CENTRE FOR HEALTH
PROGRAM EVALUATION

WORKING PAPER 135

Hospital Efficiency Measurement: Simple Ratios *vs* Frontier Methods

Mr Duncan Mortimer
Research Associate, School of Medicine,
Health Policy & Practice, University of East Anglia

Dr Stuart Peacock
Senior Lecturer, Health Economics Unit,
Centre for Health Program Evaluation

August, 2002
ISSN 1 876662 55 7

CENTRE PROFILE

The Centre for Health Program Evaluation (CHPE) is a research and teaching organisation established in 1990 to:

- undertake academic and applied research into health programs, health systems and current policy issues;
- develop appropriate evaluation methodologies; and
- promote the teaching of health economics and health program evaluation, in order to increase the supply of trained specialists and to improve the level of understanding of these disciplines within the health community.

The Centre comprises two independent research units, the Health Economics Unit (HEU) which is part of the Faculty of Business and Economics at Monash University, and the Program Evaluation Unit (PEU) which is part of the Department of Public Health at The University of Melbourne. The two units undertake their own individual work programs as well as collaborative research and teaching activities.

PUBLICATIONS

The views expressed in Centre publications are those of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of publications are encouraged to contact the author(s) with comments, criticisms and suggestions.

A list of the Centre's papers is provided inside the back cover. Further information and copies of the papers may be obtained by contacting:

The Co-ordinator
Centre for Health Program Evaluation
PO Box 477
West Heidelberg Vic 3081, Australia
Telephone + 61 3 9496 4433/4434 **Facsimile** + 61 3 9496 4424
E-mail CHPE@BusEco.monash.edu.au
Or by downloading from our website
Web Address http://chpe.buseco.monash.edu.au

ACKNOWLEDGMENTS

The Health Economics Unit of the CHPE is supported by Monash University.

The Program Evaluation Unit of the CHPE is supported by The University of Melbourne.

Both units obtain supplementary funding through national competitive grants and contract research.

The research described in this paper is made possible through the support of these bodies.

AUTHOR ACKNOWLEDGMENTS

The research reported herein was, in part, completed towards the author's Master of Economics (Honours) degree, awarded by the Department of Economics at Monash University. Professor Jeff Richardson and Dr. Stuart Peacock provided much helpful support and advice during completion of this degree. The Australian Health Ministers Advisory Council (AHMAC), the Monash University Research Fund (MURF), the National Health and Medical Research Council (NH&MRC) and the Centre for Health Program Evaluation (CHPE) provided the necessary financial support to allow completion of this report and related research projects. A version of this paper was presented at the 2nd Australia / New Zealand Health Services Research & Policy Conference, Wellington on the 4th December 2001.

Table of Contents

List of Tables

Abstract

Objectives: To provide a critical appraisal of the methods, results and policy value of two frontier methods for hospital efficiency measurement: data envelopment analysis (DEA) and stochastic frontier analysis (SFA). To compare the policy value of DEA- and SFA-based measures against more commonly used indicators of hospital performance.

Methods: Comparative analysis of DEA and SFA in estimating the relative efficiency of public hospitals in Victoria. Possible sources of measured inefficiency are investigated via the Battese and Coelli (1995) effects model in the case of SFA-based efficiency scores and via second-stage regressions in the case of DEA-based efficiency measures. The content and consistency of DEA- and SFA-based targets and measures are then compared against simple cost/output ratios.

Results: Moderate correspondence between DEA- and SFA-based efficiency scores with *measured* inefficiency at least partially attributable to between-hospital differences in casemix, stay-mix, quality of care, teaching/research activity and location. More or less the *same* set of hospital characteristics turned out to be important in explaining between-hospital variation in both DEA- and SFA-based measures of hospital efficiency. In short, results provided some reassurance that DEA and SFA are measuring closely related constructs, along similar dimensions. Perhaps surprisingly, there is *at least* as much common ground between simple cost/output ratios and SFA-based measures of hospital efficiency, as there is between DEA- and SFA-based alternatives.

Conclusions: Frontier-based measures of hospital performance are broadly consistent with simpler, more commonly available performance measures. However, consistency and precision are not the only considerations in selecting policy-ready performance measures and deliberations as to the policy-value of frontier- and ratio-based options should take account of both precision *and* content.

Hospital Efficiency Measurement: Simple Ratios *vs* Frontier Methods

Introduction

The success of policy in guiding the hospital sector towards best-practice depends on an ability to distinguish efficient from inefficient service providers. Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) provide tools with which to estimate the best-practice frontier and quantify industry-average and hospital-specific inefficiency. Whilst both DEA and SFA have potential applications in setting production targets, prospective payment rate setting, hospital 'report cards', and in evaluating the impact of past policy intervention; the relative precision and policy value of frontier methods remains in some doubt.

DEA and SFA employ quite distinct methodologies for frontier estimation and efficiency measurement, each with associated strengths and weaknesses, so that a trade-off exists in selecting the 'correct' approach. DEA offers the advantage of being relatively flexible in modelling the underlying production technology but makes no adjustment for random error. SFA explicitly models random error, but assumes a structure for the underlying production technology and then fits a curve. For the measurement of hospital efficiency (and many other applications), an ability to model random error *and* a flexible frontier are likely to be important. In other words, neither DEA nor SFA is ideally suited for the measurement of hospital efficiency.

It should, however, be emphasised that perfection is very often the enemy of the good. More commonly available performance indicators, such as cost per casemix-adjusted separation and observed to casemix-expected length of stay, are also subject to a number of shortcomings that limit interpretation and precision. DEA- and SFA-based measures offer two main advantages over simple ratios:

- **Frontier *vs* average indicators:** Frontier methods deliver benchmarks that reflect industry best-practice and efficiency scores that reflect deviations between observed and potential performance. In contrast, simple ratios benchmark hospitals against average industry behaviour. Industry-average benchmarks are quite useful for setting reimbursement rates but problems arise if we want to say anything about the gap between observed and potential performance, or about how hospital decision-makers might go about bridging this gap.

- **Multi-variate *vs* bi-variate indicators:** Whereas DEA and SFA evaluate hospital performance over multiple dimensions, simple ratios take account of only two summary dimensions: typically input (or cost) and output. Because it is more difficult to make multi-attribute comparisons using simple ratios, there is no guarantee that inefficient (efficient) hospitals lie above (below) the industry-average level of cost per output.

On the other hand, simple input-output ratios have the principle advantage of simplicity:

> “While a very simple and clearly defined objective *(or in this case indicator)* may not capture all the dimensions of interest, it has the countervailing advantage of making it

relatively easy to monitor success in those dimensions that it does capture" (Sugden and Williams, 1978 p. 6).

These strengths and weaknesses apply to greater or lesser extent in different policy and management applications. This study therefore has two primary concerns: (1) to try to clarify the relative merits and limitations of DEA- and SFA-based efficiency measures for a policy-relevant real-world application, and (2) to compare the merits and limitations of DEA- and SFA-based efficiency measures against simple cost/output ratios and other commonly used indicators of hospital performance.

Methods

DEA and SFA have much in common. Both yield relative efficiency ratings on a 0 (worst-practice) to 1 (best-practice) scale based on a comparison between the observed performance of individual production units and a best-practice frontier. More frequently, however, the literature emphasizes points of divergence between DEA and SFA. For the uninitiated, detailed descriptions of competing methods for frontier estimation and efficiency measurement are given in Lovell (1993), Coelli, Rao and Battese (1998), and Kumbhakar and Lovell (2000). DEA and SFA differ across two major dimensions:

- **Non-parametric *vs* parametric frontier estimation:** DEA employs flexible, non-parametric methods to construct the best-practice frontier and so allows the data to 'speak for themselves' (Bates, Baines and Whynes, 1996). In contrast, parametric methods such as SFA assume a structure for the best practice frontier and then fit a curve.

- **Deterministic *vs* stochastic efficiency measurement:** DEA assumes away random error and characterises deviations from the best-practice frontier as entirely due to inefficiency. In contrast, the stochastic frontier approach treats deviations from best-practice as comprising both random error (white noise) and inefficiency.

In an attempt to clarify the practical importance of these points of divergence, this paper derives DEA- and SFA-based efficiency measures for a sample of thirty-eight large Victorian public hospitals over the period 1993/94. The sample comprises all Group A and Group B hospitals operating in Victoria during 1993/94. Group A includes all Melbourne metropolitan teaching hospitals and Group B is the set of large regional base, and smaller suburban, acute care hospitals. Group A and B hospitals accounted for 87.9% of weighted inlier-equivalent separations (WIES), 88.4% of acute separations, and 82.6% of total operating expenditure for all Victorian acute care public hospitals.

To facilitate direct comparison between DEA- and SFA-based efficiency indices for multi-input/multi-output technology, we estimated input-oriented overall economic efficiency (EE_x comprising pure technical, scale and allocative efficiency effects) against DEA- and SFA-derived minimum cost frontiers. The resulting DEA cost-minimisation problem assumed a cost-reducing orientation, variable returns to scale (VRTS), and treated input-slacks as a component of allocative inefficiency. Input-prices used in DEA models were industry averages under the assumption that Victorian public hospitals face a standard set of exogenously determined input-prices. Data were predominantly sourced from *Victorian Hospital Comparative Data 1993/94* (DHCS, 1994). Variable definitions for the 'all output' DEA cost-minimisation model were:

- ***Outputs:*** WIES, Emergency/Casualty Occasions of Service, and Other Outpatient Occasions of Service.
- ***Inputs:*** Available Beds and Total Nursing, Medical/Clinical, & Non-medical Staff Cost.
- ***Input prices:*** Industry-average Nursing, Medical/Clinical, & Non-medical Salaries, Capital Expenditure per Available Bed.

Results from a narrower 'admitted inpatient' model - with WIES as the sole output and admitted inpatient (rather than total) staff costs reflecting the level and mix of labour inputs - are also reported. DEA models were run with the aid of one of two specialist DEA packages: DEAP Version 2.1 (Coelli, 1996a) and Warwick-DEA (WDEA) for Windows Version 1.02 (Thanassoulis, Athanassopoulos and Dyson, 1996).

The SFA model was operationalised to ensure comparability with 'all output' and 'admitted inpatient' versions of the DEA model outlined above, whilst also providing a flexible characterisation of the technology and operating environment facing Victorian public hospitals. As a starting point, the single-equation unrestricted translog cost frontier was estimated before paring back cross-products and higher-order terms to provide a more parsimonious model of hospital production. In general terms, the unrestricted model adopts a cost-minimising orientation, imposes no *a priori* restrictions on input-substitution, and permits VRTS. This approach sought to allow greater flexibility in the functional form used to fit the best-practice frontier under SFA. Variable definitions for the 'all output' and 'admitted inpatient' SFA models were as follows:

- ***Production Cost:*** Total Operating Expenditure *or* Admitted Inpatient Expenditure.
- ***Outputs:*** WIES, Emergency/Casualty Occasions of Service, and Other Outpatient Occasions of Service.
- ***Input Prices:*** Industry-average Nursing, Medical/Clinical, & Non-medical Salaries.
- ***Quasi-fixed Inputs*:** Available Beds.

Each of the SFA models approximates a pure technological cost function: mapping the relationship between production cost, outputs and input-prices. Note that input-prices will drop out of the regression when hospitals are assumed to face a standard set of exogenously determined input-prices. Note also that treating available beds as a quasi-fixed input and estimating the *variable* cost frontier quarantines measured inefficiency from between-hospital variation in capacity (Kumbhakar and Lovell, 2000). Stochastic frontier models were estimated with the aid of FRONTIER Version 4.1 (Coelli, 1996b).

For comparison purposes, ratios of observed to casemix-expected length of stay (acute LOS ratio), total operating expenditure (TOE) per WIES, and admitted inpatient expenditure (AIE) per WIES were taken directly from *Victorian Hospital Comparative Data 1993/94* (DHCS, 1994). Measures of hospital performance were then derived to give hospital performance as a percentage of industry-average performance (rather than group-average performance) for each ratio. A value of 100% on ratio-based performance measures indicates industry-average performance; higher (lower) values could be interpreted as indicating below (above) average levels of efficiency.

Results: DEA vs SFA

Both DEA and SFA have the potential to incorrectly estimate frontiers and deliver biased efficiency scores. Given the possibility of fairly substantial divergence between DEA and SFA results, stability within and correspondence between DEA- and SFA-based estimates should provide some reassurance that the two methods are measuring the same latent variable.

Over 250 DEA models were run to evaluate the practical significance of *a priori* assumptions and variable definitions in specifying and operationalising the final DEA model. DEA efficiency ranks were relatively stable under fairly severe changes in model specification. That is, rankings of hospitals by DEA efficiency scores did not change significantly under a range of modeling assumptions. Efficiency ranks and ratings derived from the 'all output' model, for example, varied only slightly under alternative modelling assumptions[1]:

• Input-reducing *vs* Output-increasing	Pearson's r ≥ 0.954** Spearman's rho ≥ 0.993**
• CRTS *vs* VRTS	Pearson's r ≥ 0.885** Spearman's rho ≥ 0.717**
• Non-discretionary *vs* Discretionary Capital	Pearson's r ≥ 0.898** Spearman's rho ≥ 0.824**
• Labour Cost *vs* Labour EFTs	Pearson's r ≥ 0.919** Spearman's rho ≥ 0.829**
• Hospital-specific *vs* Industry-av input-prices	Pearson's r ≥ 0.997** Spearman's rho ≥ 0.990**

Of the various possible SFA-based frontiers, the unrestricted translog with industry-average prices provides the closest approximation to the assumptions of the final DEA model. However, the additional complexity and flexibility of translog and restricted translog models returned implausible parameter estimates and very little in the way of additional explanatory power. More specifically, the inclusion of cross products and higher-order terms failed to significantly improve on model fit ($\chi^2_{\alpha = .05,\, df = 7} = 14.067 > 11.2564 = LR$) and output coefficients for translog and unrestricted translog models implied *negative* returns to scale. In short, the Cobb-Douglas frontier was a better choice with respect to both parsimony *and* interpretation.

Despite differences in parameter estimates describing the frontier technology, SFA-based efficiency ranks and ratings were relatively stable across translog, restricted translog and Cobb-Douglas models. Industry-average EE_x varied from 0.793 to 0.870 for 'all output' models, reflecting the relatively close correspondence between hospital-specific ranks and ratings under alternative model specifications (Pearson's r: 0.634 to 0.989, df = 36, p < 0.01; Spearman's rho: 0.545 to 0.988, n = 38, p < 0.01).

Industry-average and hospital-specific measures of overall economic efficiency derived from final DEA and SFA models are compared in tables 1.1 and 1.2 below. Results are reported for 'all output' and 'admitted inpatient' variants, with each DEA model matched to an equivalent SFA model. All hospital-specific scores have been de-identified, and hospital results are presented in random order.

[1] Where ** implies Pearson's r > 0.413, df = 36, p < 0.01; Spearman's rho > 0.415, n = 38, p < 0.01.

Table 1.1 DEA *vs* SFA - EE_x Summary

	N	Minimum	Maximum	Mean	SD
'All Output' DEA Model	38	0.455	1.000	0.8619	0.1548
'All Output' SFA Model	38	0.447	0.998	0.8117	0.1546
			Pearson's r = 0.654, df = 36, p < 0.01		
			Spearman's rho = 0.549, N = 38, p < 0.01		
			Mean Difference = 0.05, t = 2.403, df = 37, p < 0.05		
			Wilcoxen Signed-Ranks: z = -2.433, N = 38, p < 0.05		
'Admitted Inpatient' DEA Model	38	0.587	1.000	0.8331	0.1243
'Admitted Inpatient' SFA Model	38	0.349	1.000	0.8610	0.1395
			Pearson's r = 0.613, df = 36, p < 0.01		
			Spearman's rho = 0.478, N = 38, p < 0.01		
			Mean Difference = -0.03, t = -1.475, df = 37, p = 0.15		
			Wilcoxen Signed Ranks: z = -0.811, N = 38, p = 0.42		

Table 1.2 DEA *vs* SFA - EE_x Comparison

Hospital	'All Output' Model		'Admitted Inpatient' Model	
	DEA	SFA	DEA	SFA
1	1	0.998	0.83	0.95
2	0.699	0.871	0.832	0.875
3	1	0.862	0.745	0.85
4	0.836	0.944	0.896	0.906
5	1	0.859	1	0.877
6	1	0.695	0.587	0.811
7	0.812	0.912	0.68	0.908
8	0.604	0.719	0.823	0.947
9	1	0.965	1	0.97
10	0.833	0.819	1	0.786
11	0.948	0.967	1	0.999
12	1	0.96	0.939	0.975
13	0.575	0.572	0.701	0.805
14	0.905	0.789	0.905	0.888
15	0.655	0.698	0.681	0.722
16	0.858	0.785	0.86	0.839
17	1	0.975	0.789	1
18	0.806	0.964	0.875	0.913
19	1	0.97	0.948	0.949
20	0.455	0.593	0.638	0.349
21	0.807	0.745	0.822	0.75
22	1	0.656	0.599	0.558
23	0.844	0.755	0.795	0.773
24	1	0.904	1	0.899
25	1	0.996	1	0.942
26	1	0.683	0.829	0.701
27	1	0.939	0.975	0.956
28	0.897	0.97	0.915	0.853
29	0.919	0.97	0.884	0.984
30	0.766	0.689	0.858	0.872
31	0.633	0.454	0.632	0.712
32	0.944	0.641	0.783	0.873
33	0.497	0.447	0.938	1
34	0.918	0.768	0.863	1
35	0.905	0.778	0.613	0.612
36	1	0.991	0.77	1
37	0.846	0.885	0.858	0.928
38	0.789	0.656	0.793	0.986

The correspondence between DEA and SFA efficiency scores is broadly in line with results reported in the literature (see Mortimer, 2002 for a review). As expected, SFA models yield fewer frontier hospitals than their DEA counterparts simply by virtue of the link between dimensionality and discriminatory power in DEA models. DEA *vs* SFA comparisons delivered positive and statistically significant correlations at the 0.01 level for both 'all output' (Pearson's r = 0.654, df = 36, p < 0.01; Spearman's rho = 0.552, n = 38, p < 0.01) and 'admitted inpatient' models (Pearson's r = 0.613, df = 36, p < 0.01; Spearman's rho = 0.481, n = 38, p < 0.01). The mean difference between paired scores highlighted persistent gaps between DEA and SFA estimates of hospital efficiency. Hospital-specific efficiency scores obtained from DEA models were generally higher than SFA-based estimates for 'all output' models (t = 2.403, df = 37, p < 0.01) and generally lower than their SFA-based counterparts for 'admitted inpatient' models (t = - 1.475, df = 37, p = 0.15). Note that gaps between hospital-specific scores mirror differences between industry-average efficiency scores given in table 1.1 above.

Table 2.1 SFA Cobb-Douglas Hospital-Specific Effects Frontier[2]

Parameters	'All Output' Model		'Admitted Inpatient' Model	
β_0	0.669**	(0.10)	1.334**	(0.00)
β_1 ln WIES$_j$	0.983**	(0.03)	0.796**	(0.00)
β_2 ln E/COS$_j$	- 0.024**	(0.01)	restrict = 0	
β_3 ln OthOut$_j$	0.042*	(0.02)	restrict = 0	
β_4 ln AvBeds$_j$	0.058	(0.03)	0.252**	(0.00)
κ_0	- 0.446	(0.22)	- 1.091**	(0.18)
κ_1 TStaff/Bed	0.290**	(0.04)	0.246**	(0.04)
κ_2 %SameDaySeps	- 0.008*	(0.00)	0.009**	(0.00)
κ_3 %Non-PublicSeps	- 0.019**	(0.00)	- 0.006	(0.00)
κ_4 Regional	0.378**	(0.09)	0.095	(0.07)
κ_5 Specialist	0.172	(0.16)	0.182*	(0.09)
κ_6 DRG Weight	restrict = 0		restrict = 0	
κ_7 %LStayOutliers	restrict = 0		restrict = 0	
γ	1.000	(0.00)	1.000	(0.00)
σ_S^2	0.031	(0.01)	0.018	(0.00)
Log-likelihood	38.3921		52.1871	

All Models: $* \rightarrow |t \geq 2.056|$, df = 26, $p < 0.05$; $** \rightarrow |t \geq 2.779|$, df = 26, $p < 0.01$.

Results obtained from secondary Tobit regressions, in the case of DEA, and from the Battese and Coelli (1995) effects model, in the case of SFA, allow us to make inroads in identifying the source of *measured* inefficiency. The interpretation of the κ-parameters outlined in tables 2.1 and 2.2 is fairly straightforward[3]. In the 'all output' DEA model, higher levels of inefficiency are related to

[2] Note that the dependent variable is total operating expenditure + capital (TOE + K) for the 'all output' model and admitted inpatient expenditure (AIE) for the 'admitted inpatient' model.

[3] Strictly speaking, efficiency measurement against the SFA minimum cost frontier yields a ratio of actual to potential production cost. The resulting efficiency scores take a value greater than or equal to unity, with higher scores indicating lower levels of efficiency. In order to admit a consistent interpretation of DEA and SFA efficiency scores, EEx is reported as the ratio of potential to actual production cost with a value between zero and one. Note, however, that the sign of the SFA κ-parameters should be reversed in order to allow direct comparison against parameter estimates obtained from second-stage Tobit regressions on DEA efficiency scores.

higher ratios of total staff per bed[4], lower shares of same day and non-public separations, and regional (as opposed to metropolitan) location. More or less the *same* set of hospital characteristics turned out to be important in explaining between hospital variation in SFA-based measures of hospital efficiency. In the 'all output' SFA model, higher levels of inefficiency are related to higher ratios of total staff per bed, lower shares of same day and non-public separations, and regional (as opposed to metropolitan) location. In short, *measured* inefficiency is at least partially attributable to between-hospital variation in case-mix, quality of care, location and/or teaching and research activity.

Table 2.2 DEA Second-Stage Tobit Regressions

Parameters	'All Output' Model		'Admitted Inpatient' Model	
κ_0	1.658**	(0.25)	0.866**	(0.18)
κ_1 TStaff/Bed	-0.115**	(0.03)	-0.043*	(0.02)
κ_2 %SameDaySeps	0.010**	(0.00)	0.001	(0.00)
κ_3 %Non-PublicSeps	0.000	(0.00)	0.001	(0.00)
κ_4 Regional	-0.189**	(0.05)	-0.077	(0.04)
κ_5 Specialist	-0.041	(0.08)	-0.230**	(0.06)
κ_6 DRG Weight	-0.584*	(0.24)	0.142	(0.19)
κ_7 %LStayOutliers	-0.180*	(0.01)	0.003	(0.01)
σ^2	0.010	(0.01)	0.094	(0.01)
Log-likelihood	97.8452		123.1136	

All Models: $* \rightarrow |z \geq 1.960|, p < 0.05$; $** \rightarrow |z \geq 2.576|, p < 0.01$.

Results also flagged a number of issues for efficiency measurement in the hospital sector, particularly related to the stochastic frontier approach. The variance parameter in the SFA model: $\gamma \equiv \sigma^2 / \sigma_S^2$ reflects the divergence between deterministic and stochastic frontiers. "A value for γ of zero indicates that the deviations from the frontier are due entirely to noise, while a value of one would indicate that all deviations are due to inefficiency" (Coelli, Rao and Battese, 1998 p. 188). As such, γ also indirectly reflects the importance of random error in interpreting discrepancies between DEA and SFA estimates of (in)efficiency. Interestingly, $\gamma \cong 1$ for every one of the SFA models. The implication is that SFA attributes *all* residual variation in production costs to inefficiency rather than to error. Note that this is unlikely to reflect a lack of measurement error in the data or the absence of random fluctuations in hospital performance. It seems more likely that SFA is unable to distinguish between error and inefficiency in the current application. Banker, Gahd and Gorr (1993) report similar problems for simulated data:

> "...SFA fails to decompose deviations into efficiency and measurement error components (it assumes that deviations from the frontier are either totally due to measurement errors or totally due to inefficiency). …failure to decompose errors very likely led to overall poor performance relative to DEA" (Banker, Gahd and Gorr, 1993 pp. 332 & 341).

Based on *a priori* deliberations and the DEA *vs* SFA comparisons detailed above, the following general conclusions might be drawn. Firstly, SFA's failure to separate noise and inefficiency

[4] Note that the ratio of staff per bed is included primarily as a proxy for quality of care but admits an alternative interpretation as a labour productivity ratio if the number of hospital beds reflects throughput. Similarly, the regional/metro dummy might reflect uncontrolled variation in teaching and research activity rather than the assumed location effect.

eliminates one of its chief selling points: the ability to cope with random error. Secondly, the extent of error due to incorrect functional form in SFA is unobservable without further information as to the true structure of the underlying production technology. Thirdly, DEA efficiency ranks are relatively robust with respect to fairly severe changes in model specification. Finally, the moderate correspondence between DEA- and SFA-based measures of EE_X, together with broadly comparable results in identifying the source of measured inefficiency, suggest that the two approaches are measuring 'similar' latent efficiency varibles. Unfortunately, the extent of correspondence and corroboration is not nearly strong enough to suggest that DEA- and SFA-based measures are interchangeable and at least one approach delivers biased approximations of a common underlying construct.

Results: DEA / SFA *vs* Simple Ratios

Tables 3.1 and 3.2 report industry-average and hospital-specific measures of total operating expenditure (TOE) per WIES, admitted inpatient expenditure (AIE) per WIES, and observed to casemix-expected length of stay (acute LOS ratio). All hospital-specific data have again been de-identified and are presented in random order. Table 3.2 reports hospital performance as a percentage of industry-average performance for each ratio. No statistically significant correlations linked the acute LOS ratio with DEA-based, SFA-based or other ratio-based measures of hospital performance. For all remaining pair-wise comparisons, correlations confirmed the expected directional relationships.

Table 3.1 Simple Ratios – Summary

	N	Minimum	Maximum	Mean	SD
TOE per WIES	38	$2,410	$5,805	$3,374	$767.59
AIE per WIES	38	$2,025	$3,881	$2,461	$374.54
Acute LOS Ratio	38	0.82	1.23	0.97	0.01

A positive and statistically significant relationship was evident between TOE per WIES- and AIE per WIES-based performance measures (Pearson's r: 0.616, df = 36, $p < 0.01$). Correlations between DEA-based efficiency scores and both TOE per WIES- and AIE per WIES-based performance measures were perhaps weaker than expected but statistically significant and in the expected direction (Pearson's r: -0.520 to -0.544, df = 36, $p < 0.01$; Spearman's rho: -0.436 to -0.521, n = 38, $p < 0.01$). Correlations between SFA-based efficiency scores and both TOE per WIES- and AIE per WIES-based measures revealed the strongest link between any of our competing measures (Pearson's r: -0.734 to -0.819, df = 36, $p < 0.01$; Spearman's rho: -0.690 to -0.902, n = 38, $p < 0.01$).

It is worth emphasizing that SFA-based efficiency scores were most closely aligned to TOE per WIES- and AIE per WIES-based performance measures. In other words, results indicate that there is *at least* as much common ground between SFA-based efficiency scores and performance measures based on simple cost/output ratios, as there is between DEA- and SFA-based alternatives.

Table 3.2 Simple Ratios & Ratio-based Performance Measures

Hospital	TOE per WIES		AIE per WIES		Acute LOS Ratio	
1	$5,193	154%	$2,197	89%	1.17	121%
2	$3,873	115%	$2,848	116%	0.86	89%
3	$3,693	109%	$3,421	139%	0.98	101%
4	$2,986	89%	$2,455	100%	0.86	89%
5	$4,354	129%	$2,798	114%	0.84	87%
6	$2,852	85%	$2,279	93%	0.92	95%
7	$3,625	107%	$2,248	91%	0.97	100%
8	$4,014	119%	$2,419	98%	0.91	94%
9	$3,240	96%	$2,339	95%	0.82	85%
10	$3,692	109%	$2,279	93%	1.02	105%
11	$2,867	85%	$2,150	87%	0.93	96%
12	$2,767	82%	$2,303	94%	0.89	92%
13	$3,743	111%	$2,429	99%	0.88	91%
14	$3,396	101%	$2,459	100%	0.87	90%
15	$3,660	108%	$2,670	108%	0.96	99%
16	$2,643	78%	$2,025	82%	0.84	87%
17	$2,558	76%	$2,089	85%	0.94	97%
18	$2,718	81%	$2,210	90%	0.91	94%
19	$2,942	87%	$2,246	91%	0.97	100%
20	$2,823	84%	$2,097	85%	0.9	93%
21	$3,485	103%	$2,722	111%	1.16	120%
22	$2,892	86%	$2,351	96%	1	103%
23	$3,634	108%	$2,718	110%	0.89	92%
24	$3,200	95%	$2,517	102%	1.07	110%
25	$5,805	172%	$2,814	114%	1.07	110%
26	$2,840	84%	$2,257	92%	0.92	95%
27	$2,410	71%	$2,238	91%	0.97	100%
28	$5,353	159%	$3,881	158%	1.01	104%
29	$3,076	91%	$2,668	108%	1.23	127%
30	$2,830	84%	$2,089	85%	1.03	106%
31	$2,569	76%	$2,191	89%	1	103%
32	$3,090	92%	$2,421	98%	1.02	105%
33	$3,040	90%	$2,403	98%	0.99	102%
34	$3,848	114%	$3,015	123%	1.06	109%
35	$3,084	91%	$2,418	98%	0.99	102%
36	$2,909	86%	$2,323	94%	1.01	104%
37	$3,115	92%	$2,353	96%	0.82	85%
38	$3,400	101%	$2,160	88%	1.02	105%

Adapted from DHCS (1994), Victorian Hospital Comparative Data 1993/94.

Discussion

So what does all this mean for policy? Firstly, it should be relatively unsurprising that the raw DEA- and SFA-based scores reported above fail to provide perfect policy-ready measures of hospital efficiency. Various errors and biases weaken the link between DEA- and SFA-based estimates and the underlying 'target' construct of hospital efficiency. Blunt adjustments for quality of care, for example, fail to translate an index of hospital activity (casemix-adjusted throughput) into an index of health gain (outcome).

More generally, between-hospital variation in case-mix, stay-mix, and quality of care, as well as teaching/research activity and/or location, has been carried through and preserved in the efficiency scores reported above. Note, however, that DEA- and SFA-based measures are most usefully held against existing measurement options - rather than some hypothetical gold standard - for comparison. It is entirely possible, for example, that imperfect DEA- and SFA-based efficiency measures will still out-perform simple cost/output ratios. Not much can be said about criterion validity without having access to our hypothetical gold standard. However, a good deal can be said about content and construct[5] validity and each option's *potential* to deliver policy-ready performance measures for use in real-world policy and management applications.

Monitoring, evaluation and benchmarking

For applications requiring *summary* measures of managerial performance, it is important to control for between-hospital variation in factors that lie beyond managerial control (such as casemix, stay-mix, teaching/research activity and hospital location). Simple cost/output ratios are limited in the extent to which they can control for exogenous hospital characteristics or features of the production environment. One option is to adjust the denominator of simple cost/output ratios for between-hospital variation in casemix and/or stay-mix (ie. to use casemix-adjusted separations – or, in the case of Victoria, WIES – to index hospital output). A second option is to separate industry samples into smaller comparison groups that are much more homogenous with respect to production scale, production scope, teaching/research activity, and/or hospital location.

On the DEA front, hospital characteristics – such as capacity, quality of care and teaching/research activity – can be explicitly included in DEA models as discretionary inputs/outputs, non-discretionary inputs/outputs, non-discretionary 'control' variables, or by adding or modifying constraints (Coelli, Rao and Battese, 1998). Unfortunately, the inclusion of additional variables and constraints tends to reduce discriminatory power for a given sample size (Productivity Commission, 1997). A second possibility is to run a second stage to the basic DEA model, allowing simultaneous adjustment for any number of continuous and categorical variables (Coelli, Rao and Battese, 1998). In a similar vein, control variables can either be explicitly included in SFA-models and/or accommodated via post-hoc adjustment of efficiency scores.

[5] Construct validity comprises both convergent validity: the extent to which a measure behaves in concordance with the underlying construct, and discriminant validity: the extent to which a measure is independent of theoretically unrelated variables. The frontier- and ratio-based measures outlined above map a purely technical relationship between production cost, outputs and input-prices. In other words, each of our measures targets (approximately) the same underlying construct. If performance measures are to be of any value whatsoever, competing measures taking aim at the same target should converge in the same direction (ie. towards the target construct). In a similar vein, conceptually different performance measures (taking aim at different targets) might be expected diverge from one another as they converge on their own respective targets. In other words, results should differ if different measures access different constructs.

Note that, whilst both DEA and SFA are better equipped to deal with exogenous factors that might confound performance measures, they are also both subject to the sort of biases and errors that arise due to common data problems. In other words, DEA and SFA are able to incorporate currently available indicators of quality of care and teaching/research activity in a summary measure of hospital performance. However, neither method is able to remedy the usual problems associated with measuring quality of care and teaching/research activity in the first place.

Micro-management and target-setting

At first glance, indicators of factor productivity such as nursing staff per WIES, average length of stay, and occupancy rates provide quite useful tools for hospital management in setting targets for staff/patient ratios, in formulating discharge policies, and in adjusting hospital capacity. Note, however, that single-factor productivity ratios provide no information as to the rate at which nursing staff might be substituted for hotel or medical support staff with output held constant. Nor are such indicators likely to be useful for strategic decisions regarding production scale and production scope. DEA, in particular, is much more informative in setting targets and in informing micro-management decisions.

By constructing frontiers under different returns to scale assumptions, DEA is able to separate hospital-specific measures of pure technical, scale and allocative efficiency. Moreover, it is able to specify whether scale inefficient hospitals are currently subject to increasing or decreasing returns to scale. DEA also provides best-practice input and output targets, delineating the sort of adjustments in the level and mix of either inputs or outputs that would be required to achieve best-practice (Cooper, Seiford and Tone, 2000). Perhaps most importantly for managerial applications, DEA identifies each hospital's best-practice 'peers' to designate role models for under-performing hospitals.

Prospective payment rate-setting

A number of authors have argued that DEA- and SFA-based measures are insufficiently precise to be applied in prospective payment rate-setting (eg. Newhouse, 1994). The relative merits of frontier- and ratio-based measures outlined above stand even with respect to this most contentious of applications. Note, however, that rate-setting against the best-practice frontier would usually imply a more stringent set of incentives than existing, average performance standards. Results from the purely technical models outlined above, for example, imply that Victorian public hospitals incurred costs that were, in some cases, substantially higher than frontier costs. Any move to set reimbursement rates against the best-practice frontier should therefore cushion the impact of imposing a more stringent set of incentives (even after correcting biases to obtain policy-ready measures of hospital efficiency).

Conclusions

Frontier-based measures of hospital performance are broadly consistent with simpler, but commonly available, cost/output ratios. Unfortunately, the degree of correspondence between DEA-based, SFA-based and simple ratio measures is not so great as to suggest that they are interchangeable. Calls for parallel application of competing methods (to cross check results) are already being heeded, with the result that various studies have confirmed that the precision (and, therefore, policy value) of competing approaches is context specific (see Mortimer, 2002 for a review). Fewer studies have recognised that policy value depends on both precision and content.

Results indicate that no clear 'winner' - with respect to precision - can be identified for the measurement of hospital efficiency. It is however, quite likely, that each of our competing measures will have a comparative advantage - with respect to content - in a particular policy or management application. DEA, for example, has certain advantages for micro-management and target-setting purposes. In such circumstances, DEA might be installed as the front-runner for micro-management and target setting purposes and should be preferred unless cross-checking reveals substantial, *unexplained* deviations from ratio- or SFA-based alternatives. More generally, the practice of cross checking and deliberations on the policy-value of competing measures should follow a more principled approach to consider differences in both the precision *and* content of all available options.

References

Banker, R.D., Gadh, V.M., Gorr, W.L. (1993). A Monte Carlo comparison of two production frontier estimation methods: Corrected ordinary least squares and data envelopment analysis, *European Journal of Operational Research* 67 (3) 332-343.

Bates, J.M., Baines, D., Whynes, D.K. (1996). Measuring the efficiency of prescribing by general practitioners, *Journal of the Operational Research Society* 47 (12) 1443-51.

Battese, G.E., Coelli, T.J. (1995). A model for technical efficiency effects in a stochastic frontier production function for panel data, *Empirical Economics* 20 325-332.

Coelli, T.J. (1996a). *A Guide to DEAP Version 2.1: A Data Envelopment Analysis Computer Program.* CEPA Working Paper 96/08, Armidale: Department of Econometrics, University of New England.

Coelli, T.J. (1996b). *A Guide to Frontier Version 4.1: A Computer Program for Frontier Production Function Estimation*. CEPA Working Paper 96/07, Armidale: Department of Econometrics, University of New England.

Coelli, T., Rao, D.S.P., Battese, G.E. (1998). *An Introduction to Efficiency and Productivity Analysis*, Boston/Dordrecht/London: Kluwer Academic.

Cooper, W.W., Seiford, L.M., Tone, K. (2000). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, Boston/Dordrecht/London: Kluwer Academic.

DHCS (1994). *Victorian Hospital Comparative Data 1993/94*, Melbourne: Department of Health and Community Services.

Kumbhakar, S.C., Lovell, C.A.K. (2000). *Stochastic Frontier Analysis*, Cambridge and Melbourne: Cambrigde University Press.

Lovell, C.A.K. (1993). Production frontiers and productive efficiency, In: Fried, H.O., Lovell, C.A.K. & Schmidt, S.S. (Eds.) *The measurement of productive efficiency: Techniques and applications*, Oxford University Press, New York & Oxford, pp. 3-67.

Newhouse, J.P. (1994). Frontier estimation: How useful a tool for health economics?, *Journal of Health Economics* 13 317-322.

Mortimer, D. (2002). *Competing Methods for Efficiency Measurement: A Systematic Review of Direct DEA vs SFA/DFA Comparisons*. Centre for Health Program Evaluation Working Paper: Monash University.

Productivity Commission, Steering Committee for the Review of Commonwealth/State Service Provision (1997). *Data envelopment analysis: A technique for measuring the efficiency of government service delivery*, Canberra: AGPS.

Sugden, R., Williams, A. (1978). *The principles of practical cost-benefit analysis,* New York: Oxford University Press.

Thanassoulis, E., Athanassopoulos, A., Dyson, R.G. (1996). *Warwick-DEA (WDEA) for Windows Version 1.02*, United Kingdom: University of Warwick.