Random vs. Explained Inefficiency in SFA: The Case of Queensland Hospitals

Zhichao Wang [1] Valentin Zelenyuk [2]

[1]School of Economics, The University of Queensland

[2]School of Economics and Centre for Efficiency and Productivity Analysis (CEPA), The University of Queensland

Australian Conference of Economists, July 2022

Table of contents

1.1 Foundation of a ‘healthy society’

- Healthcare is an indispensable foundation of a ‘healthy society’, as it is a solid pillar of both a healthy population and a prosperous economy.
- According to OECD (2021), the proportion of total expenditure on healthcare accounted for the GDP in 2019 is:
  - About 16.8% in the US;
  - About 9.3% in Australia;
  - About 8.3% on average.
- Hospital, as the main provider of health services, is naturally a popular subject of study in this context.
- The empirical focus of this paper is the efficiency of Queensland public hospitals during FY 2012/13 to FY 2015/16 and try to identify the factors that explain its variation.

1.2 Related reviews

- Selected studies focused on the efficiency of hospitals in Australia:
  - The Productivity Commission (2009; 2010):
    - ⋆ Method(s): Stochastic Frontier Analysis (SFA);
    - ⋆ Period: FY 2004/05 to FY 2006/07;
    - ⋆ Conclusions: a nationwide 10 to 20% technical inefficiency.
  - Gabbitas and Jeffs (2009):
    - ⋆ Method(s): SFA models for both random effects and fixed effects;
    - ⋆ Period: FY 1995/96 to FY2005/06;
    - ⋆ Conclusions: an achievable productivity improvement of 10% on aggregate nationwide.

1.2 Related reviews

- Regarding the hospitals in Queensland, Australia:
  - O'Donnell and Nguyen (2013):
    - Method(s): SFA;
    - Period: FY 1995/96 to FY 2003/04;
    - Conclusions: The average level of inefficiency comparing to a benchmark large general hospital in the beginning of the study period (FY1995/96), was estimated at 32.9%.
  - Nguyen and Zelenyuk (2021):
    - Method(s): Data Envelopment Analysis (DEA), SFA and partial quantile frontier model;
    - Period: FY 2016/17;
    - Conclusions: The more hospitals in small size and remote area could partially explain the low efficiency level of some districts.
  - Gabbitas and Jeffs (2009):
    - Method(s): SFA models for both random effects and fixed effects;
    - Period: FY 1995/96 to FY2005/06;
    - Conclusions: The inefficiency level of Queensland public hospital system was estimated at about 20 percent, which is one of the highest among the State and Territory in Australia.

Table of contents

2.1 Brief background of efficiency analysis

- Efficiency Analysis
  - Measuring the efficiency of a group of organizations has attracted a lot of interest of researchers and policy makers in various fields, e.g., banking, healthcare, transportation, energy, etc.
  - The roots of the many approaches and applications for such measurements can be traced back to the seminal work of Farrell (1957).
- DEA
  - The measurement of productive efficiency proposed in Farrell (1957) was later elaborated as DEA by Charnes et al. (1978).
  - DEA has been further developed in many ways in past decades and applied in various research fields and industries.
- SFA
  - SFA was proposed by Aigner et al. (1977) and Meeusen and van den Broeck (1977).
  - A wide range of SFA models have been developed for specific research questions, e.g., for panel data and determinants of inefficiency.

2.2 Basic SFM

- The stochastic frontier model (SFM) can be formulated as

$$Y_i = \psi(X_i|\beta)\exp(v_i)\exp(-u_i), \quad i = 1, \ldots, n, \tag{1}$$

  - $X \in \Re^p_+$ is the vector of p inputs;
  - Output is measured by $Y \in \Re_+$;
  - Technology characterized by the production function $\psi(X|\beta)$;
  - β is the unknown vector of parameters;
  - $\exp(v_i)$ stands for statistical noise;
  - $\exp(-u_i)$ represents efficiency of unit i.

2.2 Aigner et al. (1977)

- A linear form of Equation (1) can be derived as

$$y_i = f(x_i|\beta) + \varepsilon_i, i = 1, \ldots, n, \tag{2}$$

  - Let $y_i = \ln(Y_i)$ and $x_i = \ln(X_i)$;
  - ε_i is the the error term composed as $\varepsilon_i = v_i - u_i$.
- Various SFMs put different assumptions about v_i, u_i, and $f(x_i|\beta)$.
- We start with the classic and still popular SFM proposed by Aigner et al. (1977) (hereafter ALS77).
  - $f(x_i|\beta)$ is assumed as functional form (e.g., Cobb-Douglas, Translog);
  - Distribution of v_i and u_i is assumed as

$$v_i \sim iid\mathcal{N}(0, \sigma_v^2), \tag{3}$$

$$u_i \sim iid\mathcal{N}^+(0, \sigma_u^2). \tag{4}$$

2.2 SFA Models with Determinant of Inefficiency

- An attractive advancement of the basic SFM is to consider the exogenous variables (or the 'environmental factors') with the (in)efficiency.
- We use the pioneer and still popular model introduced in Kumbhakar et al. (1991) (hereafter KGM91) for this purpose, which can be formulated as

$$
\begin{aligned}
y_i &= x_i'\beta + v_i - u_i, i = 1, \ldots, n, \\
v_i &\sim iid\mathcal{N}(0, \sigma_v^2), \\
u_i &\sim iid\mathcal{N}^+(\tau_i, \sigma_u^2), \\
\tau_i &= z_i'\gamma,
\end{aligned}
\tag{5}
$$

  - $z_i \in \Re^m$ represents a vector of m exogenous variables
  - z_i 'determines' the inefficiency via the mean of u_i, via γ, the vector of parameters to be estimated.

2.2 Estimation of Expected and Individual (In)Efficiency

- The models with specific assumptions can then be estimated with the maximum likelihood estimation (MLE);
  - Using the estimated parameters, we can further estimate the expected level of efficiency ($E[\exp(-u)] \approx 1 - E(u)$) and expected inefficiency ($E(u) = \sqrt{\frac{2}{\pi}}\sigma_u$).
- The individual level of inefficiency of unit i, $\hat{u}_i$,can be estimated following Jondrow et al. (1982) (hereafter JLMS), by estimating $E(\hat{u}_i|\varepsilon_i)$.
  - To derive a more interpretable inefficiency level, we may use $\tilde{u}_i = 1 - \exp(-\hat{u}_i)$.

Table of contents

3.1 Statistics of hospitals

Table 1: Descriptive statistics of variables

Variables	Description	Mean	Std Dev	Min	Q1	Q3	Max
Input							
Agglabours	Aggregated labor input	0.76	1.57	0.01	0.07	0.42	8.71
MEO	Salaried medical officers*	49.18	109.62	0	2.18	14.75	630.91
TNUR	Nurses*	147.94	297.45	2.28	15.19	82.43	1641.75
DHP	Diagnostic and health professionals*	30.09	74.40	0.00	0.00	15.18	468.00
ACS	Administrative and clerical staff*	46.54	101.53	0.18	2.11	22.66	627.80
OPCS	Other personal care staff*	7.39	17.55	0.00	0.01	4.00	125.42
DOS	Domestic and other staff*	49.31	85.62	0.00	8.62	42.45	509.30
SUPP	Consumable input**	7.83	19.20	0.03	0.16	2.53	164.00
DSUP	Drug supplies expenditure**	2.16	5.48	0.00	0.04	0.81	36.60
MSSUP	Medical and surgical supplies**	5.95	15.00	0.02	0.12	1.74	148.00
BEDS	Number of beds	74.92	133.78	3.00	11.00	48.50	680.00
Output							
Aggout	Aggregated output	0.54	1.03	0.01	0.04	0.32	5.05
Wepisodes	Case-mix weighted episodes	8012.15	16837.94	31.08	341.26	4001.78	82815.42
OUT	Number of outpatient visits	77741.66	137870.60	1248.00	6361.50	48525.50	677604.00
Environmental							
Continuous							
Occup	Occupancy rate	0.56	0.29	0.05	0.33	0.76	2.31
Inout	Inpatient-outpatient ratio	0.09	0.04	0.01	0.05	0.11	0.29
AveLOS	Average length of stay	3.54	4.04	1.28	2.21	3.62	54.24
Tover	Bed turn-over rate	71.05	41.05	7.79	38.16	99.84	189.59
Binary		**Number**	**Percentage**				
Remote	Located in remote areas	108	28.42%				
Non-remote	Located in non-remote areas	272	71.58%				
Small	Small hospitals	300	78.95%				
Big	Large hospitals	80	21.05%				
TEACH	Teaching hospitals	70	18.42%				
Non-teaching	No teaching function	310	81.58%				

* Full-time equivalent staffs;
** AU$1,000,000 in constant price of FY2012/2013.

3.1 Statistics of hospitals

- Twice more hospitals in non-remote areas than in remote areas;
- ‘Small hospitals’ are the majority, few ‘large hospitals’ locate in remote areas;
- ‘Teaching hospitals’ are usually of large size, so that they usually locate in non-remote areas.

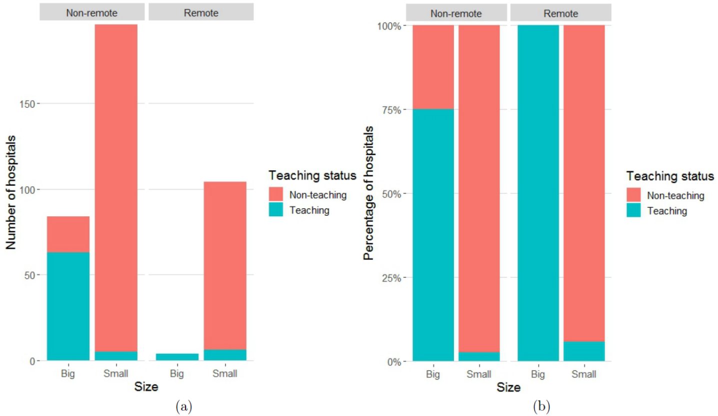

Figure: Characteristics of hospitals

3.2 PCA and outlier

- High level of correlation among each pair of input and output variables;
- Suggests a reliable aggregation is possible.

	MEO	TNUR	OPCS	DHP	ACS	DOS	DSUP	MSSUP	OUT	BEDS	Wepisodes
MEO	1	1	0.95	0.97	0.98	0.91	0.92	0.92	0.97	0.99	0.99
TNUR	1	1	0.95	0.98	0.98	0.92	0.94	0.92	0.97	0.99	0.99
OPCS	0.95	0.95	1	0.96	0.94	0.82	0.94	0.94	0.92	0.96	0.96
DHP	0.97	0.98	0.96	1	0.97	0.89	0.94	0.91	0.92	0.97	0.97
ACS	0.98	0.98	0.94	0.97	1	0.92	0.94	0.89	0.95	0.97	0.97
DOS	0.91	0.92	0.82	0.89	0.92	1	0.79	0.73	0.89	0.9	0.89
DSUP	0.92	0.94	0.94	0.94	0.94	0.79	1	0.92	0.89	0.92	0.93
MSSUP	0.92	0.92	0.94	0.91	0.89	0.73	0.92	1	0.9	0.93	0.94
OUT	0.97	0.97	0.92	0.92	0.95	0.89	0.89	0.9	1	0.98	0.98
BEDS	0.99	0.99	0.96	0.97	0.97	0.9	0.92	0.93	0.98	1	0.99
Wepisodes	0.99	0.99	0.96	0.97	0.97	0.89	0.93	0.94	0.98	0.99	1

Figure: Correlation matrix of original inputs and outputs

3.2 PCA and outlier

- Apply the PCA-based approach for aggregation proposed by Daraio and Simar (2007) and Nguyen and Zelenyuk (2021).

Table 2: Eigenvalues of PCA-based aggregation

	Eigenvalue	Information Explained (%)	Cumulative (%)
Labor			
	5.7154	95.2561	95.2561
	0.1919	3.1984	98.4546
	0.0364	0.6075	99.0621
	0.0309	0.5157	99.5778
	0.0229	0.3809	99.9587
	0.0025	0.0413	100.0000
Output			
	1.9751	98.7539	98.7539
	0.0249	1.2461	100.0000

Table 3: Correlations between the aggregated variable and original variables

Labor						
	MEO	TNUR	OPCS	DHP	ACS	DOS
AggLABOR	0.9925	0.9955	0.9593	0.9846	0.9898	0.9327
Output						
	OUT	WEPS				
AggOUT	0.9940	0.9935				

Table of contents

4.1 Model selection

- Likelihood ratio (LR) test between models and production functions:
  - Cobb-Douglas v.s. Translog;
  - ALS77 v.s. KGM91;
  - Restricted v.s. Unrestricted .

Table 5: Likelihood ratio test between restricted and unrestricted models

Models			*LR*	Degree of freedom	Decision
Production functions					
ALS77-CD	v.s.	ALS77-Translog	34.3222	6	Reject***
KGM91-CD	v.s.	KGM91-Translog	190.5308	6	Reject***
SFMs					
ALS77-CD	v.s.	KGM91-CD	259.563	8	Reject***
ALS77-Translog	v.s.	KGM91-Translog	415.7716	8	Reject***
Parsimonious KGM91					
KGM91-Translog-$\gamma = 0$[1]	v.s.	KGM91-Translog	414.369	7	Reject***
KGM91-CD-Restricted[2]	v.s.	KGM91-CD	210.1166	9	Reject***
KGM91-Translog-Restricted[2]	v.s.	KGM91-Translog	6.899	7	Not Reject

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

[1] Assuming all the coefficients of environmental factors, except the constant term, as zero.

[2] Assuming variables that are insignificant in the estimates of the unrestricted model as zero.

4.2 Production relationship

Table 6: Estimation of parameters

Model	OLS (robust s.e.)		ALS77		KGM91		
	Cobb-Douglas	Translog	Cobb-Douglas	Translog	Cobb-Douglas	Translog	Parsimonious-Translog
Frontier function							
ln BEDS	0.25***	2.27	0.25***	2.27	0.52***	5.40***	4.83***
	-0.07	-1.85	-0.06	-1.64	-0.05	-1.12	-1.02
ln Agglabours	0.03	2.35	0.03	2.35	0.04	2.76***	3.53***
	-0.06	-1.85	-0.05	-1.59	-0.04	-1.04	-1.02
ln SUPP	0.60***	-0.92	0.60***	-0.92	0.33***	-4.03***	-4.30***
	-0.05	-1.63	-0.04	-1.28	-0.03	-0.89	-0.87
ln BEDS * ln AggLABOR		0.08		0.08		0.20**	0.29***
		-0.17		-0.15		-0.09	-0.09
ln BEDS * ln SUPP		-0.21		-0.21		-0.39***	-0.27***
		-0.16		-0.14		-0.1	-0.07
ln AggLABOR * ln SUPP		-0.16		-0.16		-0.21***	-0.29***
		-0.12		-0.1		-0.06	-0.06
(ln BEDS)2		0.34		0.34		0.28*	
		-0.25		-0.25		-0.15	
(ln SUPP)2		0.14		0.14		0.38***	0.36***
		-0.13		-0.11		-0.07	-0.07
(ln AggLABOR)2		0.19		0.19		0.24**	0.26**
		-0.19		-0.17		-0.11	-0.11
Y2013	0.17***	0.14**	0.17***	0.14***	0.06	0.04	
	-0.06	-0.06	-0.05	-0.05	-0.04	-0.03	
Y2014	0.14**	0.11**	0.14***	0.11**	0.06*	0.04	
	-0.06	-0.06	-0.05	-0.05	-0.04	-0.03	
Y2015	0.13**	0.11**	0.13**	0.11**	0.05	0.04	
	-0.06	-0.05	-0.05	-0.05	-0.04	-0.03	
Constant	-11.10***	-1.58	-11.09***	-1.58	-7.91***	17.02***	20.52***
	-0.69	-11.03	-0.61	-8.93	-0.44	-6.04	-6.03

(Continued . . .)

4.2 Inefficiency function

Table 6 (Continued)

Model	OLS (robust s.e.)		ALS77		KGM91		
	Cobb-Douglas	Translog	Cobb-Douglas	Translog	Cobb-Douglas	Translog	Parsimonious-Translog
τ_i							
TEACH					-6.89	0.06	
					(.)	-0.09	
Remote					0.17*	0.06	
					-0.09	-0.04	
Small					-1.60***	3.64	
					-0.4	(.)	
Occup					-0.67**	-0.42***	-0.45***
					-0.27	-0.12	-0.11
Inout					13.28	7.67***	7.55***
					(.)	-0.41	-0.4
AveLOS					0.02**	0.03***	0.03***
					-0.01	-0.01	0
Tover					-0.03***	-0.02***	-0.02***
					0	0	0
Constant					2.37***	-2.77***	0.99***
					-0.49	-0.09	-0.07
σ_u							
Constant			-9.18	-9.59	-2.65***	-3.74***	-3.63***
			-70.11	-57.86	-0.21	-0.23	-0.21
σ_v							
Constant			-2.06***	-2.15***	-3.06***	-3.65***	-3.70***
			-0.08	-0.07	-0.1	-0.13	-0.14

Standard errors beneath
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

4.3 Inefficiency estimation

Table 7: Descriptive statistics for the estimated individual inefficiency

Models	Mean	Std Dev	Min	Q1	Q2	Q3	Max
Cobb-Douglas							
ALS77	0.81%	0.01%	0.77%	0.80%	0.81%	0.81%	0.85%
KGM91	21.44%	23.09%	0.67%	3.19%	9.52%	38.42%	77.79%
Translog							
ALS77	0.66%	0.01%	0.63%	0.65%	0.66%	0.66%	0.69%
KGM91	30.24%	25.95%	0.42%	3.45%	27.13%	52.49%	82.52%
KGM-P	32.48%	25.39%	1.20%	6.49%	30.48%	53.30%	83.57%

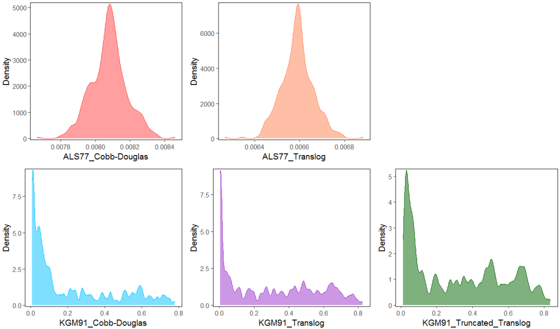

Figure: Estimated inefficiency level densities of selected models

4.3 Impact of the environmental variables

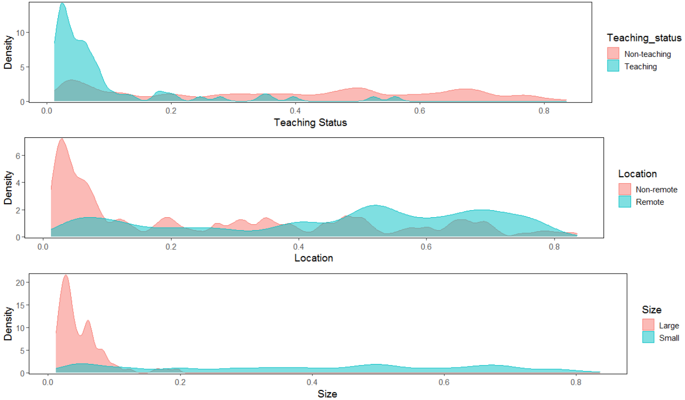

Figure: Inefficiency level densities of hospital groups by characteristics

Table 8: Li-test for distributions of estimated inefficiency by different groups

Groups	Li-test statistics	Bootstrap p-value	Decision
Non-teaching & Teaching	32.0051	0	Reject H_0
Non-remote & Remote	15.4081	0	Reject H_0
Large & Small	42.0575	0	Reject H_0

4.3 Impact of the environmental variables

- Though the character variables are restricted due to insignificance...
- Clear tendency could be found in the KPI variables of different groups.

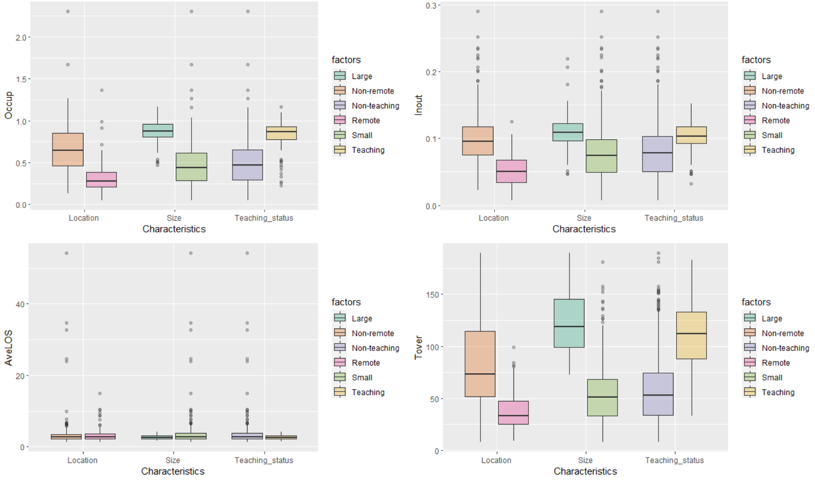

Figure: KPIs by characteristics

4.4 Key findings and suggestions

- Hospitals in larger size, in non-remote areas and with teaching function, tend to perform better than their peers, which indicates a scale effect existing in the hospital performance;
- The better performed groups may benefit from their higher occupancy and bed turnover rates, while their efficiency level can be further improved if they could reduce the proportion of inpatient admissions;
- Accordingly, one suggestion is decentralizing the inpatient care from large hospitals to more local hospitals.
  - The small and regional hospitals could then improve their efficiency by increasing the occupancy and turnover rates of beds;
  - Meanwhile, the large major hospitals may also benefit from a moderate proportion of inpatient cares.

Thank You!