
G Referee reports on modelling

G.1 Report from Adjunct Professor Tim Coelli

This study of hospital performance in Australia sets a new benchmark in terms of the sample coverage and the range of input, output and control variables included in the econometric model. The Productivity Commission team has worked hard at producing the best empirical model possible, in the face of challenging time and data constraints. However, no empirical study is perfect, so in my brief discussion below I provide my assessment of the analysis, pointing out what is to be commended and what can perhaps be improved.

Data sample: The sample size of 508 is more than sufficient to allow the Productivity Commission (PC) to estimate an econometric model that involves a flexible functional form and a number of important input, output and control variables. The main concern with the sample, as it stands, relates to a high non-response rate on the part of not-for-profit hospitals. Hence the private hospitals in the sample are mostly for-profit hospitals, and the results obtained should be viewed in this light. However, my experience with not-for-profit hospitals (mostly run by church groups) is that they tend to put extra resources into non-medical services and hence I do not expect them to normally have much influence on the position of the best-practice frontier. Hence, I expect that their low sample representation is unlikely to significantly affect the efficiency scores of the remaining hospitals in the sample.

Frontier methodology: There are two frontier estimation methods that are commonly used in the literature: data envelopment analysis (DEA) and stochastic frontier analysis (SFA). DEA is a linear programming method that has the advantage that no particular functional form needs to be specified. However, SFA is an econometric method that is less susceptible to the effects of data noise and outliers and which also allows one to easily incorporate control variables that involve categorical and ratio data. Hence the choice of SFA is appropriate for this study.

Functional form: The translog function form is a flexible second-order functional form that can accommodate a range of scale and substitution possibilities, and hence is a good choice in my assessment.

Output measures: The output measures involve a number of categories of admitted and non-admitted separations, with the former casemix-adjusted. The level of detail is substantially better than many past studies of hospital efficiency. The authors emphasize the point that these are measures of intermediate outputs rather than incremental health benefits derived from the services. However, this is standard practice in this literature, given the very substantial challenges that would be involved in attempting to derive these latter output measures.

Input measures: The input measures include three categories of staff members (nursing, diagnostic and other), three monetary measures of non-staff variable inputs (drugs, medical and surgical supplies and other) along with the number of beds. This group of input measures is better than that used in the majority of past studies, but can still be improved upon (given access to better data). In particular, the beds measure treats an intensive care bed no differently to a standard bed, and the staff measures exclude doctors. These issues could introduce some biases in efficiency estimates if the casemix weights (used to define the output measures) include allowances for the extra capital costs associated with complex cases, and if there are differences among hospitals in the degree to which doctors versus nurses undertake certain “grey area” tasks.

Quality measures: Quality issues have been often overlooked in past studies of health sector efficiency. The PC is to be commended for their efforts in this regard. The inclusion of a mortality rate measure that is adjusted for patient risk characteristics is not a perfect measure, but should go a long way to capturing any notable variations in the effects of service quality upon efficiency potentials.

Control measures: The PC has considered a wide range of exogenous control measures that could potentially be affecting efficiency potentials, including network membership, accident and emergency rates, and so on. These measures help the analyst to avoid labelling a hospital as being “inefficient” when they may be using more resources per unit output because they face different operating conditions relative to other members of the sample.

Finally, I should emphasize a number of points. First, most if not all of the comments made above are also mentioned in the main report. Second, time and data constraints have clearly placed limits on the empirical analysis in this report. Third, I look forward to seeing what is produced in the supplementary report that is due to be released in March 2010, which will involve data from additional years and will

also involve the investigation of some alternative models and a more detailed investigation of the effects of scale on hospital performance.

G.2 Report from Professor Jim Butler

The estimation of hospital production functions and hospital cost functions is a complicated exercise. These complications arise not just because hospitals are multi-product organisations, but because of the large range and diversity of the outputs they produce. The ‘treated patient’ is not a homogeneous unit of output but differs according to the illness or illnesses with which they present, the severity of those illnesses, the range of treatments available and which are selected, and patient characteristics such as age, sex and frailty. The econometric modeller then faces a quandary. Working with a sufficiently large number of output categories to minimise heterogeneity within those categories will lead to a large number of parameters to be estimated, especially if the specification involves a flexible functional form. However, the pursuit of parameter parsimony, which requires a smaller number of output categories, introduces more heterogeneity into the output categories.

In addition to this type of conceptual difficulty, there are difficulties arising out of the institutional arrangements for the provision of hospital services in Australia (e.g. the inclusion of medical service costs in hospital costs for public hospitals but not private hospitals), the lack of a ‘pure’ separation between type of hospital ownership and the funding status of patients, and the paucity of data on capital costs.

Notwithstanding these difficulties, and the tight timelines within which the Commission was working, it has produced a high quality and interesting piece of work comparing technical efficiency in public and private hospitals using data on 508 hospitals for 2006-07. Using stochastic frontier analysis and several functional forms for a production function, the analysis concludes that the technical efficiency of public hospitals and private hospitals is similar. Across all hospital size groupings, public hospitals have a mean efficiency score of 0.797 and private hospitals 0.750 suggesting slightly superior performance by public hospitals (table 8.5 — there was virtually no difference between the scores for for-profit and not-for-profit private hospitals). Given the difficulties associated with empirical work in this area mentioned above (and that list is not exhaustive), the description of these mean scores for public and private hospitals as ‘similar’ is a judicious call of the results. The only size grouping where a more marked difference between the scores for public and private hospitals emerges is the ‘small and very small’ hospitals category with mean public and private hospital scores of 0.788 and 0.641

respectively (table 8.5). But with these results and others, one should bear in mind the possibility of self-selection bias in the private hospital sample of 122 hospitals as participation in the study by private hospitals was voluntary. The Report does specifically mention this limitation.

An innovative aspect of this study is its distinction between public hospitals and public contract hospitals. The latter are privately owned institutions whose caseload comprises a large proportion of public patients treated under contract from government (on average, public patients comprise 77.9% of the caseload of public hospitals cf. 77.4% for public contract hospitals – see Table 8.2). There is virtually no difference in the technical efficiency scores between these two types of institution (0.797 for public hospitals, 0.800 for public contract hospitals — table 8.5). While the sample size for public contract hospitals is not large (n=18), this result is of some interest.

A result which has perhaps been somewhat underplayed is the absence of any significant effect of risk-adjusted mortality ratios in the production model. The Commission has constructed a predicted value of the mortality rate (proportion of patients discharged dead) for each hospital using a Tobit regression with various factors exogenous to the hospital as regressors. These predicted values are then used to construct a risk-adjusted mortality ratio for each hospital which is used in the production function to investigate possible quantity/quality trade-offs in hospital production. The absence of a statistically significant effect here is potentially a policy-significant result — there is no evidence that hospitals in this study attain higher output levels or improve technical efficiency by allowing quality to deteriorate.

In concluding, two important limitations of this study should be noted. First, it is based upon data for only one year. Replication of the analyses using data from other years may instil more confidence in the results. Second, the study investigates only technical efficiency and not cost efficiency. As the Commission notes, a hospital's performance with respect to technical efficiency may differ from its performance with respect to cost efficiency, so results on the latter would provide a more complete picture of hospital performance. The Report indicates that both of these limitations will be addressed in further analyses, the results of which will be available in a supplementary report in March 2010. If those analyses are conducted to the same standard as the analyses presented in this Report, they will undoubtedly provide a very useful addition to the stock of knowledge in this field.