) To summarize, *flabour* is constructed as follows

$$\mathit{flabour} = 0.29\mathit{meo} + 0.84\mathit{nur} + 0.24\mathit{dhp} + 0.05\mathit{opcs} + 0.29\mathit{acs} + 0.26\mathit{dos}. \tag{5.1}$$

The constructed 'labour factor' explains 98.98% of total inertia of the original data and has high correlations with each of the labour categories (see the Online Supporting Information). As a result, this dimension reduction approach allows us to mitigate the impact of the 'curse of dimensionality' without loosing much information, and the factor *flabour* is a good representative of hospital labour inputs.

5.1.2 Capital input

Capital inputs are usually referred to as long-term assets such as buildings, medical equipment, and vehicles. However, in the hospital efficiency analysis, the interest of researchers is more about the contribution of capital inputs to a hospital production process rather than the amount of capital stock owned by each hospital. Incorporating capital stock as an input in the efficiency analysis may lead to overestimating the utilization of capital in the production process, resulting in efficiency estimates that do not accurately reflect the underlying technology (Worthington, 2004). Measures of capital utilization are, however, usually not available in practice. To overcome this problem, researchers often choose alternative measures that are proportional to capital usage. The number of beds is such a measure and is commonly used in hospital efficiency studies (O'Neill et al., 2008). Following the common practice in the literature, we use the number of beds as a proxy for the utilization of capital in this study.

5.1.3 Consumable input

Drug, surgical and medical supplies are the main types of consumables that are usually incorporated in the hospital efficiency analysis. These components account for a substantial portion of hospitals' non-salary recurrent expenditure. Although consumables are considered as an important category of hospital inputs, the use of this input category in hospital efficiency studies is not common, mainly due to the problem of data availability. Since there is limited information about the amount of drug and medical supplies used, consumable inputs are usually included in hospital efficiency studies in monetary rather than physical units (e.g., Biørn et al., 2003; Productivity Commission, 2010; Chua et al., 2011; Besstremyannaya, 2013; Chowdhury and Zelenyuk, 2016). Similarly, we use expenditures on drug, surgical and medical supplies to represent hospital consumable input in this study. Drug, surgical and medical supply expenditures are recorded at the current price at the end of each financial year. We use the consumer price index (CPI) from the Australian Bureau of Statistics to deflate nominal dollar values to constant dollar values, with FY 2012/13 as the base year.

5.2 Outputs

Ideally, outputs in hospital sector should be measured by the improvement in patients' health, but it was infeasible to obtain this measure in practice. Most of the hospital efficiency studies use quantities of services as an alternative measure of hospital outputs. Hospital services are usually disaggregated further into two main sub-categories, namely inpatient and outpatient. Outpatient outputs are measured by the number of outpatient visits and/or the number of emergency department attendances (e.g., Hao and Pegels, 1994; Zuckerman et al., 1994; Magnussen, 1996; Harris et al., 2000; Nayar and Ozcan, 2008; Garcia-Lacalle and Martin, 2010; Hu et al., 2012; Nayar et al., 2013; Chowdhury and Zelenyuk, 2016).

For inpatient outputs, in order to account for the difference in resources required and the complexity of different diagnoses and treatments, quantities of inpatient services are adjusted according to various weighting criteria. The most common is casemix cost weight, which reflects the complexity as well as the resources required for an episode within a particular diagnostic-related category (DRG) relative to the average of all episodes. Studies using this method usually aggregate all inpatient episodes/days in different DRGs into a single inpatient output measure, weighting by casemix cost weight of each DRG (e.g., see Burgess and Wilson, 1996, 1998; Harris et al., 2000; Hofmarcher et al., 2002; Biørn et al., 2003; Clement et al., 2008; Nayar and Ozcan, 2008; Berta et al., 2010; Ferrier and Trivitt, 2013; Nayar et al., 2013; Chowdhury and Zelenyuk, 2016).

Following the common practice in the literature, outpatient outputs of Queensland public hospitals are measured by the number of non-admitted occasions of service including both outpatient visits and emergency department services. The number of casemix weighted inpatient episodes is used to measure inpatient outputs. Specifically, we multiply the number of inpatient episodes in each DRG by the DRG cost weight obtained from the National Hospital Cost Data Collection, to arrive at the casemix weighted inpatient episode number.[26]

5.3 Control variables

5.3.1 Hospital size

Hospital size is widely accepted as an important factor influencing hospital efficiency. Work processes in large hospitals—in comparison with small hospitals—tend to be standardized and managerial roles in those hospitals are usually undertaken by specialized staff. This typically results in better communication and coordination among hospital facilities, which may lead to efficiency enhancement (Munson and Zuckerman, 1983). Moreover, large hospitals may benefit from the high level of specialization, which allows the labour force and hospital equipment to operate at their full capacity. Empirical studies, however, show mixed results about the relationship between hospital size and hospital

[26]We use the constant cost weights of FY 2012/13.

efficiency. For example, Watcharasriroj and Tang (2004) and McCallion et al. (1999) found that small hospitals were on average less efficient than large hospitals in Thailand and Northern Ireland, respectively. In contrast, Chowdhury and Zelenyuk (2016) found evidence that small hospitals on average were more efficient than large hospitals when investigating the performance of public acute hospitals in Ontario, Canada. Moreover, Hao and Pegels (1994) obtained mixed results about the relationship between hospital size and hospital efficiency when using various proxies for hospital size in their study of 97 Veterans Affairs acute hospitals in the US.

In this study, we incorporate hospital size as an explanatory factor for hospital efficiency differentials. According to the Australian hospital peer groups developed by Australian Institute of Health and Welfare (2015), public acute hospitals in Australia are divided into five groups based on hospitals' service profile characteristics. The five groups are listed in descending order of activity volumes 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. As stated in Australian Institute of Health and Welfare (2015), hospitals in the first three groups are larger than those in the last two. As such, we classify principal referral hospitals, public acute group A and B hospitals as large hospitals, whereas public acute group C and D hospitals are classified as small hospitals. We create a dummy variable representing large hospitals to include as a control variable in our regression models.

5.3.2 Geographic location

Geographic location is another factor that many researchers have taken into account when examining the difference in hospital efficiency. It is argued that disadvantageous conditions in rural areas, such as high chronic illness rate, stagnation in the rural economy, and shortages of medical staff, may prevent rural hospitals from providing healthcare services as efficiently as urban hospitals (Weisgrau, 1995).

However, as with hospital size, the previous studies did not provide a consensus view of the relationship between geographic location and hospital performance. For example, Mick and Morlock (1990) reviewed studies on the performance of rural hospitals in the US during the 1980s and indicated that rural hospitals were usually found to be worse than urban hospitals in terms of both organizational and financial performance. Similarly, Nayar and Ozcan (2008) found evidence that on average urban hospitals performed more efficiently than their rural counterparts when studying 117 non-federal acute care hospitals in Virginia, the US. In contrast, Garcia-Lacalle and Martin (2010) showed that rural and urban hospitals were not significantly different in terms of technical efficiency when comparing efficiency scores of hospitals in Andalusia, Spain.

In our study, hospitals are categorized into two groups according to their geographic location, namely major city hospitals (located in major cities), and regional and remote hospitals (located in regional, remote and very remote areas). We create a dummy variable representing hospitals located in major cities to include as a control variables in our regression models.

5.3.3 Teaching status

Generally, there are opposite forces driving the relative efficiency between teaching and non-teaching hospitals. On the one hand, teaching hospitals are sometimes expected to operate less efficiently than non-teaching ones because they have to spend additional resources on teaching and research activities. On the other hand, it is believed that medical staff in teaching hospitals are beneficial from being involved in research and teaching activities, and thus they become more effective in performing medical procedures.[27]

The empirical evidence about the relative efficiency between teaching and non-teaching hospitals is mixed in the literature. For example, Grosskopf et al. (2001) and Chowdhury and Zelenyuk (2016) found evidence that teaching hospitals, on average and ceteris paribus, were less efficient than non-teaching hospitals when investigating the efficiency of hospitals in the US and Canada, respectively. In contrast, Nayar et al. (2013) found that the efficiency level of acute hospitals in the US was positively associated with the teaching status. Recently, Nguyen and Zelenyuk (2021) found evidence that the relative efficiency between teaching and non-teaching hospitals largely depended on the reference technology when analyzing the efficiency of hospitals in Queensland, Australia.

In Australia, a hospital is defined as a teaching hospital if it is affiliated with universities to provide medical education and training as advised by the relevant state health authority. We create a dummy variable representing teaching hospitals to include as a control variable in our regression models. It is also important to note that although there are some hospitals which changed their teaching status during our study period, the teaching status of the majority of hospitals in our sample is constant over time. As a result, we do not use the time average of teaching dummy as a Mundlak adjustment to avoid nearly perfect multicollinearity.

5.3.4 Other control variables

We also include the following variables in the regression models to control for hospital heterogeneity: (i) occupancy rate, (ii) casemix index, (iii) the ratio of outpatient volume to inpatient volume, and (iv) the unit producing personnel ratio. These variables are

[27]Relevant discussions can be found in Nguyen and Zelenyuk (2021).

included in the regression models in the logarithmic form, and their time averages are also included as Mundlak adjustments.

The occupancy rate is measured as the ratio of hospital inpatient days to total bed days available and is expected to positively relate to hospital efficiency. It is argued that a hospital with a high occupancy rate utilizes its current capacity better, i.e., it can serve more patients with a similar number of beds and medical staff compared to its peers with lower occupancy (Litvak and Bisognano, 2011). The argument is consistent with the empirical findings in the previous studies in Australia (e.g., Yong and Harris, 1999; Productivity Commission, 2010) as well as in other countries (e.g., Zuckerman et al., 1994; Herwartz and Strumann, 2014; Chowdhury and Zelenyuk, 2016), in which the evidence of the positive relationship between occupancy rate and hospital efficiency was found.

The casemix index is measured as the ratio of total casemix weighted inpatient episodes to total inpatient episodes and incorporated in the regression models to control for differences in the complexity and severity of illness of patient cases treated. We expect that the casemix index negatively affects hospital performance because hospitals facing more severe and complex cases may consume more resources, and thus appear to be relatively less efficient, as found in Chilingerian (1995).

The difference in composition of services provided (outpatient vs. inpatient services) is also argued to influence hospital performance (Kirigia and Asbu, 2013; Chowdhury and Zelenyuk, 2016). This is because outpatient services usually require less resources than inpatient services, thus a hospital that provides a higher proportion of outpatient care tends to deliver relatively more outputs from a given level of inputs and appears to be more efficient. In our study, the ratio of outpatient volume to inpatient volume is measured as the ratio of non-admitted occasions of service to total inpatient episodes.

Finally, the unit producing personnel ratio is measured as the ratio of full-time equivalent staff in four directly producing labour categories (salaried medical officers, nurses, diagnostic and health professionals, and other personal care staff) to total full-time equivalent staff. We expect that hospitals with a higher proportion of unit producing personnel tend to operate more efficiently since unit producing personnel staff are directly involved in diagnostic and treatment procedures, thus the greater the share of these staff, the more patient care hospitals can provide. Moreover, it is also consistent with the empirical findings of Burgess and Wilson (1998) and Chowdhury and Zelenyuk (2016), who found evidence that the performance of hospitals in respectively the US and Canada were positively associated with the ratio of clinical staff to non-clinical staff.

In Table 1, we present descriptive statistics for the sample as well as for different cohorts of hospitals based on the timing of ABF adoption. It can been seen that the cohort of hospitals adopting ABF in FY 2011/12 are much larger than the remaining cohorts

Table 1: Summary statistics

	All sample	Hospital cohorts based on the timing of ABF adoption			
		FY 2011/12	FY 2013/14	FY 2014/15	Never adopt
Labour factor	378.72	756.77	68.61	57.08	34.73
	(657.97)	(801.66)	(22.02)	(9.29)	(16.13)
Total beds	160.85	308.25	37.58	37.96	26.71
	(244.03)	(290.46)	(8.25)	(5.03)	(11.51)
Drug and medical supply expenditure	18.49	38.08	1.91	1.64	0.72
($1,000,000s in 2012/13 constant price)	(37.18)	(46.80)	(0.80)	(0.68)	(0.67)
Non-admitted occasions of service (1,000s)	165.88	321.26	42.56	45.96	23.23
	(235.51)	(266.63)	(8.58)	(6.47)	(10.72)
Casemix weighted inpatient episodes	17.63	35.62	3.36	2.48	1.21
(1,000s)	(29.91)	(35.70)	(0.63)	(0.32)	(0.55)
Casemix weighted average length of stay	8.25	10.23	5.62	5.82	6.60
	(3.31)	(3.32)	(1.48)	(1.35)	(2.12)
Large hospitals	0.40	0.85	0.00	0.00	0.00
	(0.49)	(0.36)	(0.00)	(0.00)	(0.00)
Major city hospitals	0.19	0.41	0.00	0.00	0.00
	(0.39)	(0.49)	(0.00)	(0.00)	(0.00)
Teaching hospitals	0.32	0.63	0.04	0.00	0.05
	(0.47)	(0.48)	(0.20)	(0.00)	(0.22)
Occupancy rate	0.67	0.85	0.68	0.51	0.50
	(0.26)	(0.21)	(0.12)	(0.06)	(0.20)
Casemix index	0.92	1.02	0.79	0.80	0.83
	(0.22)	(0.25)	(0.11)	(0.05)	(0.11)
The ratio of outpatient to inpatient	13.97	11.05	10.14	15.12	17.20
	(6.60)	(3.15)	(2.62)	(3.39)	(7.95)
The proportion of unit producing	0.66	0.70	0.70	0.64	0.62
personnel	(0.07)	(0.05)	(0.04)	(0.05)	(0.07)
Number of observations	684	324	24	24	312

Notes: Means for the variables are reported with their standard deviations in parentheses.

in terms of inputs and outputs. However, there are also huge variations in quantities of inputs and outputs among hospitals in this cohort. Moreover, compared to the other cohorts, the cohort adopting ABF in FY 2011/12 has a larger proportion of hospitals having teaching status and located in major cities.

6 Results and discussion

Before discussing the results, it is worth recalling here that for dependent variables we utilize the weighted average length of stay and the Farrell-type output-oriented technical efficiency estimated from DEA models. As noted earlier in Section 3.2, in this study we consider the two most popular variants of DEA models: CRS-DEA and VRS-DEA.[28]

6.1 Exploratory analyses

Table 2 provides the descriptive statistics for the weighted average length of stay as well as the estimated efficiency scores from both CRS-DEA and VRS-DEA models of different groups of hospitals. Of particular interest are the ABF hospitals before and after the implementation of ABF.[29]

We can see that the mean value of the weighted average length of stay of ABF hospitals decreased slightly after adopting ABF. Meanwhile, the mean values of estimated efficiency scores from both CRS-DEA and VRS-DEA models increased slightly.[30] For a further comparison, we estimate the densities of these efficiency measures for ABF hospitals before and after the implementation of ABF. As can be seen from Figure 1, after the implementation of ABF, the estimated density of weighted average length of stay of ABF hospitals shifted slightly to the left. Meanwhile, the estimated densities of efficiency scores from both CRS-DEA and VRS-DEA models of these hospitals shifted slightly to the right after adopting ABF (see Figure 2 and Figure 3). Moreover, we perform the Li's (1996) test to test whether the changes in the densities are significant (see Table 3).[31] We find evidence about the significant changes in the densities of the weighted average length of stay and efficiency scores from VRS-DEA models, but we do not find significant evidence about the change in the density of efficiency scores from CRS-DEA models (at the conventional 5% level of significance).

To gain a better understanding of the temporal variation in hospital performance, we also compare the efficiency measures of non-ABF hospitals before and after the begin-

[28]We provide more discussions about the choice of reference technologies in the Online Supporting Information.

[29]Hereafter, we use the term "ABF hospitals" to refer to a group of all hospitals that were exposed to ABF (i.e., a combination of all ever-treated cohorts). Accordingly, the term "non-ABF hospitals" will be used to refer to a group of hospitals that have never been funded by ABF (i.e., the never-treated cohort).

[30]Note here that the technical efficiency scores are measured in output-orientation, so a hospital is said to be fully technically efficient if its efficiency score equals one, and the higher the value of these efficiency measures, the less efficient the hospital is.

[31]The application of kernel density estimation and the Li's (1996) test requires an adaptation to the context of DEA (see discussion in Simar and Zelenyuk, 2006).

ning of FY 2011/12 (i.e., July 2011).[32] From Table 2 we can see that the mean values of estimated efficiency scores from both CRS-DEA and VRS-DEA models of non-ABF hospitals in periods after July 2011 were substantially higher than those in periods before July 2011. Moreover, the density estimation and the adapted Li's (1996) test also show that the distribution of efficiency scores of non-ABF hospitals after July 2011 changed significantly (and became less favourable) compared to their densities before July 2011. Meanwhile, the weighted average length of stay of non-ABF hospitals appeared to be lower after July 2011 in both central tendency as well as the whole distribution.

We also compare these efficiency measures for different hospital groups characterised by size, teaching status and geographical locations. Specifically, large hospitals, teaching hospitals, and major city hospitals had lower mean values of estimated efficiency scores (from both CRS-DEA and VRS-DEA models) but higher mean values of weighted average length of stay compared to small hospitals, non-teaching hospitals, and regional and remote hospitals, respectively.

Table 2: Descriptive statistics of dependent variables

	No. of obs.	Weighted average length of stay		Estimated efficiency scores (CRS-DEA)		Estimated efficiency scores (VRS-DEA)	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
All sample	684	8.25	3.31	1.64	0.42	1.38	0.39
ABF hospitals before ABF	196	10.05	3.46	1.52	0.23	1.18	0.15
ABF hospitals after ABF	176	9.19	3.45	1.55	0.23	1.24	0.22
Non-ABF hospitals before July 2011	156	6.92	1.38	1.66	0.46	1.47	0.38
Non-ABF hospitals after July 2011	156	6.27	2.63	1.87	0.60	1.68	0.53
Large hospitals	276	10.73	3.30	1.51	0.22	1.15	0.13
Small hospitals	408	6.58	2.00	1.72	0.49	1.53	0.43
Major city hospitals	132	11.29	3.16	1.45	0.25	1.09	0.08
Regional and Remote hospitals	552	7.53	2.91	1.68	0.44	1.44	0.4
Teaching hospitals	220	11.37	3.31	1.55	0.25	1.16	0.18
Non-teaching hospitals	464	6.78	2.04	1.68	0.47	1.48	0.42

[32] Although the timing of the implementation of ABF varies across cohorts of hospitals, the majority of ABF hospitals started adopting the new funding model from the beginning of FY 2011/12, so for the purpose of comparison in this exploratory analysis, we look at the change in efficiency measures for non-ABF hospitals through this time.

Table 3: Li's (1996) test for ABF and non-ABF hospitals

	ABF hospitals: before vs. after ABF adoption		Non-ABF hospitals: before vs. after July 2011	
	Test statistics	p-value	Test statistics	p-value
Weighted average length of stay	2.272	0.011	13.982	0.000
Estimated efficiency scores (CRS-DEA)	0.990	0.134	3.062	0.001
Estimated efficiency scores (VRS-DEA)	4.350	0.000	5.351	0.000

Notes: Computations were done in the Matlab adopting the code from Simar and Zelenyuk (2006), with 2000 bootstrap replications using Gaussian kernel, and Silverman (1986) robust rule of thumb bandwidth.

6.2 Main regression results

Table 4 provides a summary of the main results from the regression analysis.[33] Note that for the technical efficiency measures, we report both estimated coefficients and average partial effects.[34] Meanwhile, for the weighted average length of stay (WLOS), only the estimated coefficients are reported since they are also the estimates of the corresponding average partial effects.

We first examine the impact of activity based funding on the weighted average length of stay. From the first column of Table 4, we can see that the coefficient of the policy dummy is not statistically different from zero. However, we cannot conclude about the causal effect since the parallel trends assumption appeared not satisfied. (We provide results of the test for parallel trends assumption and other robustness checks in the Online Supporting Information.)

We next turn to the impact of activity based funding on the technical efficiency measures. From the estimated results in Table 4, we find evidence that ABF helped to improve hospital efficiency. Specifically, if a hospital adopted ABF, its technical efficiency score reduced by 5.4% (if benchmarking to the CRS reference technology) or 4.8% (if benchmarking to the VRS reference technology) compared to without being funded by ABF, on average and ceteris paribus. Our finding is consistent with the empirical evidence in Norway (Biørn et al., 2003; Biørn et al., 2010), Sweden (Gerdtham, Löthgren, Tambour and Rehnberg, 1999; Gerdtham, Rehnberg and Tambour, 1999), and Italy (Cavalieri

[33]Full regression results (e.g., with estimated coefficients for year dummies and Mundlak adjustments) are reported in the Online Supporting Information.

[34]Consistent with the discussion in Section 4, Table 4 shows that the estimated coefficients overestimate the corresponding average partial effects due to the truncated nature of the error terms.

et al., 2018), where the implementation of ABF was found to positively correlate with the efficiency improvement. However, it is worth noting here that the empirical findings in those studies were about association rather than causal as in our study.

Regarding the time-invariant characteristics of hospitals, the coefficients and average partial effects of the major city dummy were negative and significantly different from zero at the conventional 5% level of significance in both the cases of CRS-DEA and VRS-DEA models. As a result, we conclude that hospitals located in major cities were, on average and ceteris paribus, more efficient than hospitals located in regional and remote areas. Our findings about the relationship between geographical location and hospital efficiency is consistent with the findings of the other studies undertaken in the Australian context. For example, Paul (2002) utilized stochastic frontier analysis (SFA) to study public hospitals in New South Wales, Australia and found evidence that hospitals operating in urban areas appeared to have a higher level of efficiency than rural hospitals. Similarly, when using SFA with the output distance function, a Productivity Commission study (Productivity Commission, 2010) revealed that Australian public hospitals were more productive if they operated in areas being relatively close to a major city.[35] We do not find significant evidence about the impact of teaching status on hospital efficiency after controlling for other factors. Meanwhile, hospital size was found to positively correlate with the efficiency level in the VRS-DEA model, but was not significantly associated with the efficiency level in the CRS-DEA model (at the conventional 5% level of significance), on average and ceteris paribus.

On the time-varying covariates, since we include the time average of these variables into the regression model as Mundlak adjustment terms, which control for the between-hospital variation, the coefficients of these variables (and the corresponding partial effects) can be interpreted as the impact of the within-hospital variation. As expected, we find statistical evidence (at the conventional 5% level of significance) that on average and ceteris paribus, an increase in the ratio of outpatient to inpatient or an increase in the occupancy rate of a hospital was positively correlated with the hospital's improvement in efficiency (for both CRS-DEA and VRS-DEA models). Meanwhile, the casemix index was negatively associated with the efficiency level in the CRS-DEA model, but was not significantly correlated with the efficiency level in the VRS-DEA model (at the conventional 5% level of significance), on average and ceteris paribus. For the unit producing personnel ratio, it was found to positively correlate with the efficiency level in the VRS-DEA model. For the CRS-DEA model, the positive relationship between hospital efficiency and the unit producing personnel ratio was captured mainly by the between-hospital variation, and we

[35]Note that these studies utilized different datasets at different points in time and with different methods.

Table 4: Regression results - Technical efficiency and WLOS

	WLOS (log)	Technical efficiency (CRS-DEA, log)		Technical efficiency (VRS-DEA, log)	
	Coeff.	Coeff.	APE	Coeff.	APE
Acitivity-based funding	-0.002	-0.056^{***}	-0.054^{***}	-0.064^{***}	-0.048^{**}
	(0.038)	(0.013)	(0.013)	(0.022)	(0.017)
Major city hospitals	-0.090	-0.131^{***}	-0.125^{***}	-0.117^{**}	-0.091^{**}
	(0.075)	(0.039)	(0.036)	(0.051)	(0.038)
Large hospitals	0.163	-0.018	-0.018	-0.130^{**}	-0.106^{**}
	(0.100)	(0.030)	(0.029)	(0.056)	(0.046)
Teaching hospitals	-0.033	0.005	0.005	-0.046	-0.037
	(0.049)	(0.010)	(0.010)	(0.030)	(0.024)
The ratio of outpatient to	-0.140	-0.467^{***}	-0.450^{***}	-0.279^{***}	-0.223^{***}
inpatient (log)	(0.161)	(0.066)	(0.063)	(0.102)	(0.082)
The proportion of unit	0.111^{**}	-0.019	-0.018	-0.199^{***}	-0.159^{***}
producing personnel (log)	(0.054)	(0.017)	(0.016)	(0.029)	(0.023)
Casemix index (log)	0.328^{**}	0.150^{**}	0.144^{**}	0.130	0.104
	(0.146)	(0.051)	(0.049)	(0.085)	(0.068)
Occupancy rate (log)	0.234^{***}	-0.163^{***}	-0.157^{***}	-0.086^{**}	-0.069^{**}
	(0.051)	(0.023)	(0.022)	(0.033)	(0.026)
Constant	2.532^{***}	0.331^{**}	n.a	0.498^{***}	n.a
	(0.213)	(0.107)	n.a	(0.124)	n.a
$\hat{\sigma}_a$	$0.281^{n.a}$	0.138^{***}	n.a	0.096^{***}	n.a
	n.a	(0.013)	n.a	(0.013)	n.a
$\hat{\sigma}_\varepsilon$	$0.172^{n.a}$	0.124^{***}	n.a	0.130^{***}	n.a
	n.a	(0.003)	n.a	(0.005)	n.a
Number of observations	684		684		684
LogLikelihood	n.a		394.177		513.912
AIC	n.a		-734.353		-973.824
BIC	n.a		-612.098		-851.569
R-square	0.296		n.a		n.a
Adjusted R-square	0.268		n.a		n.a

Notes: (i) The models for the technical efficiency were estimated using the adapted double bootstrap procedure with $B_1 = 200$ iterations for the first round of bootstrapping and $B_2 = 2000$ iterations for the second round of bootstrapping. The computations were done using Matlab codes programmed by the authors, involving standard Matlab library and some codes adopting from Matlab codes of Léopold Simar. Both coefficients and average partial effects (APE) are reported. Bootstrapped standard errors are reported in the parentheses. (ii) The model for the weighted average length of stay (WLOS) was estimated using Feasible Generalized Least Square estimator, utilizing the `plm` package in R (Croissant and Millo, 2008). Clustered standard errors (clustering at individual level) are reported in the parentheses. (iii) * p-value < 0.1, ** p-value < 0.05, ***p-value < 0.01.

do not find significant evidence about the relationship between hospital efficiency and the within-hospital variation in the unit producing personnel ratio.[36]

7 Concluding remarks

This study investigates the impact of the ABF reform on the performance of public hospitals in Queensland, Australia in the period from FY 2005/06 to FY 2016/17. Specifically, to identify the causal effect, we exploit a natural experiment that only a fraction of public hospitals was exposed to the new funding method and at different points in time. To do so, we develop the difference-in-differences method for the popular two-stage DEA framework with the double bootstrapped truncated regression of Simar and Wilson (2007). We also adapt the framework to the context of panel data to account for individual heterogeneity.

We find evidence that the implementation of ABF had a positive impact on hospital efficiency. Specifically, based on difference-in-differences analysis, if a hospital was funded by ABF, our model and data suggest that, on average and ceteris paribus, such a hospital tended to become more efficient than without being funded by ABF. We also find that remote and regional hospitals were, on average and ceteris paribus, less efficient than hospitals located in cities. Other organizational characteristics were also found to be significantly associated with hospital efficiency (or inefficiency). Occupancy rate, the ratio of outpatient volumes to inpatient volumes, and the proportion of producing personnel to total staff were positively correlated with hospitals' efficiency. Meanwhile, the relationships between hospital efficiency and hospital size as well as between hospital efficiency and casemix index were found to depend on the assumption about the returns to scale of the reference technology.

Our findings have policy implication for health policymakers. In Queensland, there are still many public hospitals (predominantly small and rural hospitals) being funded through block grants. It is under the decision of the health authorities in Queensland since the NHRA allows states to select either ABF or block funding for small hospitals who meet the "low volume thresholds." Our findings show that the application of ABF helped to improve the performance of public hospitals. This suggests that it might be beneficial to implement ABF for small hospitals (when it is practically possible); further investigation may be worthwhile for Queensland Health to address this issue.

[36]Care must be undertaken when making inference about the impact of between- and within-hospital variations of these variables since they are highly correlated with their Mundlak terms, and thus subject to multicollinearity issues. As such, it might not be possible to separate the effects of their between- and within-hospital variations. (See the Online Supporting Information for the analysis of multicollinearity.) Also, it is worth noting here that the multicollinearity among these covariates does not impact the inference for the policy variable as well as other time-invariant variables discussed above.

The main limitation of this study, similar to many other studies in the literature, is the non-feasibility to account for the potential yet unobserved difference in output quality when investigating hospital efficiency due to a lack of available data. Another limitation of this study relates to the two-stage regression results. As discussed in Simar and Wilson (2007), an important requirement for a valid two-stage DEA approach is the separability assumption that environmental variables influence efficiency distribution but do not alter the production frontier. Our a priori goal was to measure efficiency with respect to the unconditional frontier of the observed best practice, where this assumption is satisfied by definition. Alternative views may require analysis of conditional frontiers, where various conditional variables may be allowed to influence not only the efficiency but also the frontier (e.g., Bădin et al., 2012; Simar et al., 2017). Such an approach would require consideration and testing of many conditional variables (and their various combinations) as potential 'influencers' of both efficiency and frontiers and is a subject in itself which we leave for future endeavours.

References

Australian Institute of Health and Welfare (2015), 'Australian Hospital Peer Groups'. Last visited: 10 August 2018.

Bădin, L., Daraio, C. and Simar, L. (2012), 'How to measure the impact of environmental factors in a nonparametric production model', *European Journal of Operational Research* **223**(3), 818–833.

Badunenko, O. and Tauchmann, H. (2019), 'Simar and Wilson two-stage efficiency analysis for Stata', *The Stata Journal* **19**(4), 950–988.

Böcking, W., Ahrens, U., Kirch, W. and Milakovic, M. (2005), 'First results of the introduction of DRGs in Germany and overview of experience from other DRG countries', *Journal of Public Health* **13**(3), 128–137.

Berta, P., Callea, G., Martini, G. and Vittadini, G. (2010), 'The effects of upcoding, cream skimming and readmissions on the Italian hospitals efficiency: A population-based investigation', *Economic Modelling* **27**(4), 812–821.

Besstremyannaya, G. (2013), 'The impact of Japanese hospital financing reform on hospital efficiency: A difference-in-difference approach', *The Japanese Economic Review* **64**(3), 337–362.

Biørn, E., Hagen, T. P., Iversen, T. and Magnussen, J. (2010), 'How different are hospitals' responses to a financial reform? the impact on efficiency of activity-based financing', *Health Care Management Science* **13**(1), 1–16.

Biørn, E., Hagen, T. P., Iversen, T. and Magnussen, J. (2003), 'The effect of Activity-Based Financing on hospital efficiency: A Panel Data Analysis of DEA efficiency scores 1992–2000', *Health Care Management Science* **6**(4), 271–283.

Borden, J. P. (1988), 'An assessment of the impact of diagnosis-related group (DRG)-based reimbursement on the technical efficiency of New Jersey hospitals using data envelopment analysis', *Journal of Accounting and Public Policy* **7**(2), 77–96.

Burgess, J. F. and Wilson, P. W. (1996), 'Hospital ownership and technical inefficiency', *Management Science* **42**(1), 110–123.

Burgess, J. F. and Wilson, P. W. (1998), 'Variation in inefficiency among US hospitals', *INFOR: Information Systems and Operational Research* **36**(3), 84–102.

Busse, R., Schreyögg, J. and Smith, P. C. (2006), 'Hospital case payment systems in europe', *Health Care Management Science* **9**(3), 211–213.

Callaway, B. and Sant'Anna, P. H. (2021), 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2), 200–230.

Card, D. and Krueger, A. B. (1994), 'Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania', *American Economic Review* **84**(4), 772–93.

Cavalieri, M., Guccio, C., Lisi, D. and Pignataro, G. (2018), 'Does the extent of per case payment system affect hospital efficiency? evidence from the Italian NHS', *Public Finance Review* **46**(1), 117–149.

Chern, J.-Y. and Wan, T. T. (2000), 'The impact of the prospective payment system on the technical efficiency of hospitals', *Journal of Medical Systems* **24**(3), 159–172.

Chilingerian, J. A. (1995), 'Evaluating physician efficiency in hospitals: A multivariate analysis of best practices', *European Journal of Operational Research* **80**(3), 548–574.

Chilingerian, J. A. and Sherman, H. D. (2004), Health-care applications: From hospitals to physicians, from productive efficiency to quality frontiers, *in* W. W. Cooper, L. M. Seiford and J. Zhu, eds, 'Handbook on Data Envelopment Analysis', Springer US, pp. 445–493.

Chowdhury, H. and Zelenyuk, V. (2016), 'Performance of hospital services in Ontario: DEA with truncated regression approach', *Omega* **63**, 111–122.

Chua, C. L., Palangkaraya, A. and Yong, J. (2011), 'Hospital competition, technical efficiency and quality', *Economic Record* **87**(277), 252–268.

Clement, J. P., Valdmanis, V. G., Bazzoli, G. J., Zhao, M. and Chukmaitov, A. (2008), 'Is more better? an analysis of hospital outcomes and efficiency with a DEA model of output congestion', *Health Care Management Science* **11**(1), 67–77.

Croissant, Y. and Millo, G. (2008), 'Panel data econometrics in r: The plm package', *Journal of Statistical Software* **27**(2).

Daraio, C. and Simar, L. (2007), Economies of scale, scope and experience in the italian motor-vehicle sector, *in* C. Daraio and L. Simar, eds, 'Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications', Springer Science & Business Media, pp. 135–165.

de Chaisemartin, C. and d'Haultfoeuille, X. (2020), 'Two-way fixed effects estimators with heterogeneous treatment effects', *American Economic Review* **110**(9), 2964–96.

de Chaisemartin, C. and D'Haultfoeuille, X. (2022), 'Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey', *NBER Working Paper* .

Duckett, S. J. (1995), 'Hospital payment arrangements to encourage efficiency: The case of Victoria, Australia', *Health Policy* **34**(2), 113–134.

Färe, R. and Lovell, C. K. (1978), 'Measuring the technical efficiency of production', *Journal of Economic Theory* **19**(1), 150–162.

Farrar, S., Yi, D., Sutton, M., Chalkley, M., Sussex, J. and Scott, A. (2009), 'Has payment by results affected the way that English hospitals provide care? difference-in-differences analysis', *British Medical Journal* **339**(7720), 554–556.

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

Ferrier, G. D. and Trivitt, J. S. (2013), 'Incorporating quality into the measurement of hospital efficiency: A double DEA approach', *Journal of Productivity Analysis* **40**(3), 337–355.

Garcia-Lacalle, J. and Martin, E. (2010), 'Rural vs urban hospital performance in a 'competitive' public health service', *Social Science & Medicine* **71**(6), 1131–1140.

Gerdtham, U.-G., Löthgren, M., Tambour, M. and Rehnberg, C. (1999), 'Internal markets and health care efficiency: a multiple-output stochastic frontier analysis', *Health Economics* **8**(2), 151–164.

Gerdtham, U.-G., Rehnberg, C. and Tambour, M. (1999), 'The impact of internal markets on health care efficiency: evidence from health care reforms in Sweden', *Applied Economics* **31**(8), 935–945.

Goodman-Bacon, A. (2021), 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* . forthcoming.

Government of Queensland (2012), 'State Budget 2012-13 Service Delivery Statements Queensland Health'. Last visited: 10 August 2018.

Grosskopf, S., Margaritis, D. and Valdmanis, V. (2001), 'Comparing teaching and non-teaching hospitals: A frontier approach (teaching vs. non-teaching hospitals)', *Health Care Management Science* **4**(2), 83–90.

Guterman, S., Eggers, P. W., Riley, G., Greene, T. F. and Terrell, S. A. (1988), 'The first 3 years of Medicare prospective payment: an overview', *Health Care Financing Review* **9**(3), 67–77.

Hall, J. (2010), 'Health-care reform in Australia: Advancing or side-stepping?', *Health Economics* **19**(11), 1259–1263.

Hao, S. S. and Pegels, C. C. (1994), 'Evaluating relative efficiencies of Veterans Affairs medical centers using Data Envelopment, ratio, and multiple regression analysis', *Journal of Medical Systems* **18**(2), 55–67.

Harris, J., Ozgen, H. and Ozcan, Y. (2000), 'Do mergers enhance the performance of hospital efficiency?', *The Journal of the Operational Research Society* **51**(7), 801–811.

Herwartz, H. and Strumann, C. (2014), 'Hospital efficiency under prospective reimbursement schemes: An empirical assessment for the case of Germany', *The European Journal of Health Economics* **15**(2), 175–186.

Hofmarcher, M. M., Paterson, I. and Riedel, M. (2002), 'Measuring hospital efficiency in Austria – a DEA approach', *Health Care Management Science* **5**(1), 7–14.

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

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

Hu, H. H., Qi, Q. and Yang, C. H. (2012), 'Evaluation of China's regional hospital efficiency: DEA approach with undesirable output', *Journal of the Operational Research Society* **63**(6), 715–725.

Imbens, G. W. and Wooldridge, J. M. (2009), 'Recent developments in the econometrics of program evaluation', *Journal of Economic Literature* **47**(1), 5–86.

Jacobson, L. S., LaLonde, R. J. and Sullivan, D. G. (1993), 'Earnings losses of displaced workers', *The American Economic Review* **83**(4), 685–709.

Kirigia, J. M. and Asbu, E. Z. (2013), 'Technical and scale efficiency of public community hospitals in Eritrea: An exploratory study', *Health Economics Review* **3**(1), 1–16.

Kjerstad, E. (2003), 'Prospective funding of general hospitals in Norway—incentives for higher production?', *International Journal of Health Care Finance and Economics* **3**(4), 231–251.

Kneip, A., Simar, L. and 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. and 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.

Korostelev, A., Simar, L. and Tsybakov, A. B. (1995*a*), 'Efficient estimation of monotone boundaries', *Annals of Statistics* **23**(2), 476–489.

Korostelev, A., Simar, L. and Tsybakov, A. B. (1995*b*), 'On estimation of monotone and convex boundaries', *Publications de l'Institut de Statistique des Universités de Paris* **39**(1), 3–18.

Kutner, M. H., Nachtsheim, C. J. and Neter, J. (2004), *Applied Linear Regression Models*, McGraw-Hill/Irwin.

Li, Q. (1996), 'Nonparametric testing of closeness between two unknown distribution functions', *Econometric Reviews* **15**(3), 261–274.

Litvak, E. and Bisognano, M. (2011), 'More patients, less payment: Increasing hospital efficiency in the aftermath of health reform', *Health Affairs* **30**(1), 76–80.

Magnussen, J. (1996), 'Efficiency measurement and the operationalization of hospital production', *Health Services Research* **31**(1), 21–37.

McCallion, G., McKillop, D. G., Glass, J. C. and Kerr, C. (1999), 'Rationalizing Northern Ireland hospital services towards larger providers: Best-practice efficiency studies and current policy', *Public Money & Management* **19**(2), 27–32.

Mick, S. S. and Morlock, L. L. (1990), 'America's rural hospitals: A selective review of 1980s research', *The Journal of Rural Health* **6**(4), 437–466.

Moreno-Serra, R. and Wagstaff, A. (2010), 'System-wide impacts of hospital payment reforms: evidence from central and eastern europe and central asia', *Journal of Health Economics* **29**(4), 585–602.

Mundlak, Y. (1978), 'On the pooling of time series and cross section data', *Econometrica* **46**(1), 69–85.

Munson, F. C. and Zuckerman, H. S. (1983), The managerial role, *in* S. M. Shortell and A. D. Kaluzny, eds, 'Health Care Management: A Text in Organization Theory and Behavior', John Wiley & Sons, pp. 48–58.

Nayar, P. and Ozcan, Y. A. (2008), 'Data Envelopment Analysis comparison of hospital efficiency and quality', *Journal of Medical Systems* **32**(3), 193–199.

Nayar, P., Ozcan, Y. A., Yu, F. and Nguyen, A. T. (2013), 'Benchmarking urban acute care hospitals: Efficiency and quality perspectives', *Health Care Management Review* **38**(2), 137–145.

Nghiem, S., Coelli, T. J. and Barber, S. (2011), 'Sources of productivity growth in health services: A case study of Queensland public hospitals', *Economic Analysis and Policy* **41**(1), 37–48.

Nguyen, B. H. and Zelenyuk, V. (2021), 'Aggregate efficiency of industry and its groups: The case of Queensland public hospitals', *Empirical Economics* **60**(6), 2795–2836.

O'Donnell, C. J. and Nguyen, K. (2013), 'An econometric approach to estimating support prices and measures of productivity change in public hospitals', *Journal of Productivity Analysis* **40**(3), 323–335.

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

Parliament of Australia (2018), 'Recent developments in federal government funding for public hospitals: a quick guide'. Last visited: 23 September 2021.

Paul, C. J. M. (2002), Productive structure and efficiency of public hospitals, *in* K. J. Fox, ed., 'Efficiency in the Public Sector', Springer, Boston, MA, pp. 219–248.

Productivity Commission (2010), 'Public and Private Hospital: Multivariate Analysis'. Last visited: 10 August 2018.

Queensland Health (2017), 'Health Funding Principles and Guidelines 2017-18'. Last visited: 10 August 2018.

Roth, J., Sant'Anna, P. H., Bilinski, A. and Poe, J. (2022), 'What's trending in difference-in-differences? a synthesis of the recent econometrics literature', *arXiv preprint* .

Sickles, R. C. and Zelenyuk, V. (2019), *Measurement of Productivity and Efficiency*, Cambridge University Press, New York.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London ; New York.

Simar, L., Van Keilegom, I. and Zelenyuk, V. (2017), 'Nonparametric least squares methods for stochastic frontier models', *Journal of Productivity Analysis* **47**(3), 189–204.

Simar, L. and Wilson, P. W. (2007), 'Estimation and inference in two-stage, semi-parametric models of production processes', *Journal of Econometrics* **136**(1), 31–64.

Simar, L. and Zelenyuk, V. (2006), 'On testing equality of distributions of technical efficiency scores', *Econometric Reviews* **25**(4), 497–522.

Staat, M. (2006), 'Efficiency of hospitals in Germany: A DEA-bootstrap approach', *Applied Economics* **38**(19), 2255–2263.

Street, A., O'Reilly, J., Ward, P. and Mason, A. (2011), DRG-based hospital payment and efficiency: theory, evidence, and challenges, *in* R. Busse, A. Geissler, W. Quentin and M. M. Wiley, eds, 'Diagnosis-related groups in Europe: Moving towards transparency, efficiency and quality in hospitals', Open University Press, pp. 93–114.

Sun, L. and Abraham, S. (2021), 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *Journal of Econometrics* **225**(2), 175–199.

Theurl, E. and Winner, H. (2007), 'The impact of hospital financing on the length of stay: evidence from Austria', *Health Policy* **82**(3), 375–389.

Tiemann, O. and Schreyögg, J. (2009), 'Effects of ownership on hospital efficiency in Germany', *Academy of Management Proceedings* **2009**(1), 1–6.

Watcharasriroj, B. and Tang, J. C. (2004), 'The effects of size and information technology on hospital efficiency', *The Journal of High Technology Management Research* **15**(1), 1–16.

Weisgrau, S. (1995), 'Issues in rural health: Access, hospitals, and reform', *Health Care Financing Review* **17**(1), 1–14.

Wooldridge, J. M. (2010), *Econometric analysis of cross section and panel data*, MIT press.

Wooldridge, J. M. (2019), 'Correlated random effects models with unbalanced panels', *Journal of Econometrics* **211**(1), 137–150.

Wooldridge, J. M. (2021), 'Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators', *Available at SSRN 3906345* .

Worthington, A. C. (2004), 'Frontier efficiency measurement in health care: A review of empirical techniques and selected applications', *Medical Care Research and Review* **61**(2), 135–170.

Yong, K. and Harris, A. H. (1999), *Efficiency of Hospitals in Victoria under Casemix Funding: A Stochastic Frontier Approach*, Centre for Health Program Evaluation.

Zuckerman, S., Hadley, J. and Lezzoni, L. (1994), 'Measuring hospital efficiency with frontier cost functions', *Journal of Health Economics* **13**(3), 255–280.

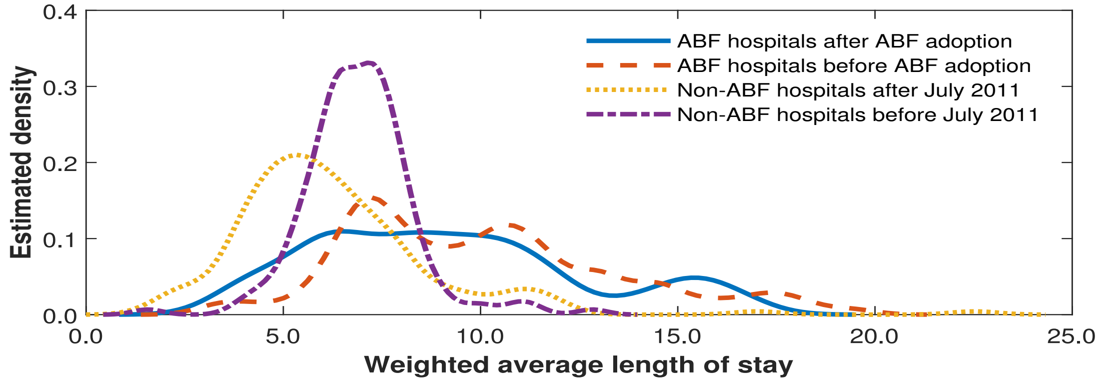


Figure 1: Estimated density of weighted average length of stay for ABF and non-ABF hospitals

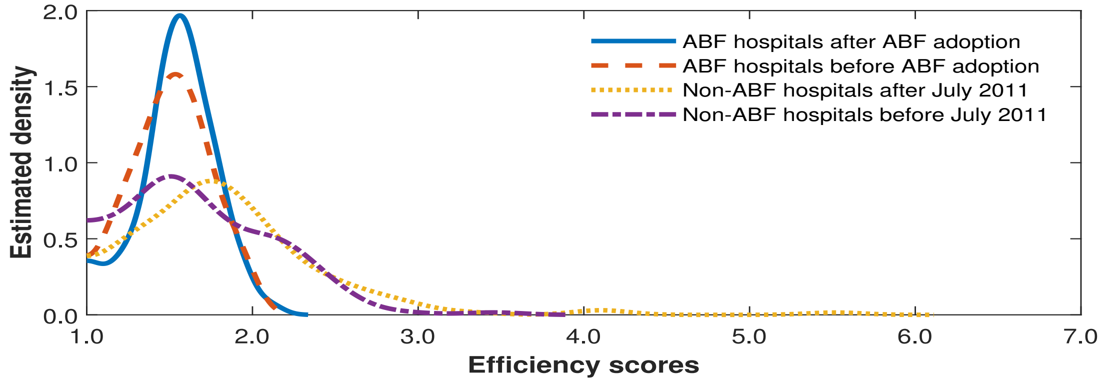


Figure 2: Estimated density of efficiency scores estimated from CRS-DEA model for ABF and non-ABF hospitals

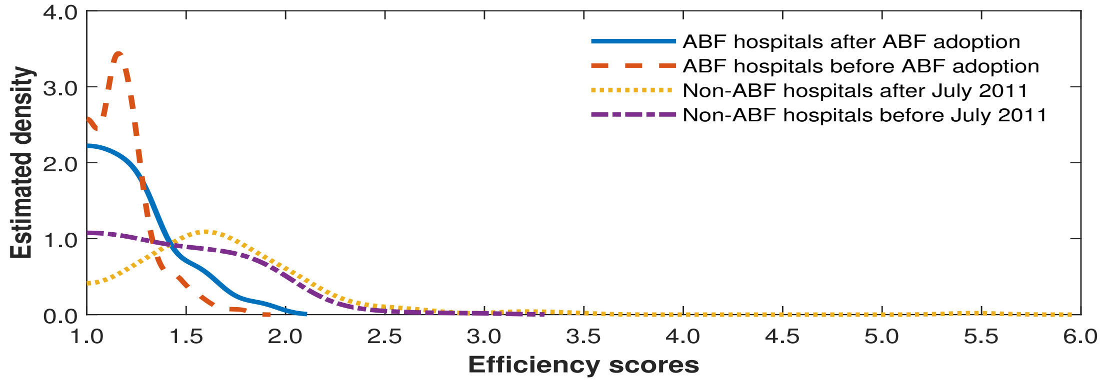


Figure 3: Estimated density of efficiency scores estimated from VRS-DEA model for ABF and non-ABF hospitals

Activity based funding reform and the performance of public hospitals: The Case of Queensland, Australia

Bao Hoang Nguyen* Shawna Grosskopf† Jongsay Yong‡
Valentin Zelenyuk§

Online Supporting Information

*Corresponding author. School of Economics and Centre for the Business and Economics of Health, University of Queensland, Brisbane, Qld 4072, Australia. Email address: b.h.nguyen@uq.edu.au.

†Economics, School of Public Policy, Oregon State University, Corvallis, OR 97331, USA

‡Melbourne Institute of Applied Economic and Social Research, The University of Melbourne, Parkville, Australia

§School of Economics and Centre for Efficiency and Productivity Analysis, University of Queensland, Brisbane, Qld 4072, Australia

1 Robustness Checks

1.1 Event-study setup and parallel trends test

As discussed in Section 4, the key assumption for the identification of the causal effect in the difference-in-differences framework is the parallel trends assumption. Thus, to ensure the robustness of our main results, in this section we examine the parallel trends assumption by looking at pretreatment trends for both outcome variables: the technical efficiency scores and the weighted average lengths of stay. Also, because the timing of ABF adoption varies across hospital cohorts, we examine the pretrends using an event study setup (Jacobson et al., 1993). Specifically, we estimate the following model

$$\eta_{it} = c_0 + \sum_{\ell=-4,\ell\neq-1}^{\ell=5} \delta_\ell L_{it}^{\ell} + W_i'\zeta + Z_{it}'\beta + \sum_{t=2}^{T} \alpha_t Year_t + \sum_{g\in\mathcal{G}} \lambda_g G_g + \overline{Z_{i\cdot}}'\tau + a_i + \varepsilon_{it}, \tag{1.1}$$

where η_{it} is the logarithm of the outcome variables, and other covariates are the same as those described in equation (4.4) in Section 4, except for L_{it}^{ℓ}. L_{it}^{ℓ} is a dummy variable representing the duration that hospital i at the financial year t has been funded by ABF. Specifically, if denoting g_i as the financial year that hospital i started adopting ABF, we defined $L_{it}^{-4} = \mathbb{1}\{t - g_i + 1 \leq -4\}$, $L_{it}^{\ell} = \mathbb{1}\{t - g_i + 1 = \ell\}$ for $\ell = -3, \ldots, 4$, and $L_{it}^{5} = \mathbb{1}\{t - g_i + 1 \geq 5\}$, where $\mathbb{1}\{A\}$ is an indicator function taking value 1 if the statement A is true and 0 otherwise.[1] For non-ABF hospitals, $L_{it}^{\ell} = 0$ for all t and ℓ.[2] In this event study setup, the parallel trends assumption is not violated if $\delta_\ell = 0$ for all $\ell \leq 0$.

Table 1 presents the regression results and Figure 1 provides a dynamic visualization for the results by plotting point estimates of coefficients of the dummies (together with their 95% confidence interval) against the time duration since ABF adoption. For the technical efficiency for both the CRS-DEA and VRS-DEA models, we do not find any statistical evidence about the violation of parallel trends assumptions since all the estimates of δ_ℓ for $\ell \leq 0$ were not significantly different from zero (at the conventional 5% level of significance). Meanwhile, for the weighted average lengths of stay, the parallel trends assumption seemed to be violated since the estimates of δ_{-2} and δ_0 were significantly different from zero.

[1] We calculate the duration that a hospital has been exposed to ABF as $t - g_i + 1$ because the adoption of ABF starts at the beginning of a financial year, but the data are collected at the end of the financial year.

[2] As a convention, in (1.1) the dummy for $\ell = -1$ is omitted to avoid perfect multicollinearity.

Table 1: Regression results for event study setup

	WLOS (log)	Technical efficiency (CRS-DEA, log)		Technical efficiency (VRS-DEA, log)	
	Coeff.	Coeff.	APE	Coeff.	APE
$\geq$ 4 years before	-0.053	-0.030	-0.029	0.012	0.010
	(0.047)	(0.027)	(0.026)	(0.029)	(0.023)
3 years before	-0.060	0.019	0.019	0.002	0.001
	(0.043)	(0.030)	(0.030)	(0.008)	(0.006)
2 years before	-0.146^{***}	-0.052	-0.050	-0.021	-0.017
	(0.043)	(0.031)	(0.030)	(0.037)	(0.029)
Starting adoption	-0.122^{**}	-0.029	-0.028	-0.006	-0.005
	(0.058)	(0.029)	(0.028)	(0.023)	(0.018)
1 year after	-0.116^{**}	-0.070^{**}	-0.066^{**}	-0.034	-0.027
	(0.057)	(0.031)	(0.029)	(0.035)	(0.027)
2 years after	-0.067	-0.080^{**}	-0.077^{**}	-0.024	-0.019
	(0.064)	(0.032)	(0.030)	(0.036)	(0.028)
3 years after	-0.090	-0.084^{**}	-0.080^{**}	-0.053	-0.041
	(0.074)	(0.032)	(0.030)	(0.038)	(0.029)
4 years after	-0.032	-0.083^{**}	-0.079^{**}	-0.069	-0.053^{*}
	(0.070)	(0.034)	(0.031)	(0.040)	(0.030)
$\geq$ 5 years later	-0.054	-0.088^{**}	-0.084^{**}	-0.091^{**}	-0.070^{***}
	(0.056)	(0.030)	(0.028)	(0.035)	(0.025)
Number of observations	684		684		684
LogLikelihood	n.a		397.272		519.352
AIC	n.a		-724.545		-968.704
BIC	n.a		-566.066		-810.226
R-square	0.300		n.a		n.a
Adjusted R-square	0.264		n.a		n.a

Notes: (i) The models for the technical efficiency were estimated using the adapted double bootstrap procedure with $B_1 = 200$ iterations for the first round of bootstrapping and $B_2 = 2000$ iterations for the second round of bootstrapping. The computations were done using Matlab codes programmed by the authors, involving standard Matlab library and some codes adopting from Matlab codes of Léopold Simar. Both coefficients and average partial effects (APE) are reported. Bootstrapped standard errors are reported in the parentheses. (ii) The model for the weighted average length of stay (WLOS) was estimated using Feasible Generalized Least Square estimator, utilizing the `plm` package in R (Croissant and Millo, 2008). Clustered standard errors (clustering at individual level) are reported in the parentheses. (iii) * p-value < 0.1, ** p-value < 0.05, ***p-value < 0.01.

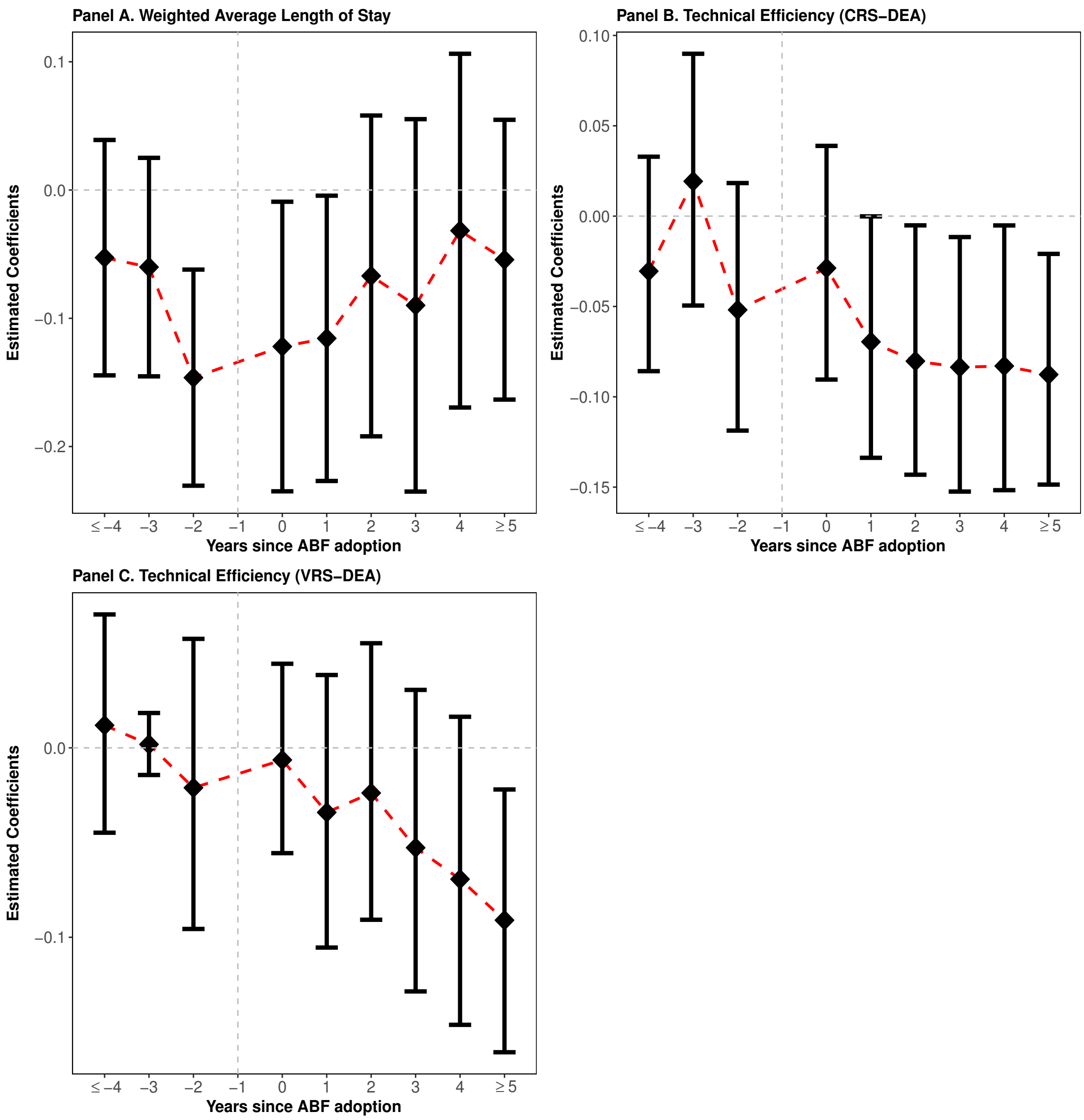


Figure 1: Event-study plots for technical efficiency and weighted average lengths of stay

1.2 Other specifications of DEA Models

In this section, we examine the robustness of our main results with respect to different specifications of DEA models. Specifically, we examine the sensitivity of our results with respect to the labour input aggregation. Recalling that to mitigate the 'curse of dimensionality' issue, we utilize the approach suggested by Daraio and Simar (2007) to aggregate the FTE staff in six labour categories into a single measure of labour input. Now, instead of aggregating all personnel into a single labour input measure, we aggregate the FTE staff into two different measures of labour input. The first labour input measure is an aggregation of medical officers, nurses, diagnostic and health professionals, and other personal care staff. Meanwhile, the second labour input measure is an aggregation of administrative and clerical staff, and domestic and other staff. We then use these two measures of labour input in the CRS-DEA model and the VRS-DEA model.

The results for the alternative specifications of DEA models are reported in Table 2. We can see that the CRS-DEA model is more robust (compared to the VRS-DEA model): the estimated partial effect of ABF only changed slightly compared to our main results. Meanwhile, for the VRS-DEA model, although the estimated partial effect of ABF decreased in magnitude, it was still negative and significant at 10% level of significance. (Also note that the parallel trends assumption was not violated for the technical efficiency with this specification of DEA models as illustrated in Figure 2).

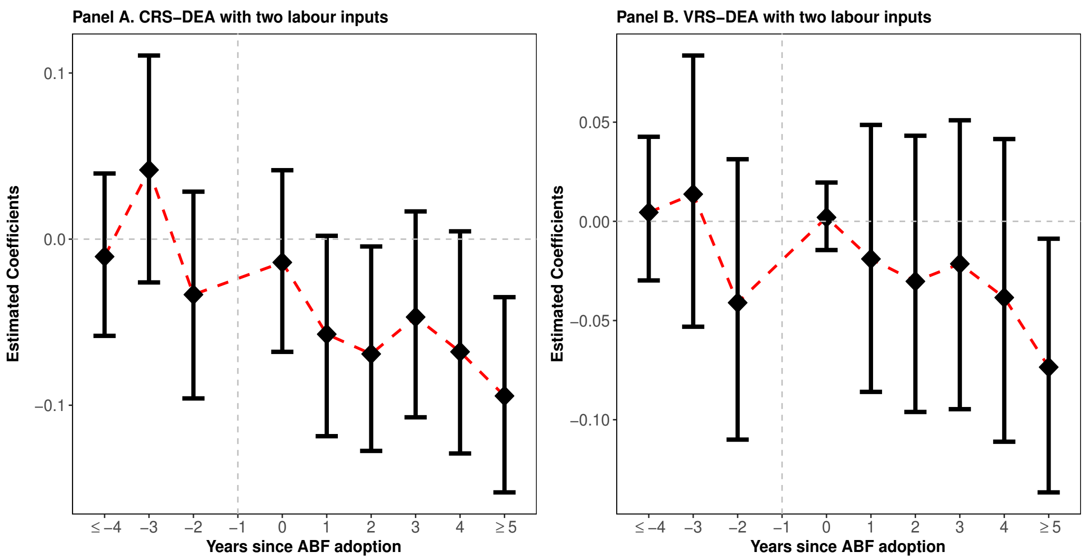


Figure 2: Event-study plots for the technical efficiency with alternative specifications of DEA models

Table 2: Regression results for the technical efficiency alternative specifications of DEA models

	CRS-DEA with two labour inputs		CRS-DEA with two labour inputs	
	Coeff.	APE	Coeff.	APE
Acitivity-based funding	−0.062***	−0.060***	−0.036*	−0.026*
	(0.018)	(0.017)	(0.018)	(0.014)
Major city hospitals	−0.127**	−0.119**	−0.139***	−0.104***
	(0.053)	(0.049)	(0.042)	(0.030)
Large hospitals	−0.005	−0.005	−0.140***	−0.112***
	(0.014)	(0.013)	(0.045)	(0.036)
Teaching hospitals	0.030	0.029	−0.010	−0.008
	(0.025)	(0.024)	(0.024)	(0.019)
The ratio of outpatient to inpatient (log)	−0.560***	−0.536***	−0.471***	−0.369***
	(0.090)	(0.086)	(0.083)	(0.066)
The proportion of unit producing personnel (log)	−0.014	−0.014	−0.178***	−0.139***
	(0.024)	(0.023)	(0.024)	(0.018)
Casemix index (log)	0.111	0.106	0.059	0.047
	(0.068)	(0.065)	(0.072)	(0.057)
Occupancy rate (log)	−0.117***	−0.112***	−0.073**	−0.057**
	(0.029)	(0.028)	(0.028)	(0.022)
Constant	0.258*	n.a	0.399***	n.a
	(0.138)	n.a	(0.087)	n.a
$\hat{\sigma}_a$	0.124***	n.a	0.102***	n.a
	(0.017)	n.a	(0.013)	n.a
$\hat{\sigma}_\varepsilon$	0.125***	n.a	0.130***	n.a
	(0.004)	n.a	(0.004)	n.a
Number of observations		684		684
LogLikelihood		397.270		526.133
AIC		−740.540		−998.265
BIC		−618.286		−876.010

Notes: (i) The models were estimated using the adapted double bootstrap procedure with $B_1 = 200$ iterations for the first round of bootstrapping and $B_2 = 2000$ iterations for the second round of bootstrapping. The computations were done using Matlab codes programmed by the authors, involving standard Matlab library and some codes adopting from Matlab codes of Léopold Simar. Both coefficients and average partial effects (APE) are reported. Bootstrapped standard errors are reported in the parentheses. (ii) * p-value < 0.1, ** p-value < 0.05, ***p-value < 0.01.

2 CRS vs. VRS

In reality, each hospital may use its own proprietary technology, which may depend on a myriad of factors that are specific to that hospital, including its geographical location, patient demographics, and so on. Many of these factors are simply infeasible to account for in practice. Such technology is far more complex to describe than either CRS, VRS, or NIRS, or any other simplified models. Hence, our application of DEA is not about the estimation of the "true" technology, rather it is about estimating the 'agreed reference' as revealed by the unconditional frontier of the observed best practice. The question therefore is what should be the 'agreed reference'?

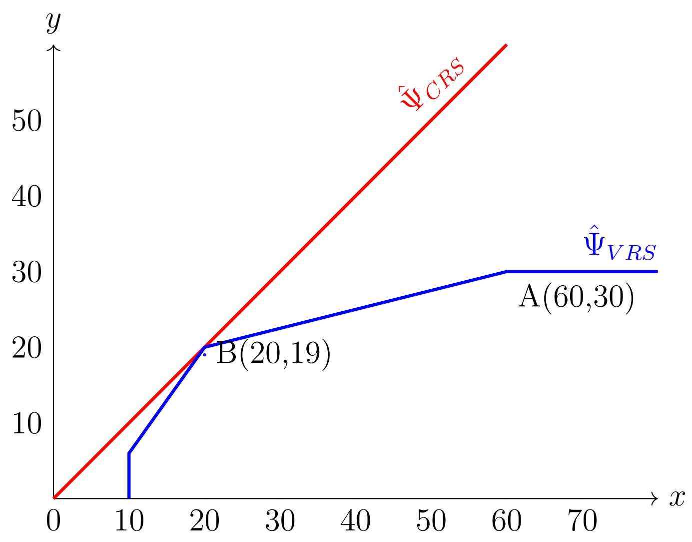
Consider a hypothetical example with the VRS reference. Suppose there are two hypothetical hospitals, A and B, using a single-input-single-output technology with their input-output combinations being $(60, 30)$ and $(20, 19)$, respectively (Figure 3).[3] With respect to the VRS reference technology, hospital A is viewed as fully efficient, i.e., it is on the estimated boundary of the VRS technology. Meanwhile, hospital B is only 95% efficient. Yet, from the social point of view, hospital B is far better than hospital A in providing healthcare services for society. Hospital A appears to waste too many resources, only producing around half of the healthcare services that would have been produced by three hospitals of the same size as hospital B (using the same total amount of inputs).

On the other hand, with respect to the CRS reference technology, the efficiency measures correspond to the socially optimal use of resources: hospital A is now only 50% efficient, meanwhile hospital B is still 95% efficient. The CRS reference technology thus can be viewed as the socially optimal reference technology, which reflects the socially optimal scale and the highest level of average productivity. However, as with any perspective, the CRS reference technology is not without caveats. Alternative views may allow hospitals to operate at sub-optimal scales because there might be factors beyond hospitals' control that prevent them from reaching the socially optimal scale, such as imperfect competition, budget constraints, as well as regulatory constraints on entry, mergers, exits, and so on.

In a nutshell, the emphasis here is that it is important to consider both CRS and VRS reference technologies, and to understand and explain the differences between them when deriving the policy implications.

[3]This hypothetical example is similar to and inspired by the example in Nguyen and Zelenyuk (2021), to which interested readers are referred for more detailed discussions.

Figure 3: CRS vs VRS assumption illustration



3 Labour Input Aggregation

We here discuss the aggregation method proposed in Daraio and Simar (2007), who argued that when inputs are highly correlated, it is reasonable and recommended to use a suitable aggregation of inputs for the efficiency analysis due to the 'curse of dimensionality'. The basic principle of their aggregation approach is to find an 'input factor' which is a linear combination of all inputs and covers the most information contained in all of the inputs. Because the input factor acts as a proxy for all inputs, it is expected to be highly positively correlated with all inputs. Moreover, Daraio and Simar (2007) also suggested that in the case that inputs are measured in different units, researchers should standardize them by dividing each input either by its mean or by its standard deviation. It is worth noting here that the standardization does not alter the efficiency analysis because the DEA estimator of the Farrell type technical efficiency measure is scale-invariant. The aggregation problem, then, can be stated mathematically as follows.

Assuming that L is an $N \times r$ matrix of (standardized) inputs, where N is the number of observations and r is the number of inputs to be aggregated (in our case for labour inputs, $N = 684$ and $r = 6$). The problem is to find an r-dimensional scalar vector, say $w \in \Re^r_+$, such that the input factor, say F_N, given by the following formula, 'best' represents the data matrix L (in terms of minimizing the sum of squares of residuals)

$$F_N = Lw = w_1 L_1 + ... + w_r L_r. \tag{3.2}$$

As discussed in Daraio and Simar (2007), adapting principal component analysis, it can be shown that the optimal solution for vector w is the first eigenvector of the matrix $L'L$, which corresponds to the largest eigenvalue λ_1. Moreover, it is important to note that eigenvalues are the second moment of the factor and the ratio of $\lambda_1/\sum_{j=1}^{r}\lambda_j$ represents the percentage of the second moment explained by the optimal input factor.

For Queensland hospitals' labour data, the correlation matrix shown in Table 3 suggests that all the labour categories are highly correlated. Thus we apply the Daraio and Simar's (2007) approach to aggregate all the labour inputs into one 'labour factor' (*flabour*). In our case, we do not standardize the data because all the labour inputs are measured in the same unit (FTE staff). The high value of $\lambda_1/\sum_{j=1}^{6}\lambda_j$ (see the first row of Table 4) indicates that *flabour* summarizes most of the information provided in all of the labour categories. Besides, the high correlations between *flabour* and each of the labour categories (Table 3) demonstrates that the aggregated variable is a good proxy for the original ones.

Table 3: Correlation matrix of labour inputs

	meo	*nur*	*dhp*	*opcs*	*acs*	*dos*	*flabour*
meo	1.00						
nur	1.00	1.00					
dhp	0.97	0.98	1.00				
opcs	0.95	0.95	0.95	1.00			
acs	0.98	0.99	0.98	0.95	1.00		
dos	0.93	0.94	0.92	0.87	0.94	1.00	
flabour	1.00	1.00	0.98	0.95	0.99	0.95	1.00

Table 4: Eigenvalues and eigenvectors of matrix $X'X$

Eigenvalues	% of second moment	Corresponding eigenvectors					
393794177	0.9898	0.29	0.84	0.24	0.05	0.29	0.26
2236416	0.0056	-0.17	-0.19	-0.16	-0.08	0.00	0.95
1139697	0.0029	-0.14	-0.36	0.74	0.06	0.54	0.04
402004	0.0010	-0.01	0.07	0.60	-0.01	-0.79	0.12
205538	0.0005	0.93	-0.35	0.02	-0.04	-0.01	0.10
68348	0.0002	0.02	-0.05	-0.06	0.99	-0.05	0.07

4 Analysis of Multicollinearity

Following Chowdhury and Zelenyuk (2016), we diagnose multicollinearity among independent variables using the variance inflation factor (VIF). There seems to be an issue of multicollinearity between time-varying covariates and their Mundlak adjustment terms, especially for casemix index and occupancy rate. Table 5 shows that the VIFs of these covariates and their Mundlak adjustment terms are around the rule of thumb of 10 (Kutner et al., 2004).

Table 5: Variance Inflation Factor

Variable names	VIF	Tolerance	R-square
Acitivity-based funding	3.11	0.32	0.68
Major city hospitals	2.06	0.49	0.51
Large hospitals	6.00	0.17	0.83
Teaching hospitals	2.67	0.37	0.63
The ratio of outpatient to inpatient (log)	4.86	0.21	0.79
The proportion of unit producing personnel (log)	5.61	0.18	0.82
Casemix index (log)	10.16	0.10	0.90
Occupancy rate (log)	7.22	0.14	0.86
FY 2006/07	1.84	0.54	0.46
FY 2007/08	1.86	0.54	0.46
FY 2008/09	1.87	0.53	0.47
FY 2009/10	1.98	0.51	0.49
FY 2010/11	2.00	0.50	0.50
FY 2011/12	2.23	0.45	0.55
FY 2012/13	2.21	0.45	0.55
FY 2013/14	2.26	0.44	0.56
FY 2014/15	2.31	0.43	0.57
FY 2015/16	2.26	0.44	0.56
FY 2016/17	2.34	0.43	0.57
2011/12 adoption cohort	5.66	0.18	0.82
2013/14 adoption cohort	1.25	0.80	0.20
2014/15 adoption cohort	1.09	0.92	0.08
The ratio of outpatient to inpatient (log, time average)	6.61	0.15	0.85
The proportion of unit producing personnel (log, time average)	6.55	0.15	0.85
Casemix index (log, time average)	10.80	0.09	0.91
Occupancy rate (log, time average)	9.31	0.11	0.89

5 Full estimation results

Table 6: Full regression results - Main models - WLOS and technical efficiency

	WLOS (log)	Technical efficiency (CRS-DEA, log)		Technical efficiency (VRS-DEA, log)	
	Coeff.	Coeff.	APE	Coeff.	APE
Acitivity-based funding	-0.002	-0.056^{***}	-0.054^{***}	-0.064^{***}	-0.048^{**}
	(0.038)	(0.013)	(0.013)	(0.022)	(0.017)
Major city hospitals	-0.090	-0.131^{***}	-0.125^{***}	-0.117^{**}	-0.091^{**}
	(0.075)	(0.039)	(0.036)	(0.051)	(0.038)
Large hospitals	0.163	-0.018	-0.018	-0.130^{**}	-0.106^{**}
	(0.100)	(0.030)	(0.029)	(0.056)	(0.046)
Teaching hospitals	-0.033	0.005	0.005	-0.046	-0.037
	(0.049)	(0.010)	(0.010)	(0.030)	(0.024)
The ratio of outpatient to	-0.140	-0.467^{***}	-0.450^{***}	-0.279^{***}	-0.223^{***}
inpatient (log)	(0.161)	(0.066)	(0.063)	(0.102)	(0.082)
The proportion of unit	0.111^{**}	-0.019	-0.018	-0.199^{***}	-0.159^{***}
producing personnel (log)	(0.054)	(0.017)	(0.016)	(0.029)	(0.023)
Casemix index (log)	0.328^{**}	0.150^{**}	0.144^{**}	0.130	0.104
	(0.146)	(0.051)	(0.049)	(0.085)	(0.068)
Occupancy rate (log)	0.234^{***}	-0.163^{***}	-0.157^{***}	-0.086^{**}	-0.069^{**}
	(0.051)	(0.023)	(0.022)	(0.033)	(0.026)
FY 2006/07	0.006	0.033^{*}	0.032^{*}	0.016	0.013
	(0.017)	(0.015)	(0.015)	(0.028)	(0.023)
FY 2007/08	-0.025	0.061^{***}	0.059^{***}	0.036	0.030
	(0.019)	(0.016)	(0.015)	(0.027)	(0.023)
FY 2008/09	0.027	0.107^{***}	0.104^{***}	0.079^{***}	0.065^{***}
	(0.021)	(0.017)	(0.016)	(0.028)	(0.024)
FY 2009/10	0.082^{***}	0.115^{***}	0.112^{***}	0.077^{***}	0.063^{**}
	(0.022)	(0.017)	(0.016)	(0.028)	(0.024)
FY 2010/11	-0.027	0.211^{***}	0.207^{***}	0.172^{***}	0.147^{***}
	(0.038)	(0.019)	(0.019)	(0.028)	(0.025)
FY 2011/12	-0.061	0.224^{***}	0.219^{***}	0.196^{***}	0.169^{***}
	(0.039)	(0.018)	(0.018)	(0.029)	(0.026)
FY 2012/13	-0.142^{***}	0.191^{***}	0.187^{***}	0.171^{***}	0.146^{***}
	(0.042)	(0.019)	(0.019)	(0.029)	(0.026)
FY 2013/14	-0.114^{**}	0.190^{***}	0.186^{***}	0.187^{***}	0.161^{***}
	(0.049)	(0.019)	(0.018)	(0.028)	(0.026)
FY 2014/15	-0.127^{***}	0.151^{***}	0.148^{***}	0.155^{***}	0.132^{***}
	(0.043)	(0.019)	(0.018)	(0.030)	(0.027)

Continued on next page

FY 2015/16	-0.139^{***}	0.187^{***}	0.183^{***}	0.200^{***}	0.172^{***}
	(0.042)	(0.019)	(0.018)	(0.029)	(0.027)
FY 2016/17	-0.089^{***}	0.201^{***}	0.196^{***}	0.194^{***}	0.167^{***}
	(0.033)	(0.019)	(0.019)	(0.029)	(0.027)
2011/12 adoption cohort	−0.041	0.132^{**}	0.125^{**}	0.012	0.010
	(0.088)	(0.045)	(0.042)	(0.044)	(0.035)
2013/14 adoption cohort	-0.217^{*}	−0.010	−0.010	−0.023	−0.018
	(0.123)	(0.026)	(0.024)	(0.072)	(0.056)
2014/15 adoption cohort	−0.098	0.004	0.004	−0.123	-0.091^{*}
	(0.116)	(0.009)	(0.009)	(0.074)	(0.050)
The ratio of outpatient to inpatient (log, time average)	−0.108	−0.022	−0.021	−0.159	−0.127
	(0.375)	(0.048)	(0.046)	(0.234)	(0.187)
The proportion of unit producing personnel (log, time average)	-0.231^{**}	-0.117^{**}	-0.113^{**}	−0.013	−0.010
	(0.106)	(0.045)	(0.043)	(0.044)	(0.035)
Casemix index (log, time average)	0.752^{***}	0.176^{*}	0.169^{*}	0.083	0.067
	(0.212)	(0.093)	(0.089)	(0.134)	(0.107)
Occupancy rate (log, time average)	0.057	-0.267^{***}	-0.257^{***}	-0.285^{***}	-0.228^{***}
	(0.108)	(0.056)	(0.053)	(0.069)	(0.055)
Constant	2.532^{***}	0.331^{**}	n.a	0.498^{***}	n.a
	(0.213)	(0.107)	n.a	(0.124)	n.a
$\hat{\sigma}_a$	$0.281^{n.a}$	0.138^{***}	n.a	0.096^{***}	n.a
	n.a	(0.013)	n.a	(0.013)	n.a
$\hat{\sigma}_\varepsilon$	$0.172^{n.a}$	0.124^{***}	n.a	0.130^{***}	n.a
	n.a	(0.003)	n.a	(0.005)	n.a
Number of observations	684		684		684
LogLikelihood	n.a		394.177		513.912
AIC	n.a		−734.353		−973.824
BIC	n.a		−612.098		−851.569
R-square	0.296		n.a		n.a
Adjusted R-square	0.268		n.a		n.a

Notes: (i) The models for the technical efficiency were estimated using the adapted double bootstrap procedure with $B_1 = 200$ iterations for the first round of bootstrapping and $B_2 = 2000$ iterations for the second round of bootstrapping. The computations were done using Matlab codes programmed by the authors, involving standard Matlab library and some codes adopting from Matlab codes of Léopold Simar. Both coefficients and average partial effects (APE) are reported. Bootstrapped standard errors are reported in the parentheses. (ii) The model for the weighted average length of stay (WLOS) was estimated using Feasible Generalized Least Square estimator, utilizing the `plm` package in R (Croissant and Millo, 2008). Clustered standard errors (clustering at individual level) are reported in the parentheses. (iii) * p-value < 0.1, ** p-value < 0.05, ***p-value < 0.01.

Table 7: Full regression results - Event-study setup - WLOS and technical efficiency

	WLOS (log)	Technical efficiency (CRS-DEA, log)		Technical efficiency (VRS-DEA, log)	
	Coeff.	Coeff.	APE	Coeff.	APE
$\geq$ 4 years before	-0.053	-0.030	-0.029	0.012	0.010
	(0.047)	(0.027)	(0.026)	(0.029)	(0.023)
3 years before	-0.060	0.019	0.019	0.002	0.001
	(0.043)	(0.030)	(0.030)	(0.008)	(0.006)
2 years before	-0.146^{***}	-0.052	-0.050	-0.021	-0.017
	(0.043)	(0.031)	(0.030)	(0.037)	(0.029)
Starting adoption	-0.122^{**}	-0.029	-0.028	-0.006	-0.005
	(0.058)	(0.029)	(0.028)	(0.023)	(0.018)
1 year after	-0.116^{**}	-0.070^{**}	-0.066^{**}	-0.034	-0.027
	(0.057)	(0.031)	(0.029)	(0.035)	(0.027)
2 years after	-0.067	-0.080^{**}	-0.077^{**}	-0.024	-0.019
	(0.064)	(0.032)	(0.030)	(0.036)	(0.028)
3 years after	-0.090	-0.084^{**}	-0.080^{**}	-0.053	-0.041
	(0.074)	(0.032)	(0.030)	(0.038)	(0.029)
4 years after	-0.032	-0.083^{**}	-0.079^{**}	-0.069	-0.053^{*}
	(0.070)	(0.034)	(0.031)	(0.040)	(0.030)
$\geq$ 5 years later	-0.054	-0.088^{**}	-0.084^{**}	-0.091^{**}	-0.070^{***}
	(0.056)	(0.030)	(0.028)	(0.035)	(0.025)
Major city hospitals	-0.089	-0.130^{**}	-0.124^{**}	-0.196^{***}	-0.148^{***}
	(0.076)	(0.049)	(0.046)	(0.049)	(0.034)
Large hospitals	0.167^{*}	-0.015	-0.014	-0.120^{**}	-0.097^{**}
	(0.101)	(0.036)	(0.034)	(0.057)	(0.047)
Teaching hospitals	-0.040	0.012	0.012	-0.046	-0.037
	(0.049)	(0.020)	(0.020)	(0.029)	(0.023)
The ratio of outpatient to	-0.155	-0.460^{***}	-0.445^{***}	-0.348^{***}	-0.278^{***}
inpatient (log)	(0.162)	(0.080)	(0.078)	(0.097)	(0.078)
The proportion of unit	0.113^{**}	-0.021	-0.020	-0.212^{***}	-0.169^{***}
producing personnel (log)	(0.055)	(0.022)	(0.021)	(0.029)	(0.023)
Casemix index (log)	0.331^{**}	0.144^{**}	0.139^{**}	0.179^{**}	0.143^{**}
	(0.147)	(0.061)	(0.059)	(0.081)	(0.065)
Occupancy rate (log)	0.220^{***}	-0.163^{***}	-0.158^{***}	-0.117^{***}	-0.094^{***}
	(0.052)	(0.027)	(0.026)	(0.032)	(0.026)

Continued on next page

FY 2006/07	0.006	0.036*	0.035*	0.017	0.013
	(0.017)	(0.020)	(0.020)	(0.026)	(0.021)
FY 2007/08	−0.021	0.039	0.038	0.056*	0.046*
	(0.027)	(0.023)	(0.022)	(0.029)	(0.024)
FY 2008/09	0.073**	0.119***	0.116***	0.118***	0.098***
	(0.030)	(0.024)	(0.024)	(0.029)	(0.025)
FY 2009/10	0.058*	0.098***	0.096***	0.106***	0.088***
	(0.032)	(0.023)	(0.023)	(0.029)	(0.025)
FY 2010/11	0.011	0.211***	0.207***	0.201***	0.173***
	(0.050)	(0.026)	(0.025)	(0.030)	(0.028)
FY 2011/12	−0.029	0.216***	0.213***	0.210***	0.181***
	(0.048)	(0.026)	(0.025)	(0.032)	(0.029)
FY 2012/13	−0.135**	0.189***	0.185***	0.180***	0.154***
	(0.053)	(0.027)	(0.026)	(0.031)	(0.028)
FY 2013/14	−0.092	0.190***	0.186***	0.204***	0.175***
	(0.058)	(0.026)	(0.026)	(0.032)	(0.030)
FY 2014/15	−0.132**	0.150***	0.147***	0.177***	0.151***
	(0.052)	(0.027)	(0.026)	(0.031)	(0.029)
FY 2015/16	−0.136***	0.189***	0.185***	0.235***	0.204***
	(0.046)	(0.025)	(0.025)	(0.030)	(0.028)
FY 2016/17	−0.085**	0.201***	0.197***	0.229***	0.198***
	(0.038)	(0.026)	(0.025)	(0.030)	(0.028)
2011/12 adoption cohort	0.028	0.130**	0.121**	−0.022	−0.016
	(0.094)	(0.051)	(0.045)	(0.052)	(0.038)
2013/14 adoption cohort	−0.147	0.001	0.001	−0.037	−0.029
	(0.133)	(0.004)	(0.004)	(0.071)	(0.054)
2014/15 adoption cohort	−0.026	−0.008	−0.008	−0.118	−0.088*
	(0.127)	(0.025)	(0.024)	(0.070)	(0.048)
The ratio of outpatient to inpatient (log, time average)	−0.089 (0.379)	−0.021 (0.059)	−0.020 (0.057)	−0.086 (0.208)	−0.069 (0.167)
The proportion of unit producing personnel (log, time average)	−0.232** (0.107)	−0.179*** (0.055)	−0.174*** (0.053)	−0.017 (0.047)	−0.014 (0.038)
Casemix index (log, time average)	0.754*** (0.214)	0.173 (0.113)	0.167 (0.109)	0.217* (0.127)	0.173* (0.102)
Occupancy rate (log, time average)	0.070 (0.109)	−0.315*** (0.074)	−0.305*** (0.071)	−0.245*** (0.066)	−0.195*** (0.053)
Constant	2.521***	0.497***	n.a	0.573***	n.a
	(0.216)	(0.125)	n.a	(0.121)	n.a

Continued on next page

$\hat{\sigma}_a$	$0.305^{n.a}$	0.129^{***}	n.a	0.094^{***}	n.a
	n.a	(0.015)	n.a	(0.013)	n.a
$\hat{\sigma}_\varepsilon$	$0.171^{n.a}$	0.124^{***}	n.a	0.130^{***}	n.a
	n.a	(0.004)	n.a	(0.004)	n.a
Number of observations	684		684		684
LogLikelihood	n.a		397.272		519.352
AIC	n.a		−724.545		−968.704
BIC	n.a		−566.066		−810.226
R-square	0.300		n.a		n.a
Adjusted R-square	0.264		n.a		n.a

Notes: (i) (i) The models for the technical efficiency were estimated using the adapted double bootstrap procedure with $B_1 = 200$ iterations for the first round of bootstrapping and $B_2 = 2000$ iterations for the second round of bootstrapping. The computations were done using Matlab codes programmed by the authors, involving standard Matlab library and some codes adopting from Matlab codes of Léopold Simar. Both coefficients and average partial effects (APE) are reported. Bootstrapped standard errors are reported in the parentheses. (ii) The model for the weighted average length of stay (WLOS) was estimated using Feasible Generalized Least Square estimator, utilizing the `plm` package in R (Croissant and Millo, 2008). Clustered standard errors (clustering at individual level) are reported in the parentheses. (iii) * p-value < 0.1, ** p-value < 0.05, ***p-value < 0.01.

Table 8: Full regression results - Robustness check - DEA model

	CRS-DEA with two labour inputs		VRS-DEA with two labour inputs	
	Coeff.	APE	Coeff.	APE
Acitivity-based funding	−0.062***	−0.060***	−0.036*	−0.026*
	(0.018)	(0.017)	(0.018)	(0.014)
Major city hospitals	−0.127**	−0.119**	−0.139***	−0.104***
	(0.053)	(0.049)	(0.042)	(0.030)
Large hospitals	−0.005	−0.005	−0.140***	−0.112***
	(0.014)	(0.013)	(0.045)	(0.036)
Teaching hospitals	0.030	0.029	−0.010	−0.008
	(0.025)	(0.024)	(0.024)	(0.019)
The ratio of outpatient to inpatient (log)	−0.560***	−0.536***	−0.471***	−0.369***
	(0.090)	(0.086)	(0.083)	(0.066)
The proportion of unit producing personnel (log)	−0.014	−0.014	−0.178***	−0.139***
	(0.024)	(0.023)	(0.024)	(0.018)
Casemix index (log)	0.111	0.106	0.059	0.047
	(0.068)	(0.065)	(0.072)	(0.057)
Occupancy rate (log)	−0.117***	−0.112***	−0.073**	−0.057**
	(0.029)	(0.028)	(0.028)	(0.022)
FY 2006/07	0.023	0.023	0.010	0.008
	(0.022)	(0.021)	(0.023)	(0.019)
FY 2007/08	0.054**	0.052**	0.038	0.030
	(0.021)	(0.021)	(0.023)	(0.019)
FY 2008/09	0.110***	0.107***	0.104***	0.085***
	(0.023)	(0.022)	(0.024)	(0.021)
FY 2009/10	0.124***	0.120***	0.094***	0.077***
	(0.022)	(0.022)	(0.024)	(0.020)
FY 2010/11	0.216***	0.212***	0.180***	0.152***
	(0.024)	(0.023)	(0.025)	(0.022)
FY 2011/12	0.231***	0.226***	0.188***	0.160***
	(0.025)	(0.024)	(0.025)	(0.023)
FY 2012/13	0.205***	0.200***	0.174***	0.147***
	(0.025)	(0.024)	(0.026)	(0.023)
FY 2013/14	0.193***	0.188***	0.170***	0.143***
	(0.024)	(0.024)	(0.025)	(0.023)
FY 2014/15	0.165***	0.161***	0.152***	0.128***
	(0.025)	(0.025)	(0.026)	(0.023)
FY 2015/16	0.186***	0.181***	0.187***	0.158***
	(0.025)	(0.024)	(0.026)	(0.024)
FY 2016/17	0.209***	0.204***	0.193***	0.164***
	(0.025)	(0.025)	(0.026)	(0.023)

Continued on next page

2011/12 adoption cohort	0.127^{**}	0.120^{**}	0.007	0.006
	(0.052)	(0.048)	(0.030)	(0.024)
2013/14 adoption cohort	−0.064	−0.060	−0.036	−0.027
	(0.081)	(0.075)	(0.065)	(0.048)
2014/15 adoption cohort	−0.006	−0.006	-0.117^{*}	-0.085^{**}
	(0.017)	(0.017)	(0.064)	(0.043)
The ratio of outpatient to inpatient (log, time average)	0.460^{*}	0.440^{*}	−0.002	−0.002
	(0.250)	(0.239)	(0.009)	(0.007)
The proportion of unit producing personnel (log, time average)	−0.049	−0.046	−0.008	−0.006
	(0.062)	(0.059)	(0.029)	(0.023)
Casemix index (log, time average)	0.301^{**}	0.288^{**}	0.174	0.136
	(0.132)	(0.126)	(0.110)	(0.087)
Occupancy rate (log, time average)	-0.356^{***}	-0.340^{***}	-0.271^{***}	-0.212^{***}
	(0.078)	(0.074)	(0.058)	(0.045)
Constant	0.258^{*}	n.a	0.399^{***}	n.a
	(0.138)	n.a	(0.087)	n.a
$\hat{\sigma}_a$	0.124^{***}	n.a	0.102^{***}	n.a
	(0.017)	n.a	(0.013)	n.a
$\hat{\sigma}_\varepsilon$	0.125^{***}	n.a	0.130^{***}	n.a
	(0.004)	n.a	(0.004)	n.a
Number of observations		684		684
LogLikelihood		397.270		526.133
AIC		−740.540		−998.265
BIC		−618.286		−876.010

Notes: (i) The models were estimated using the adapted double bootstrap procedure with $B_1 = 200$ iterations for the first round of bootstrapping and $B_2 = 2000$ iterations for the second round of bootstrapping. The computations were done using Matlab codes programmed by the authors, involving standard Matlab library and some codes adopting from Matlab codes of Léopold Simar. Both coefficients and average partial effects (APE) are reported. Bootstrapped standard errors are reported in the parentheses. (ii) * p-value < 0.1, ** p-value < 0.05, ***p-value < 0.01.

Table 9: Event-study - Robustness check - DEA model

	CRS-DEA with two labour inputs		VRS-DEA with two labour inputs	
	Coeff.	APE	Coeff.	APE
≥ 4 years before	−0.011	−0.010	0.005	0.004
	(0.021)	(0.020)	(0.016)	(0.013)
3 years before	0.042	0.040	0.014	0.011
	(0.030)	(0.029)	(0.030)	(0.024)
2 years before	−0.033	−0.032	−0.041	−0.031
	(0.029)	(0.027)	(0.033)	(0.025)
Starting adoption	−0.014	−0.013	0.002	0.002
	(0.025)	(0.024)	(0.008)	(0.006)
1 year after	−0.057*	−0.054**	−0.019	−0.015
	(0.029)	(0.027)	(0.030)	(0.023)
2 years after	−0.069**	−0.065**	−0.030	−0.023
	(0.030)	(0.027)	(0.032)	(0.024)
3 years after	−0.047	−0.044	−0.021	−0.017
	(0.030)	(0.028)	(0.033)	(0.025)
4 years after	−0.068*	−0.064*	−0.038	−0.029
	(0.032)	(0.029)	(0.034)	(0.026)
≥ 5 years later	−0.094***	−0.089***	−0.074**	−0.055**
	(0.028)	(0.026)	(0.030)	(0.022)
Major city hospitals	−0.129**	−0.121**	−0.190***	−0.139***
	(0.053)	(0.048)	(0.043)	(0.029)
Large hospitals	−0.001	−0.001	−0.122**	−0.097**
	(0.002)	(0.002)	(0.050)	(0.040)
Teaching hospitals	0.035	0.033	−0.046*	−0.036*
	(0.024)	(0.023)	(0.025)	(0.020)
The ratio of outpatient to inpatient	−0.568***	−0.544***	−0.384***	−0.300***
	(0.085)	(0.082)	(0.079)	(0.062)
The proportion of unit producing personnel	−0.012	−0.012	−0.184***	−0.144***
	(0.022)	(0.021)	(0.025)	(0.020)
Casemix index	0.105	0.100	0.123	0.096
	(0.062)	(0.060)	(0.069)	(0.054)
Occupancy rate	−0.107***	−0.102***	−0.090***	−0.070***
	(0.028)	(0.026)	(0.028)	(0.022)

Continued on next page

FY 2006/07	0.021	0.020	−0.002	−0.002
	(0.019)	(0.018)	(0.010)	(0.008)
FY 2007/08	0.028	0.026	0.036	0.028
	(0.024)	(0.023)	(0.025)	(0.020)
FY 2008/09	0.119***	0.116***	0.117***	0.096***
	(0.025)	(0.024)	(0.024)	(0.021)
FY 2009/10	0.116***	0.112***	0.099***	0.080***
	(0.024)	(0.024)	(0.023)	(0.019)
FY 2010/11	0.215***	0.210***	0.182***	0.153***
	(0.027)	(0.027)	(0.025)	(0.022)
FY 2011/12	0.223***	0.218***	0.187***	0.158***
	(0.028)	(0.028)	(0.027)	(0.024)
FY 2012/13	0.202***	0.198***	0.173***	0.145***
	(0.027)	(0.026)	(0.025)	(0.023)
FY 2013/14	0.179***	0.175***	0.162***	0.135***
	(0.028)	(0.027)	(0.027)	(0.024)
FY 2014/15	0.161***	0.156***	0.157***	0.131***
	(0.028)	(0.027)	(0.026)	(0.024)
FY 2015/16	0.193***	0.188***	0.198***	0.167***
	(0.026)	(0.026)	(0.025)	(0.023)
FY 2016/17	0.217***	0.211***	0.205***	0.174***
	(0.026)	(0.026)	(0.026)	(0.024)
2011/12 adoption cohort	0.136**	0.125***	−0.015	−0.011
	(0.050)	(0.044)	(0.043)	(0.030)
2013/14 adoption cohort	−0.066	−0.063	−0.038	−0.029
	(0.078)	(0.072)	(0.061)	(0.046)
2014/15 adoption cohort	0.001	0.001	−0.062	−0.047
	(0.004)	(0.004)	(0.060)	(0.044)
The ratio of outpatient to inpatient	0.463*	0.443*	0.001	0.001
(log, time average)	(0.244)	(0.234)	(0.005)	(0.004)
The proportion of unit producing	−0.049	−0.047	−0.018	−0.014
personnel (log, time average)	(0.058)	(0.056)	(0.043)	(0.034)
Casemix index (log, time average)	0.297**	0.284**	0.292**	0.228**
	(0.126)	(0.120)	(0.114)	(0.090)
Occupancy rate (log, time average)	−0.368***	−0.352***	−0.267***	−0.209***
	(0.075)	(0.070)	(0.058)	(0.045)
Constant	0.249*	0.000	0.511***	n.a
	(0.128)	(0.000)	(0.100)	n.a

Continued on next page

$\hat{\sigma}_a$	0.125^{***}	n.a	0.097^{***}	n.a
	(0.016)	n.a	(0.013)	n.a
$\hat{\sigma}_\varepsilon$	0.125^{***}	n.a	0.130^{***}	n.a
	(0.004)	n.a	(0.004)	n.a
Number of observations		684		684
LogLikelihood		398.858		534.033
AIC		−727.715		−998.066
BIC		−569.237		−839.588

Notes: (i) The models were estimated using the adapted double bootstrap procedure with $B_1 = 200$ iterations for the first round of bootstrapping and $B_2 = 2000$ iterations for the second round of bootstrapping. The computations were done using Matlab codes programmed by the authors, involving standard Matlab library and some codes adopting from Matlab codes of Léopold Simar. Both coefficients and average partial effects (APE) are reported. Bootstrapped standard errors are reported in the parentheses. (ii) * p-value < 0.1, ** p-value < 0.05, ***p-value < 0.01.

References

Chowdhury, H. and Zelenyuk, V. (2016), 'Performance of hospital services in Ontario: DEA with truncated regression approach', *Omega* **63**, 111–122.

Croissant, Y. and Millo, G. (2008), 'Panel data econometrics in r: The plm package', *Journal of Statistical Software* **27**(2).

Daraio, C. and Simar, L. (2007), Economies of scale, scope and experience in the italian motor-vehicle sector, *in* C. Daraio and L. Simar, eds, 'Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications', Springer Science & Business Media, pp. 135–165.

Jacobson, L. S., LaLonde, R. J. and Sullivan, D. G. (1993), 'Earnings losses of displaced workers', *The American Economic Review* **83**(4), 685–709.

Kutner, M. H., Nachtsheim, C. J. and Neter, J. (2004), *Applied Linear Regression Models*, McGraw-Hill/Irwin.

Nguyen, B. H. and Zelenyuk, V. (2021), 'Aggregate efficiency of industry and its groups: The case of Queensland public hospitals', *Empirical Economics* **60**(6), 2795–2836.