Australian Bureau of Statistics

# WORKING PAPERS IN ECONOMETRICS AND APPLIED STATISTICS

Working Paper No. 98/3

# COMPARING TECHNIQUES FOR MEASURING THE EFFICIENCY AND PRODUCTIVITY OF AUSTRALIAN PRIVATE HOSPITALS

Richard Webster

Steven Kennedy

Leanne Johnson

November 1998

ABS Catalogue No. 1351.0

EMBARGO: 11:30 AM (CANBERRA TIME) FRI 20 NOV 1998

Working Paper No. 98/3

# Comparing Techniques for Measuring the Efficiency and Productivity of Australian Private Hospitals

Richard Webster, Steven Kennedy and Leanne Johnson

November 1998

*This Working Paper Series is intended to make the results of current research within the Australian Bureau of Statistics available to other interested parties. The aim is to present accounts of developments and research or analysis of an experimental nature so as to encourage discussion and comment.*

*Views expressed in this paper are those of the authors and do not necessarily represent those of the Australian Bureau of Statistics. Where quoted or used, they should be attributed clearly to the authors.*

**ABS Catalogue No. 1351.0**

Produced by the Australian Bureau of Statistics

# Contents

*INQUIRIES*

- *For further information about the contents of this Working Paper, please contact the authors:*

**Steven Kennedy** — *telephone:* Canberra 02 6252 6100 *or facsimile:* 02 6252 8015 *or e-mail:* steven.kennedy@abs.gov.au
**Richard Webster** — *telephone:* Canberra 02 6252 6804 *or facsimile:* 02 6252 8015 *or e-mail:* richard.webster@abs.gov.au


- *For information on the* **Working Papers in Econometrics and Applied Statistics**, *please contact the Managing Editor:*
**Mark T. Jones** — *telephone:* Canberra 02 6252 7516 *or facsimile:* 02 6252 8015 *or e-mail:* mark.jones@abs.gov.au *or write* c/- Analytical Services Section (W42b), Australian Bureau of Statistics, PO Box 10, Belconnen, ACT 2616, Australia.

- *For information about other ABS statistics and services* please refer to the back cover of this publication.

# List of abbreviations and acronyms

| | |
|---|---|
| ABS | Australian Bureau of Statistics |
| COLS | Corrected Ordinary Least Squares |
| CRS | Constant Returns to Scale |
| DEA | Data Envelopment Analysis |
| DMU | Decision Making Unit |
| DRG | Diagnosis Related Group |
| EE | Economic Efficiency |
| EKS | Elteto, Koves and Szulc |
| FP | For-profit |
| FTE | Full-Time Equivalent |
| HASAC | Health and Allied Services Advisory Council |
| ICD9-CM Modification | International Classification of Diseases Volume 9 — Clinical |
| ID | Identification |
| LH | Linear Homogeneity |
| LP | Linear Program |
| MLE | Maximum Likelihood Estimation |
| MPSS | Most Productive Scale Size |
| NFP | Not-for-profit |
| NFPo | Not-for-profit, other |
| NFPr | Not-for-profit, religious or charitable |
| NIRS | Non Increasing Returns to Scale |
| OBD | Occupied Bed Day |
| OBS | Observation |
| OLS | Ordinary Least Squares |
| PHEC | Private Health Establishments Collection |
| PTE | Part-Time Equivalent |
| SBD | Surgical Bed Day |
| SD | Standard Deviation |
| SE | Standard Error |
| SFA | Stochastic Frontier Analysis |
| SMO | Salaried Medical Officer |
| sq. | squared |
| TE | Technical Efficiency |
| VMO | Visiting Medical Officer |
| VRS | Variable Returns to Scale |

# Abstract

The Australian Bureau of Statistics (ABS) is trying to improve its measurement of inputs, outputs and productivity for the non-market sector and, more generally, for the services sector of the Australian economy. In 1996–97, this project focussed on the health services industry. The analysis reported in this paper applies a range of firm-level efficiency-measurement techniques to a unit record dataset for the Australian private hospital industry. Firm-level analyses of this kind are being applied by influential members of the ABS user community. This private hospitals study has three aims:

- to explore the differences in *assumptions* made by the various techniques and the differences in *results* they yield;
- to *test* the assumptions (relating to homogeneity of the industry, economies of scale, etc.) that underlie ABS standard methods for analysing *aggregate* productivity; and
- to understand the ways in which the characteristics of a *dataset* can affect the application of these analytical techniques.

Two types of techniques are used in the analyses: a non-parameteric technique known as Data Envelopment Analysis (DEA), and two parametric technique — Stochastic Frontier Analysis (SFA) and Ordinary Least Squares (OLS) regression. The benefits and shortcomings of each technique are discussed in general terms, then each is applied to a number of model specifications using different combinations of input and output variables drawn from the private hospitals dataset.

In this analysis the DEA technique is not robust to changes in the number or construction of variables. Conclusions about the relative efficiency of sub-samples and the efficiency ranking of individual hospitals change appreciably when the choice of variables is altered. Thus, if DEA is to be used for monitoring the performance of individual firms or for assessing patterns of efficiency across the whole population of firms, extrinsic judgements must be brought to bear when selecting the input and output variables.

Results from the parametric estimation techniques (OLS and SFA) also suggest a lack of robustness to changes in model specification. Conclusions about the structure of production, the pattern of productivity and the performance of individual hospitals can all change when the model is altered. Analyses of sub-populations (characterised by hospitals' size or profit-making status) indicate that individual hospitals may be engaging in substantially different activities from one another. This brings into question the validity of an aggregate productivity analysis of the kind traditionally applied by the ABS.

The analysis also highlights the inability of the dataset and our models *in combination* to completely characterise the private hospitals industry. In part, this is due to shortcomings of the frontier estimation techniques. However, it also suggests minor changes to the private hospitals census which could enhance the value to analysts who are interested in developing measures of unit level hospital efficiency.

---

An earlier version of this paper was presented to the ABS' Methodology Advisory Committee where Annette Dobson was the discussant. The authors would also like to thank Tim Coelli, Kathy Kang, Marelle Rawson, John Goss, Ken Tallis, Ben Phillips and Keith Carter for helpful comments and assistance with this research project. Responsibility for any mistakes or omissions is entirely our own.

---

# 1 Introduction

This paper examines two types of methodologies for measuring the efficiency and productivity of Australian private acute care hospitals: a non-parametric technique known as Data Envelopment Analysis (DEA); and parametric techniques including Stochastic Frontier Analysis (SFA) and Ordinary Least Squares regression (OLS).

The impetus for this paper is a current Australian Bureau of Statistics (ABS) project that aims to improve the measurement of outputs and inputs for the non-market sector, and more generally for the services sector, of the Australian economy. In the first phase of the project, the Methodology Division of the ABS constructed some experimental estimates of outputs and inputs for the health industry. The primary measure of hospital output was constructed using Diagnostic Related Group (DRG) cost weights to aggregate the treatments provided by hospitals – though it was noted that other measures such as occupied bed days and deflated patient revenue had also been used by other investigators to measure output. The rich ABS dataset on private hospitals provided a unique opportunity to compare direct or volume-based measures of output such as DRG based measures or occupied bed days with measures that arise out of the market context in which private hospitals operate, that is, patient revenue-based measures. This enables comparison of (and, to a degree can validate) our use of certain output indicators for the non-market sector and more broadly for the whole service sector of the economy.

The recent emergence of unit record or firm-based frontier techniques (which not only provide important firm-level information to managers, but also provide aggregate information on the change in efficiency and productivity over time) has given further impetus for the ABS to undertake this style of research. The ABS is in a unique position to do such analyses, given its access to unit record files like the private hospitals dataset.

Section 2 describes the techniques used in the analyses. It introduces the ideas behind the techniques, the kinds of analyses that each technique allows, and the mathematical formulations used to find solutions in each case. The ideas presented are standard formulations and this section may be skipped or skimmed by readers familiar with the techniques. Parametric techniques are used to estimate both cost and production functions for a range of functional forms, and the reasons why one estimation technique may be preferable to the other are discussed.

Section 3 introduces some issues concerned with the measurement of variables representing input quantities and prices, output quantities and total operating costs. Using data from the ABS Private Health Establishments Collection, the section describes the construction of data used in the analyses and well as any problems, both practical and conceptual, with the definitions employed.

Section 4 discusses the results of applying the DEA technique and index number measurement. Section 5 presents the results obtained by applying the SFA and OLS techniques to production and cost function estimation. The analyses have been applied to both a cross-section of the private hospitals data (for 1994–95) and a panel (for 1991–92 to 1994–95).[1] Section 6 compares selected results obtained from the non-parametric and parametric techniques. Section 7 concludes with suggestions for further work. The paper also contains two appendixes which present summary statistics for the data on which all analyses are based and describe an alternative technique for measuring capital inputs.

---

[1] One of the reasons estimation was undertaken on both the panel and cross-section was the high probability that data quality in this dataset has changed over time. Data quality has improved as respondents have become more accustomed to the survey and because of a general increase in the awareness of hospital statistics in the health sector. Therefore we were interested to see if panel results coincided with those obtained from a single cross-section 1994–95.

# 2 The techniques

## 2.1 Data Envelopment Analysis

The use of data envelopment analysis in the study of hospital efficiency, both public and private, is relatively common (for example, Banker, Conrad & Strauss 1986, Grosskopf & Valdmanis 1987, Register & Bruning 1987, Fare, Grosskopf & Valdmanis 1989 and Valdmanis 1992). Most authors cite the inherent flexibility of the DEA model as a major attraction for its use in such studies. Another reason for the use of the DEA technique arises when there is lack of realistic price data associated with hospital inputs and outputs. The DEA technique is able to handle multiple outputs of production, reducing the need for price data to form the types of composite measures of output (and even input) required for regression-based techniques. However, if one wishes to measure allocative efficiency price data is required.

While there is general agreement about the applicability of DEA to evaluate hospital efficiency, a number of features of the model may worry many researchers in the field. Two important problem areas of the model are: the assumption that there is no 'noise' (or error) in the data being studied; and the lack of a definite functional form encapsulating the production technology. The latter, whilst a strong argument for the technique in many studies, raises the problem of what method should be used to evaluate the results of a DEA study, mainly due to the inability to perform the usual diagnostic tests associated with regression estimation.

Valdmanis (1992) (based on Nunamaker 1985) suggests, as a possible answer to these problems, that a DEA researcher run a number of different models from each dataset and evaluate the sensitivity of the results to changes in model specification. These changes may take the form of alternative input and output definitions, or even different populations within each dataset. The purpose of this sensitivity analysis is to assess whether the ranking and efficiency of an individual firm is variable-specific (or model-specific) or whether the results are robust to changes in dataset specification. Valdmanis (1992) cautions that '... for a model to be considered robust, it must be shown that minor changes in the list of variables cannot alter fundamentally the conclusions of the DEA model.' Section 7 of the paper discusses possibilities for extensions and further work in this area, one of the most important of which are implementing methods to identify influention outliers (or extreme data points) in DEA.

Another method of evaluation is to compare the results of a DEA study with results from other efficiency evaluation methods applied to datasets comprising similar observations and variables. Some alternative methods include SFA and production function or cost function OLS estimation.

In this paper, it is the intention to assess practically the DEA technique by comparing results obtained using a selected input-output specification with results from different dataset specifications and alternative estimation methods. Only after this analysis is done can meaningful conclusions be drawn from the results of a DEA study. A lack of robustness to changing datasets implies that any results from this type of analysis must be analysed in reference to the data used in the study. It may also be possible to explain difference in results and conclusions between two DEA studies in terms of differences in data construction and definition.

Another aim of the paper is to apply and assess the Malmqvist index approach to analysing technological changes over time from a panel dataset. The method, introduced by Caves, Christensen and Diewert (1982) and refined by Fare et al. (1994) into a linear programming technique, calculates a productivity index based on Malmqvist (1953) indexes and Farrell (1957) distance functions using data from adjacent time periods. This method can then be used to decompose traditionally defined productivity into efficiency change and technological change.

The measurement of micro level efficiency involves a comparison between the observed and optimal usage of inputs to produce an amount of output for each observation in a sample. Optimal input or output values are determined by the potential production possibilities, that is, the best observed practice in the sample. In this context, efficiency and productivity are defined using the values and ratios of 'useful' inputs to outputs.

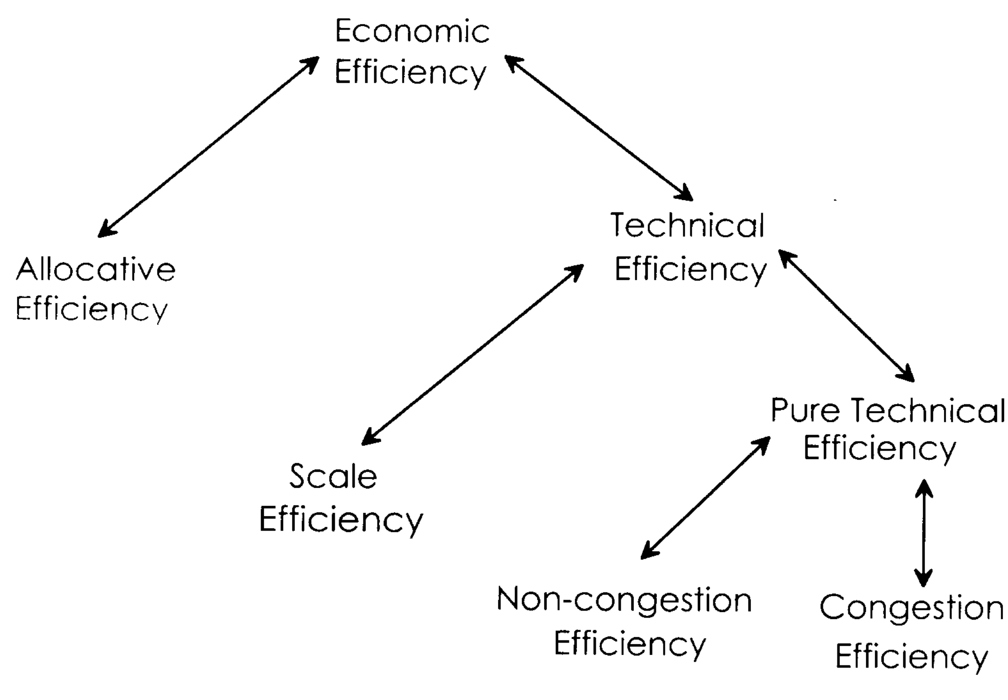
Before outlining the different versions of the DEA technique, it is useful to look at a number of 'types' of efficiency and the way in which they relate to each other. Consider the concept of **economic** efficiency,[2] which is composed of **technical** and **allocative** efficiency. Nunamaker (1985) defines technical efficiency as a measure of the ability of a micro level unit (referred to as a **firm, observation** or **decision making unit (DMU)**) to avoid waste by producing as much output as input usage will allow, or using as little input as output level will allow. Allocative efficiency measures the ability of a DMU to avoid waste by producing a level of output at the minimal possible cost.

Another decomposition occurs at the level of technical efficiency, which can be considered to be composed of **scale** and non-scale effects, the latter being referred to as **pure technical** efficiency. Scale efficiency is the measure of the ability to avoid waste by operating at, or near, to the most productive scale.

Lastly, pure technical efficiency can be considered to be composed of **congestion** efficiency and other effects. Input congestion efficiency is the measure of the component of pure technical efficiency due to the existence of negative marginal returns to input, and the inability of a firm to dispose of unwanted inputs costlessly. The inability to costlessly dispose of unwanted inputs is referred to as **weak disposability of inputs** in the discussion that follows.

The following diagram sets out the progression of efficiency measures outlined above. In the next section, these concepts are defined in terms of the DEA linear programming technique.

**Figure 2.1: A 'roadmap' of efficiency decomposition**




[2] The concepts discussed in this paper relate to the measurement of static efficiency, and are not designed to assess the dynamic component of efficiency. Economic efficiency also contains the dimension of dynamic efficiency, that is, the success of economic agents in adapting their activities to latent or emerging opportunities in production technology and actual and potential changes in consumer preferences over time.

DEA is a non-parametric mathematical programming approach to production or cost frontier estimation. The piecewise-linear convex hull approach to frontier estimation was originally proposed by Farrell (1957), but failed to gain popularity until reformulated into a mathematical programming problem in a paper by Charnes, Cooper and Rhodes (1978), which has become known as the DEA approach (Seifort &Thrall 1990).

The original Charnes, Cooper and Rhodes (1978) paper considered an input-oriented, constant returns to scale (CRS) specification, with additional modifications to the methodology including a variable returns to scale (VRS) model (Banker, Charnes & Cooper 1984) and an output-oriented model.

The Charnes, Cooper and Rhodes (1978) paper reformulated Farrell's original ideas into a mathematical programming problem, allowing the calculation of an efficiency 'score' for each observation in the sample. This score is defined as the *percentage reduction in the use of all inputs* that can be achieved to make an observation comparable with the best, similar observation(s) in the sample *with no reduction in the amount of output*.

Equation (1) below sets out the linear programming problem corresponding to the basic DEA specification of Charnes, Cooper and Rhodes (1978). This linear program (LP) is in fact the dual, *envelopment form* of an efficiency maximisation LP for each observation. The objective function seeks to minimise the efficiency score, θ, which represents the amount of radial reduction in the use of each input. The constraints on this minimisation apply to the comparable use of outputs and inputs. Firstly, the output constraint implies that the production of the $r$th output by observation $i$ cannot exceed any linear combination of output $r$ by all firms in the sample. The second constraint involves the use of input $s$ by observation $i$, and implies that the radially reduced use of input $s$ by firm $i$ $(\theta x_{is})$ cannot be less than the same linear combination of the use of input $s$ by all firms in the sample. In other words, to reduce the use of all inputs by observation $i$ to the point where input usage lies on the 'frontier' defined by the linear combination of input and output usage by the 'best' firms in the sample.

Considering a dataset containing $K$ inputs, $M$ outputs and $N$ firms, where the sets of inputs and outputs for the $i$th observation are $x_{ik}$, $k=1,\dots,K$ and $y_{im}$, $m=1,\dots,M$, the input-oriented CRS DEA LP for observation $i$ has the form:

$$\min_{\theta,\lambda} \ \theta = \theta_{CRS}$$

$$\text{such that: } -y_{ir} + \sum_{j=1}^{N} \lambda_j y_{jr} \geq 0, \ r = 1, \dots, M$$

$$\theta x_{is} - \sum_{j=1}^{N} \lambda_j x_{js} \geq 0, \ s = 1, \dots, K$$

$$\lambda_j \geq 0, j = 1, \dots, N \qquad (1)$$

where θ is a scalar and λ is an $N$ x 1 vector of constants. The value of θ obtained from the LP is the *efficiency score* for the $i$th observation, and will lie in the region (0,1]. An efficiency score of 1 indicates a point on the frontier and hence a technically efficient observation relative to the dataset.

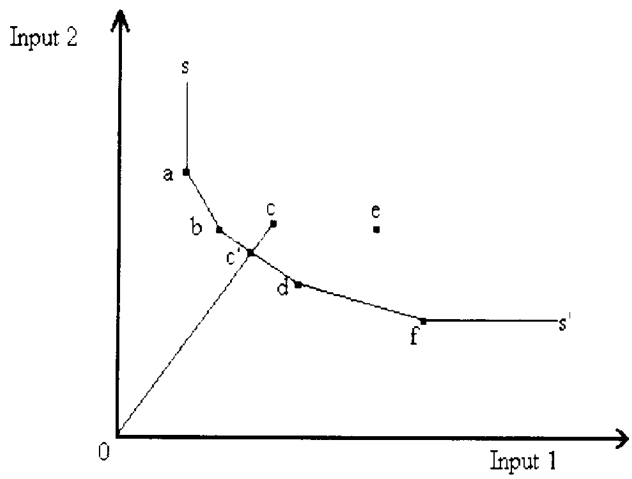
Equation (1) must be solved $N$ times, once for each observation in the sample. The efficiency scores from the set of LPs (1) indicate, given a level of output, by how much inputs can be decreased for an inefficient observation to be comparable with similar, but more efficient, members of the sample. This efficiency is often referred to as technical efficiency.

As an illustration of the technique, consider an example of six firms using two inputs (input 1 and input 2) to produce one unit of output, shown in figure 2.2. The linear programming solution produces the non-parametric piece-wise linear frontier (ss'). Firms which lie on this frontier are fully efficient (firms a, b, d, f). Firms which lie above and to the right of the frontier are inefficient (firms c and e).

The measure of the technical inefficiency of firm c ($\theta$ from equation 1) is captured by the ratio 0c'/0c. Note that point c' in figure 2.2 does not represent a firm, but the point on the frontier that firm c would occupy if it could be made fully efficient by radially reducing its use of both inputs. That is, firm c could reduce the amount of input 1 and input 2 it uses in production and still produce the same amount of output.

The input-oriented DEA technique calculates efficiency scores by the amount of radial reduction in inputs that can be achieved to move the firm towards the best practice frontier. By using the radial reduction technique (moving each inefficient firm towards the frontier by contracting towards the origin each input by the same proportion), the technique becomes invariant to the units used to measure each input.

**Figure 2.2: Illustration of the DEA technique**




Equation 1 represents the case in which the assumption of constant returns to scale is imposed on every observation in the sample. In this formulation no account is taken of factors which may make firms unique beyond the simple input-output mix, such as inefficiencies which result from operating in areas of increasing or decreasing returns to scale due to size constraints. Another assumption embodied in the LP in equation (1) is that of strong disposability of inputs. This represents the assumption that, when reducing input usage, an observation is able to dispose of the unwanted inputs costlessly. In effect, this assumption rules out the possibility of decreasing marginal products for inputs.

To further decompose the efficiency scores from equation (1) it is necessary to use a number of additional DEA formulations which relax some or all of the assumptions embodied in the basic DEA equation.

The first variation relaxes the CRS assumption by considering scale and allowing firms to exhibit both increasing and decreasing returns to scale in addition to constant returns. Known as the VRS formulation, this involves the addition of a constraint to the basic CRS formulation specifying that the sum of the linear combination parameters be equal to one ($\sum_j \lambda_j = 1$).

In practice, this most often results in a 'tighter' fitting frontier with more firms on and near to the frontier (efficiency scores closer to one).[3]

The efficiency scores from models estimating CRS and VRS, can be used to calculate **scale efficiency** for each observation ($\theta_{SC}$) using the following relationship between CRS (technical efficiency) and VRS (pure technical efficiency) efficiency scores:

$$\theta_{CRS} = \theta_{VRS}.\theta_{SC} \tag{2}$$

While it is possible to use these results to decompose technical efficiency into scale and other effects, the results offer no information about whether a observation with scale inefficiencies is operating under increasing or decreasing returns to scale.

Information about returns to scale can be obtained from a variant of the VRS formulation called the **non-increasing returns to scale formulation** (NIRS). This form of the DEA LP involves the modification of the VRS constraint from a strict equality governing the sum of the linear combination parameters to one of being less than or equal to one. Comparing variable and non-increasing returns efficiency scores allows a judgement of the nature of returns to scale for each observation in the sample.

The input-riented DEA LPs discussed thus far look at numerical combinations of inputs to yield a given amount of output. It is possible to invert the problem and look at numerical combinations of outputs which can be yielded from a given amount of input. This formulation is known as the **output-oriented approach**. Again, this particular formulation is the linear programming dual to an efficiency maximisation problem, analogous to the previous discussion for the input-oriented formulation. The scores indicate, given a set of inputs, by how much a observation can increase each output to be comparable with the 'nearest, compatible' member(s) of the sample, with no increase in the use of inputs. Analogous with the input-oriented formulation, outputs of inefficient DMUs are radially increased towards the frontier making the formulation invariant to the units used to measure each output. Alternative forms for VRS and NIRS output-oriented DEA formulations can also be solved (by including additional restraints on the weights) in the same manner as those discussed previously for the input-oriented formulation.

In all of the DEA formulations discussed, it has been assumed that a firm is able to reduce its use of inputs with no additional costs associated with input disposal. This assumption is called *strong input disposability*. By formulating a DEA LP which relaxes this assumption, it is possible to decompose efficiency scores into technical and **congestion efficiency** effects. This DEA formulation allows the frontier to 'bend back' on itself (that is, have a positive slope in the input-input plane), simulating the effect of a negative marginal product for a particular input.

To decompose an efficiency score into technical and congestion efficiency, frontiers representing strong disposability and weak disposability are estimated.

[3] Whilst this method is able to account for firm size in a technical efficiency rating, there may be a number of additional factors which distort the inter-firm comparisons necessary for the construction of the frontier. Using the efficiency scores from the DEA LP (1), a number of decomposition techniques are available which can be used to adjust efficiency for **uncontrollable or environmental factors**. Methods which adjust the efficiency scores obtained from a DEA LP are known as two-stage methods.

One such method, introduced by McCarty and Yaisawarng (1993), uses the truncated, or Tobit, regression method to control for factors not considered in the DEA model (see Maddala 1983) for an introduction to the Tobit regression technique).

An alternative (one-stage) approach to this problem is the **non-discretionary variable** specification. In this formulation the assumption that a firm is able to costlessly alter the usage of all inputs is tightened by considering a subset of inputs which are considered fixed.

The discretionary variable formulation can be used to analyse the effect on overall efficiency scores of the assumption that the use of certain inputs cannot be altered by the observation manager within a specified time frame (as would be the case for fixed capital assets).

Congestion efficiency is equal to the ratio of pure technical efficiency under strong disposability to pure technical efficiency under weak disposability, or,

$$\theta^{SD}_{VRS} = \theta^{WD}_{VRS}.\theta_{CON} \tag{3}$$

Technical efficiency can also be decomposed into congestion efficiency, scale efficiency and 'pure' technical efficiency for each observation by running three DEA models: strong disposability CRS , strong disposability VRS and weak disposability VRS. Given the definition of scale efficiency (equation 2) and congestion efficiency (equation 3), technical efficiency is related to congestion and scale efficiency in the following way:

$$\theta^{SD}_{CRS} = \theta^{WD}_{VRS}.\theta_{SC}.\theta_{CON} \tag{4}$$

The congestion efficiency model is often used in efficiency analysis, however, it should be noted that the model assumes that all of the inefficiency due to congestion is outside of the firm's control (such as labour unions controlling staff numbers, government regulation or in instances where it would be costly to reduce the use of inputs that are not needed to meet current demand). To the extent that this is not the case, the model may 'inflate' the efficiency scores for a firm by incorrectly assigning inefficiency between scale and congestion effects.

Efficiency again lies in the region (0,1], with a score of one indicating an observation with no technical or congestion inefficiencies.

To analyse the **movements in firm and overall efficiency over time** using a panel of firms, it is necessary to adapt the methods mentioned previously to allow for inter-temporal comparisons (such as comparing the input-output mix for a particular time period with the production technology implied by input and output usage for an adjacent time period). The following outlines this process, using the Farrell (1957) definition of micro level efficiency and the Malmqvist index approach to efficiency measurement of Fare et al. (1994).

The input distance function for firm $i$ with respect to two time periods, $t$ and $s$, is defined using equation (5), where $S^t = \{(x^t, y^t) : x^t \Rightarrow y^t\}$ is the production technology that governs the transformation of inputs to outputs for period $t$:

$$d^t_i(x^s, y^s) = \min_{\theta>0} \{\theta : (y^s, \theta x^s) \in S^t\} \tag{5}$$

The distance function in equation (5) measures the minimum proportional change in input usage at period $s$ required to make the period $s$ input-output set, $(x^s, y^s)$, feasible in relation to the technology $S^t$ at period $t$ (see Fare et al. 1994). The Malmqvist input productivity index comparing periods $t$ and $t+1$ can then be defined using distance functions representing the four combinations of adjacent time periods,[4]

$$m_i(y^{t+1}, x^{t+1}, y^t, x^t) = \sqrt{\frac{d^t_i(x^{t+1}, y^{t+1})}{d^t_i(x^t, y^t)}.\frac{d^{t+1}_i(x^{t+1}, y^{t+1})}{d^{t+1}_i(x^t, y^t)}} \tag{6}$$

Following Fare et al. (1994) an equivalent way of writing equation (6) is

$$m_i(y^{t+1}, x^{t+1}, y^t, x^t) = \frac{d^{t+1}_i(x^{t+1}, y^{t+1})}{d^t_i(x^t, y^t)}.\sqrt{\frac{d^t_i(x^{t+1}, y^{t+1})}{d^{t+1}_i(x^{t+1}, y^{t+1})}.\frac{d^t_i(x^t, y^t)}{d^{t+1}_i(x^t, y^t)}} \tag{7}$$

where the ratio outside the brackets measures the change in relative efficiency[5] between periods $t$ and $t+1$ and the geometric mean of the ratios in the brackets measures the shift in technology between the two periods.

---

[4] In this form, the index is the geometric mean of Malmqvist indices with time periods $t$ and $t+1$, **respectively, as the** reference technology. This form is typical of Fisher ideal indices (Fare et al. 1994).

[5] That is, the change in how far observed production is from potential production.

Using the link between the Farrell distance function and DEA efficiency scores (Charnes, Cooper & Rhodes 1978) it is possible to calculate the required distance function in equation (6) using DEA LPs of the following form (assuming input orientation and CRS):

$$[d_i^s(\mathbf{x}^r, \mathbf{y}^r)]^{-1} = \min_{\theta,\lambda} \theta$$

$$\text{such that: } -y_i^s + \sum_{j=1}^{N} \lambda_j y_{jk}^r \geq 0,\ k = 1, \ldots, M$$

$$\theta x_i^s - \sum_{j=1}^{N} \lambda_j x_{jl}^r \geq 0,\ l = 1, \ldots, K$$

$$\lambda_j \geq 0, j = 1, \ldots, N. \tag{8}$$

where $r$ and $s$ represent the possible combinations of time periods $t$ and $t+1$. Note that two of these LPs involve the use of data from both of the time periods being compared.

Using this technique, we can calculate an overall Malmqvist index and its decomposition for each observation for each pair of time periods being compared. To obtain an estimate of technical progress over time, a time specific Malmqvist index for each period is calculated as the geometric mean of indices for each observation in a period.

The index calculated in the previous analysis, whilst giving some indication of movements in efficiency over the time period, is not a transitive index. One method which can be used to reinterpret the results is to transform the calculated Malmqvist indices into a **time transitive form.**[6] One such method, due to Balk and Althin (1996), transforms the previously defined Fare et al. (1994) Malmqvist index using a method analogous to the Elteto, Koves and Szulc transitivity transformation technique.

Using a transitive index allows a better understanding and comparison of the movements in the index between each of the periods, as well as the contribution that the movement of the index in each period plays in the index covering all periods.

To this point, the measurement of micro level technical efficiency and productivity has been considered (that is, how well an individual unit avoids waste by producing as much output as input usage allows). As depicted in figure 2.1, technical efficiency is only one part of economic efficiency. Another contributing factor to economic efficiency is **allocative efficiency**. This measures the ability of an observation to avoid waste by producing outputs at their marginal cost minimising quantum.

The DEA method can be used to analyse allocative efficiency (see figure 2.1) by defining a set of input prices to match the input quantities used in the model. Economic efficiency (EE) for the $i$th observation is calculated as the ratio of minimum cost to actual cost

$$EE_i = \sum_{k=1}^{K} w_{ik} x_{ik}^* / \sum_{k=1}^{K} w_{ik} x_{ik} \tag{9}$$

where $w_{ik}$ is the price of input $k$ for the $i$th observation and $x_{ik}^*$ the cost minimising input level obtained from solving a cost minimising LP for each observation. Allocative efficiency is calculated residually as the ratio of economic efficiency to technical efficiency, where technical efficiency scores are obtained by solving an LP, such as equation (1). The results of this analysis will be very dependent on the input price definitions adopted. Since an important reason for using the DEA method is the unavailability of clearly defined price information, the use of this method and the results obtained from it must be considered in the light of the prices defined for the exercise.

[6] An index is transitive if the direct and chained versions of the index are equal for the comparison of any two time periods.

## 2.2 Parametric techniques

Recently a number of studies have applied SFA to hospital datasets, usually to measure relative efficiencies, see Newhouse (1994), Vitaliano and Toren (1994) and Zuckerman, Hadley and Iezzoni (1994). In all these studies cost functions were estimated rather than production functions. Cost functions are often estimated to control for the biases that arise in the direct estimation of production functions,[7] so as to represent a multi-product firm, or because analysts are interested in measuring allocative efficiency as well as technical efficiency or a combination of these two efficiencies, cost efficiency. This paper estimates both production and cost functions for a number of reasons: to facilitate comparisons between other techniques; because of the likely fallibility of the price data used in this study; and because of uncertainty in determining what environment private acute hospitals operate in and how their economic behaviour is affected.[8]

Stochastic frontier modelling is becoming increasingly popular primarily because of its flexibility and its ability to closely marry economic concepts with modelling reality. These techniques are also now more easily applied given improvements in computing technology and the availability of unit record datasets. Stochastic frontier modelling is often used to compare firms' relative efficiencies though it can also be used to derive estimates of productivity change over time.[9] The technique has a number of benefits when compared to standard econometric estimation (OLS) of production functions.[10] It estimates a 'true' production frontier rather than an average frontier, thus it fully represents the maximal properties of the production function. One important implication of estimating the frontier is that measured productivity change will represent pure technological change rather than a combination of efficiency change and technological change which is the case when using non-frontier techniques. However, OLS estimation of functions is still very useful when testing for standard statistical aspects of the analysis, for example, heteroskedasticity and the normality of the residuals.

SFA has some advantages over non-parametric techniques, such as DEA, for estimating frontiers, efficiency and productivity; in particular, it is able to account for measurement error. However, stochastic frontier modelling does have some constraints which DEA does not including: only one output can be accommodated when modelling production functions, and the need to select functional forms for both the production structure and error components. Thus, parametric techniques for measuring efficiency and productivity provide an alternative approach for dealing with errors, at the cost of using a more restrictive model specification than DEA. Parametric techniques such as SFA are also more likely to be appropriate if the focus is on drawing conclusions about the aggregate properties of the dataset, rather than the performance of individual units. By contrast, DEA may be more appropriate if the focus is on developing a detailed understanding of the performance of individual units within the sector or identifying DEA peer relationships among the production units.

---

[7] Under the assumption of profit maximisation or cost minimisation parameter estimates will be biased and inconsistent when the production function is directly estimated (Thomas 1985, p. 224). Direct estimation is only valid when input levels are assumed fixed or expected profit is maximised, that is, there is uncertainty about output prices or quantities. See also Coelli (1995, p. 226).

[8] This paper assumes that private hospitals behave in a manner that can be predicted by neoclassical economics, however, it is quite possible that this model of behaviour is inadequate, particularly, when explaining how private non-profit hospitals operate.

[9] See Lovell (1996) for an excellent review of the ability of SFA and DEA to measure efficiency and productivity change.

[10] This discussion will be mostly couched in terms of production function though much of the discussion also applies to cost functions.

### Production functions

A stochastic frontier production function was first proposed by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977).[11] The model took the form

$$\ln(y_i) = F(x_i : \beta) + v_i - u_i, \quad i = 1,...N \qquad (10)$$

where $y_i$ is the output of the $i$th firm, $x_i$ is a vector of inputs for the $i$th firm, $\beta$ is a vector of parameters to be estimated, $v_i$ a symmetric error term, and $u_i$ a non-negative error term representing inefficiency.

Once an assumption is made about the distribution of the error terms the parameters of this model can be estimated using maximum likelihood estimation (MLE) or a corrected form of ordinary least squares (COLS).[12] The other assumption which is required before estimation is what functional form $F(.)$ is to be estimated.

In this exercise both Cobb-Douglas and Translog production functions were estimated. Linearised versions of the Cobb-Douglas and Translog are presented below, where $x_{ij}$ represents the $j$th input for the $i$th firm (j=1...K).

Cobb-Douglas $\ln(y_i) = \ln(A) + \sum_j \alpha_j \ln(x_{ij}) + v_i - u_i$

Translog $\ln(y_i) = \ln(K) + \sum_j \alpha_j \ln(x_{ij}) + .5[\sum_j \beta_j (\ln(x_{ij}))^2 + \sum_j \sum_{k \neq j} (\gamma_{jk} \ln(x_{ij}) \ln(x_{ik}))] + v_i - u_i$

Cobb-Douglas and Translog versions of equation (11) were estimated on a cross-section of data using the specialist SFA software FRONTIER package and the econometric package LIMDEP.[13]

Estimation of Cobb-Douglas and Translog production functions were also undertaken on the panel of data using the panel data analogue of equation (11), based on Battese and Coelli (1992).

$$\ln(y_{it}) = F(x_{it} : \beta) + v_{it} - u_{it} \qquad i = 1,...N, t = 1, ...T \qquad (11)$$

This panel model can incorporate a number of extensions to the original cross-sectional model, for example, inefficiencies as represented by $u_{it}$ can be fixed or varying over time (when inefficiencies are time invariant $u_{it}$ becomes $u_i$). Where inefficiencies are assumed to vary across time a model has to be estimated to explain this variation. Battese and Coelli (1992) propose the following model for $u_{it}$ (as estimated in the FRONTIER package)[14]

$$u_{it} = \{\exp[\eta(t - T)]\} u_i \qquad (12)$$

where $\eta$ is the parameter to be estimated, $T$ represents the number of time periods over which the equation is estimated and $u_i$ are random error drawn from an assumed distribution (usual half normal or a more general truncated normal distributions are used for this purpose).

In this paper, we estimate equations of the form of (10) and (11), assuming a normal symmetric error assumed an a half normal or truncated normal inefficiency error. The inefficiency error was usually assumed to be unchanging or fixed due to the short panel, however some time-varying models were estimated using the FRONTIER package. In addition to the stochastic frontier estimation, OLS counterparts to equations (10) and (11) were also estimated using only a symmetric error term.

---

[11] As cited in Coelli (1995, p. 224).

[12] Greene (1993, p. 69).

[13] The results presented in this paper were obtained from the LIMDEP software, and these results were always consistent with the those obtained with the Coelli FRONTIER package. The FRONTIER package uses a three-step estimation method to obtain the final maximum likelihood estimates; in the first step OLS estimates are obtained, secondly a two-phase grid search of a parameter representing the ratio of inefficiency variance to the composite inefficiency and error variance whilst setting other parameters equal to their COLS counterparts, thirdly these values are used as starting values in an 'iterative process using the Davidson-Fletcher-Powell Quasi-Newton method to obtain maximum likelihood estimates', see Coelli (1995, p. 11) for a full description.

[14] Time-varying efficiencies were not estimated using the LIMDEP package.

As an alternative to equation (11) firm specific effects can be used to explain differences in inefficiency between firms.

**Cost functions**

Both cost and profit functions can be estimated using SFA. These functions are most useful when the following circumstances arise: profit maximisation or cost minimisation is suspected (in which case direct estimation of the production function will produce biased and inconsistent estimates); firms have multiple outputs; or there is an interest in predicting allocative efficiencies.

Cost functions of the following form were estimated using the FRONTIER and LIMDEP packages, on both the cross-section and panel data.[15]

$$\ln(c_{it}) = C(y_{it}, w_{it} : \alpha) + v_{it} + u_{it} \qquad i = 1,\ldots,N \; t = 1,\ldots,T \qquad (13)$$

where $c_{it}$ is the costs of the $i$th firm, $y_{it}$ a vector of outputs, $w_{it}$ a vector of input prices, and $\alpha$ a vector of unknown parameters to be estimated. Note that in this case, because inefficiencies are assumed to always increase costs, both of the error terms are preceded by positive signs.

Cost functions can be estimated as a single equation (for example, equation (13)) or in a systems equation setting where factor demands are estimated.[16] When a systems approach is used allocative efficiency can be measured directly through the factor demand equations and more efficient parameter estimates are obtained (Coelli 1995). However, both systems and single equation forms of the cost function present difficulties in obtaining allocative and technical efficiencies when functional forms other than self-dual forms such as the Cobb-Douglas are estimated. For SFA, systems estimation was not undertaken primarily because the FRONTIER and LIMDEP packages did not allow it and because the authors were not clear how to represent frontier analysis in a system equation setting.

The inefficiency scores obtained directly from a cost function represent composite cost inefficiency (within which both allocative and technical inefficiency can be contained) and therefore need to be decomposed to be directly compared to the efficiency scores derived for other methodologies, for example, from DEA.

Kopp and Diewert(1982) have developed a method with which technical and allocative efficiencies can be derived from the estimated cost efficiencies of a deterministic cost frontier. Greene (1993) has suggested that this technique could be used to decompose the inefficiency error of stochastic frontier (composite error models).[17]

We attempted to use the analytical method of decomposing cost inefficiency of the Cobb-Douglas stochastic cost functions, the only addition to the technique presented by Kopp and Diewert (1982) was to add back into frontier points the symmetric error term as suggested by Greene (1993). However, the authors are unclear about whether this process is as straight forward as adding back in the symmetric error term as the results of this exercise were not sensible.[18]

---

[15] When T = 1 the cross-sectional model applies.

[16] The estimated cost functions in this analysis are drawn from production economic theory. Cost functions have often been estimated in the health sector based on a more ad hoc representation of costs and behaviour, see Breyer (1987) for an interesting discussion of these issues.

[17] Kopp and Diewert (1982) tested their technique for decomposing technical and allocative efficiency. They found the simplest method to do this was to take deviations in the system of equations and minimise the squared sum of these differences using variant of the Davidson-Fletcher-Powell Algorithm. Their results coincided exactly with those solved for analytically.

[18] See Fare and Primont (1996) for a discussion of duality theory and SFA.

# 3 Data

Data for the analysis were obtained from the ABS Private Health Establishments Collection (PHEC), an annual census of private hospitals covering all private acute care and psychiatric hospitals operating in Australia. We chose to exclude private psychiatric hospitals due to their substantially different characteristics and mode of operation. Various characteristics of private acute hospitals can be identified in the dataset including: whether a hospital is characterised as (1) For-profit (FP), (2) Not-for-profit, religious or charitable (NFPr) or (3) Not-for-profit, other (NFPo); size proxied by the number of beds. In the case of the cross-section, data for the year 1994–95 was used and in the case of the panel, the years 1991–92 through to 1994–95. For the 1994–95 dataset (cross-section), the population of acute care private hospitals comprises 301 observations, composed of 155 FP, 70 NFPr and 76 NFPo. A balanced panel for the years 1991–92 through to 1994–95 contains 280 hospitals, an unbalanced panel contains 314 hospitals. Balanced panels were derived from the dataset to facilitate comparisons between techniques for measuring productivity in particular DEA. When estimating cost functions only 255 hospitals are used due to missing data or zeros problems.[19]

In order to check whether the population of hospitals is suitably homogeneous, some analyses were based on subsets of the cross-section and panel data as defined by hospital type or size.

## 3.1 Input variables

Input variables were constructed in order to represent labour, capital and intermediate inputs. The degree of disaggregation within these categories depended on the homogeneity of an input category, the quality of data within which to measure this input and whether most hospitals used this input. Table 3.1 details the definition of various input variables.

Labour inputs were measured by total full-time equivalent (FTE) staff. This measure included salaried medical officers (SMOs) but did not include visiting medical officers (VMOs) as data on the hours worked or days worked by VMOs were not available. A labour measure to capture the input of VMOs was based on wages paid to all staff plus the cost (contract value) of VMO services. Models with both measures of labour inputs were trialled in production function estimation. Although it was possible to disaggregate the FTE staff measure, for example, into nursing and non-nursing personnel, this was often not done when estimating parametric models due to the problem of zero values when logarithms of values are required.

Neither of these measures of labour input is ideal, the FTE measure because of the high level of aggregation of employment groups, and dollar-based measures because movements represent changes in not only volumes but also in prices (wages). In the DEA study, models including lower levels of aggregation in measuring labour inputs were able to be developed, as zero inputs in some hospitals did not have the same impact on DEA analysis. For this analysis, variables representing the inputs of SMOs, nursing staff and other staff are used in a number of models.

Two measures of the capital input were available, a measure based on the number of beds per hospital and a derived measure of capital stock. Beds are often used to proxy for capital stock in hospital studies usually because a reliable measure of the value of assets is not available. However, the PHEC dataset contains data on depreciation and gross capital expenditure, and an alternative measure of capital stock was derived through inverting the Perpetual Inventory

[19] A problem we encountered in estimating both the Translog and the Cobb-Douglas production and cost functions was the existence of zeros for some inputs for some hospitals therefore not enabling us to log this value. This problem was largely solved by moving to a higher degree of aggregation in inputs which resulted in all hospitals having some positive value. Battese (1996) uses a dummy variable technique to overcome the problems of some firms not using certain inputs, however, we also have the problem that some firms do not produce certain outputs and it is not clear if the same underlying rationale would apply. We trialled Battese's technique in our production function estimation.

Model and by incorporating a number of assumptions such as the investment history of firms.[20] These estimates were trialled in production function analysis; however, they were not persisted with due to the substantial inconsistencies in the data at the unit record level. Interestingly an aggregate (industry wide) measure of capital stock derived from this technique proved quite robust to changes in underlying assumptions.

Intermediate inputs were measured as the total dollar value of all non-labour, non-capital expenditure. This covered expenditures on drugs and medical supplies, food, repairs and maintenance, and patient transport services. When volume-based measures of labour inputs were used the contract costs of VMOs were also included in the measure of intermediate inputs and raw materials. This was not entirely unsatisfactory as VMO services were purchased similarly to other intermediate services and unlike salaried labour. However, the characteristics of VMO services indicate that they are more like a labour service rather than a capital or non-labour service.

For the DEA study, a fourth input class representing the total patient input (output) was included for some of the DEA study, to reflect the argument stated by Valdmanis (1992), '... just as raw materials such as iron are inputs to a steel mill, so "sick" people are inputs to the hospital production process'. The actual variable used was the total number of inpatient separations. This approach to explaining hospitals production processes tends to view the output of hospitals as health outcomes rather than a process type output. The output measures in this analysis are focused on the process type or production volume style estimates of output and we are only interested in outcomes in the sense that outcomes reflect the quality of outputs. In any case, the dataset does not contain measures which would enable us to measure the improvement in the health of sick persons over the course of their hospital stay. Therefore admissions were not used as an input measure in the parametric analysis.

## 3.2 Output variables

This study examined a number of measures of private acute care hospitals output. A frequently used measure of hospital output is case-weighted separations where the case weights reflect the severity of the different cases treated by the hospital. This type of output measure was constructed using disease-costing weights provided by the Australian Institute of Health and Welfare. However, it was found that the level of aggregation in grouping diseases in the PHEC dataset was too high to effectively reweight the raw separations. This led to movements in weighted separations being almost the same as movements in unweighted separations. As a measure of output, unweighted separations tend to favour hospitals which treat simpler diseases and provide quicker treatments.[21]

To avoid this problem, a composite estimate of output was constructed based on occupied bed days. Although this measure of output is less than ideal, it at least incorporates some element of adjustment for disease severity. In the DEA analysis, where it was possible to use multiple outputs, separations were used since different types of outputs are able to be accounted for in this case. The occupied bed day measure of output is focused on inpatient care provided by hospitals. In order to capture non-inpatient care and reflect this in a single measure of output, non-inpatients were weighted according to their relative costs to form an overall measure of occupied bed days.

[20] Details of the construction of the alternative capital measure are in Appendix 2.

[21] The definition of separations was amended in the 1995–96 PHEC. Previously, a patient separation was recorded only when a patient left hospital, with the total hospital stay being attributed to that separation. The new method adopts the casemix concept of 'episode of care' with a separation being recorded if there is a change in the clinical treatment a patient receives while in hospital. Consequently, two (or more) separations may now be recorded for a single patient's hospital stay, while prior to 1995–96 only one separation would have been recorded.

These weights were based on the Health and Allied Services Advisory Committee (HASAC) formula where one inpatient treatment represents on average 1/5.753 of an occupied bed day.

An additional complication that arose when measuring the non-inpatient component of outputs was that accident and emergency cases were measured on a visits basis rather than an occasion of service or treatments basis. In order to convert these visits to treatments a conversion factor of 3:1 was applied, based on the HASAC formula.[22]

When undertaking DEA analysis, occupied bed days were further disaggregated into different types, for example, surgical bed days, and include other output activity measured separately such as non-inpatient activities; however, this was not feasible in the parametric analysis, again due to the problem of logarithms of zero values.

Deflated patient revenue was also used to measure hospital output. This was done for two reasons; firstly as an alternative to the occupied bed day measure of outputs; and secondly because deflated revenues are often used in National Accounting exercises to measure output and it would be interesting to see how this measure of output compared with direct volume measures using occupied bed days. In the case of analysis conducted on the cross-section there was no need to deflate revenue, however it was still assumed that prices were the same across hospitals. The deflated patient revenue measure of output is less appropriate for NFP hospitals than FP hospitals.

Despite having alternative measures of output neither measure is ideal and it is certainly the case that some aspects of the outputs of hospitals have failed to be accounted for, for example, research and development.[23] Another aspect which has not been adequately accounted for in measuring outputs has been quality dimensions of outputs. The occupied bed day measure does not adjust at all for changes in quality between firms or over time. The patient revenue measure may adjust for quality between hospitals if difference in price between hospitals reflect differences in quality; however, because a quality-adjusted price index is not used to deflate revenue in a given year, there will be no adjustment for changes in quality over time. The authors plan to extend this analysis by either augmenting the output measure or directly including variables which measure differences in the quality of outputs both between firms and over time.[24]

A future option for developing an improved output measure is to make use of the increasing availability of information on separations and occupied bed days classified by detailed DRG. Each DRG represents a class of patients with similar clinical conditions requiring similar hospital services. A cost-weighted separations measure, based on DRG data, would adjust for changes in quality which are due to changes in the mix of cases across hospitals or over time, for example, hospitals which focus on providing complex, high-technology surgical services will record higher

[22] The HASAC formula of 5.753 treatments per bed day or 1.917 visits per bed day (implying 3 treatments per visit) was established in 1971, and although in widespread use until recently, is rather outdated. Recent estimates suggest that ratios of 7.102 treatments per bed day and 2 treatments per visit is appropriate for the 1990s. However, due to the relatively insignificant role of non-admitted patient services in Australian private hospitals, use of the more up-to-date ratios is unlikely to appreciably alter the results of this study.

[23] Newhouse (1994) discusses output measurement issues in respect of hospitals, such as the difficulties in measuring outputs and adjusting for quality when the product is heterogeneous and multi-dimensional, and the existence of omitted outputs. He also notes how existing studies have drastically aggregated both inputs and outputs since it is the only way to make headway when estimating production or cost functions.

[24] The ABS is currently undertaking a project investigating how to assess quality change within the health services industry. There are a number of industry initiatives to monitor changes in the quality of service provision, including development of Quality of Care and Patient Satisfaction Indicators, while outcome concepts such as Quality Adjusted Life Years are also gaining increasing attention. In practice, combination and weighting of these indicators to adjust activity-based measures of output is a further complex issue and the ABS is unlikely to be able to directly adjust output estimates for quality change for quite some time. However, quality indicators can readily be incorporated in DEA analysis, as illustrated by the use of unplanned readmission rates as an additional (negative) output variable for Victorian hospitals in Gregan and Bruce (1997).

output than a hospital which focuses on nursing home type care but has the same number of separations. However, changes in quality of output within DRGs will still not be captured.

DRG information is already collected for private hospitals, and published in Commonwealth Department of Health and Family Services (1996) at an aggregate level. Linking of the DRG data for individual private hospitals to the detailed financial and characteristics data collected in PHEC would enable more sophisticated measures of output to be developed and improved analysis of the performance of Australian private hospitals to be undertaken.

## 3.3 Price variables

The measurement of prices presented a number of difficulties for this study. In order to estimate a cost function, input prices must vary across firms and if profit functions are to be estimated, output price must also vary across firms. When panel data are analysed, prices do not need to vary across firms for estimation purposes though if prices do vary the data should capture this feature of the industry.

## 3.4 Input prices for cost functions

An average price of labour was calculated by dividing wages by FTE for all employees. It was possible to calculate prices for labour at a more disaggregated level, for example, nursing and non-nursing prices for the DEA analysis. Zuckerman, Hadley and Iezzoni (1994, p. 260) used an instrument for capturing the price of labour because '... average annual salary per full-time equivalent employee is used as the price of labour and this variable reflects hospitals' choices regarding the number and skill-mix of employees. Therefore, it is endogenous'. This appears to be a sound argument and will be addressed in forthcoming analysis.

Volume measures of intermediate inputs where unavailable in the current data to calculate average prices from intermediate input. As a result, dollar-based measures of intermediate inputs were calculated by dividing the composite occupied bed day measure of output. This produces a price per unit of output for intermediate inputs, and follows the methodology adopted in Ferrier and Lovell (1990).

The price of capital or the price of the flow of capital services was also a difficult concept to adequately measure. The standard way of measuring the price of capital would be to divide the value of capital inputs (i.e. the user cost of capital) by the quantity of capital inputs (e.g. the real value of the capital stock). The user cost of capital consists of depreciation, the nominal opportunity cost of holding capital and changes in the nominal price of capital.[25] However, due to data limitations the measure of the price of capital used in this paper can only be considered a rough approximation to this concept. An additional issue is the potential inappropriateness of this standard formulation of the price of capital to the non-market sector (e.g. NFP hospitals).

Two separate measures of the price of capital were developed, one was based on interest payments plus reported depreciation divided by the number of beds. A measure of the price of capital services based on our estimates of capital stock was also used to calculate a price of capital, replacing beds as the denominator, though this measure was not persisted with due to the problems (discussed earlier) with the capital stock measure.

[25] A more detailed discussion, formula and example is provided in Steering Committee on National performance Monitoring of Government Trading Enterprises (1992, p. 17–18).

## 3.5 Input and output deflators for production functions

Prices are also important to production function panel data estimation particularly in appropriately deflating dollar-based measures of volumes: that is, for revenue, intermediate inputs and labour cost measures. A price index was developed to deflate revenue based on benefit payments. In using these data, it was assumed that movements in benefits reflect movements in prices and secondly that prices are consistent across all hospitals. [26]

The deflator used Private Health Insurance Administration Council data covering the benefits paid on ordinary, reinsurance and supplementary benefit tables for the years 1991–92 to 1994–95 for private hospital procedures. The data covered the total benefit paid and the number of 'occupied bed days' claimed for a range of private hospital 'outputs': day only and overnight stays for advanced surgery, surgery/obstetrics, other medical services, psychiatric and rehabilitation for up to 14 days duration and over 14 days. We use the June quarter of each year (1992, 1993, 1994 and 1995) to represent the financial years and to remove any seasonality that may occur in the data. Index numbers were constructed by inferring a price as benefit ($) per patient day for each category, using 1991–92 as the base year.

In the production function analysis, dollar-based measures of labour were deflated using an index. This index was constructed using data on earnings and hours worked from the ABS Employees, Earnings and Hours Survey. Similarly the dollar-based measure of materials was also deflated in this case by an expenditure based deflator produced by the ABS and published by the Australian Instititute of Health and Welfare.[27]

## 3.6 Costs

Costs were calculated as the total expenditure of hospitals minus expenditure on new capital goods.

Summary descriptive statistics and correlation coefficients for a range of input and output measures are given in Appendix 1.

Table 3.1 presents a consolidated list of the definitions of input and output quantity variables used in the three estimation techniques in later sections of the paper. The more disaggregated input and output measures are used in the DEA technique, which is able to cope with zero values for variables.

[26] A Consumer Price Index measure of hospitals and medical costs was also available and despite the fact that movements in this deflator were consistent with the benefit-based deflator it was not persisted with as a component of this deflator is made up of non-hospital medical costs. Note that the use of benefits rather than insurance premiums is not inconsistent with consumer price index construction where the trend in benefits is assumed not to diverge significantly from the trend in premiums (Australian Bureau of Statistics 1987).

[27] Australian Institiute of Health and Welfare 1985, Table 12.

## Table 3.1: Variable definitions

| Variable | Definition |
|---|---|
| SMO | Total FTE professional medical officers |
| VMO | Total contract value ($) allied and medical health services |
| Nursing staff | Total FTE nursing staff (registered, enrolled and student/trainee and other nurses) |
| Other staff | Total FTE of staff other than medical professionals and nursing staff |
| Beds | Average number of total available beds (calculated on monthly figures) |
| Materials | Total value of recurrent expenditure on non-labour items (total recurrent expenditure less wages and salaries, superannuation, payroll tax, depreciation and VMO contract services) |
| Admissions | Weighted sum of inpatient, same day and accident and emergency separations, and non-inpatient occasions of service |
| Total staff I | Total FTE all staff employed |
| Total staff II | Total value ($) of wages, salaries and contracts for all staff and professionals employed |
| Acute care inpatient days | Total of inpatient and same day patient occupied beds days less nursing home type occupied bed days and surgery bed days |
| Accident and emergency treatments | Total number of accident and emergency (or casualty) treatments less patients admitted by presentation at accident/emergency department |
| Non-inpatient occasions of service | Total non-inpatient occasions of service (excluding accident/emergency and admitted patients) |
| Nursing home ID | Total nursing home type occupied bed days |
| Surgical procedures | Total number of surgical procedures performed (including advanced and minor surgery and obstetric procedures) |
| Advanced surgery | Total number of advanced surgery occupied bed days |
| Surgery | Total number of surgery occupied bed days |
| Minor surgery | Total number of minor surgery occupied bed days |
| Obstetrics | Total number of obstetrics occupied bed days |
| Psychiatric care | Total number of psychiatric occupied bed days |
| Rehabilitation | Total number of rehabilitation occupied bed days |
| Medical | Total number of medical occupied bed days |
| Inpatient separations | Total number of inpatient separations less nursing home type separations and surgery separations |
| Composite inpatient separations | Cost weighted sum of separations, with separations classified by principal diagnosis into 18 categories (in accordance with ICD9-CM Vol 1). |
| Composite output I | Weighted sum of inpatient occupied bed days and non-inpatient occasions of service |
| Revenue | Total patient revenue from admitted and non-admitted patients |

# 4 Data Envelopment Analysis results

The first part of this section presents and discusses results obtained from applying the DEA technique to a number of dataset model specifications. These specifications, based on the variable definitions in section 3, are presented below in table 4.1, which shows the input and output combinations used in 12 model specifications. Models 2 to 8 are based on slight modifications of model 1, which is used as the preferred model for studying hospital efficiency.

**Table 4.1: Model specifications**

| Variables(a) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Inputs** | | | | | | | | | | | | |
| SMO (FTE) | X | X | X | X | X | X | X | X | | | | |
| VMO ($ contract value) | X | X | X | X | X | X | X | X | | | | |
| Nursing staff (FTE) | X | X | X | X | X | | X | X | | | | |
| Other staff (FTE) | X | X | X | X | X | (b)X | X | X | | | | |
| Beds | X | X | X | X | X | X | X | (c)X | X | X | X | X |
| Materials (non-labour costs) | X | X | X | X | X | X | (d)X | X | (e)X | X | (e)X | X |
| Admissions | | | | X | | | | | | | | |
| Total staff I (FTE) | | | | | | | | | X | | X | |
| Total staff II (labour costs, $) | | | | | | | | | | X | | X |
| **Outputs** | | | | | | | | | | | | |
| Acute care inpatient days | X | X | | X | | X | X | X | | | | |
| Psychiatric care inpatient days | | | X | | | | | | | | | |
| Rehabilitation days | | | X | | | | | | | | | |
| Medical care days | | | X | | | | | | | | | |
| Surgery inpatient days | X | | | X | | X | X | X | | | | |
| Advanced surgery days | | | X | | | | | | | | | |
| Surgery days | | | X | | | | | | | | | |
| Minor surgery days | | | X | | | | | | | | | |
| Obstetrics days | | | X | | | | | | | | | |
| Non-inpatient occasions of service | X | X | X | X | X | X | X | X | | | | |
| Nursing home type inpatient days | X | X | X | X | (f)X | X | X | X | | | | |
| Surgical procedures | | X | | | X | | | | | | | |
| Acute care inpatient separations | | | | | X | | | | | | | |
| Accident/emergency | X | X | X | X | X | X | X | X | | | | |
| Composite output I | | | | | | | | | X | X | | |
| Total inpatient revenue | | | | | | | | | | | X | X |

(a) An X in the table indicates that the variables is included in the model.
(b) Other staff in this case includes nursing FTE.
(c) Non-discretionary input.
(d) Material costs (drugs and medical supplies, food and other domestic servies).
(e) Non-labour costs plus VMO contract valuation.
(f) Nursing home type separations.

Each of models 2 to 8 contains a minor definitional change (such as the inclusion or exclusion of a variable from a model) to the specification contained in model 1. For example, model 7 uses the same inputs and outputs as model 1 with the exception of a different definition of the materials variable (from non-labour costs to materials costs). Model 4 contains an additional input variable (total admissions) and model 6 contains only three labour input variables (as opposed to four in model 1). Model 1 was chosen as the preferred model because it was decided that occupied bed days was conceptually a better measure of output than separations. The model also gives a sensible spread of efficiency scores for the whole sample and contains a plausible number of variables (input plus outputs) when compared with the size of the overall sample. Models 9 to 12 are derived directly from models used in the SFA section of this paper (see section 5), and are used as comparison tools between DEA and SFA (see section 6). In the analysis that follows, the results of models 2 to 12 will be compared with those of model 1, since all of the models are derived to contain minor changes compared with model 1.

The DEA method provides relative efficiency scores for a particular sample. One important consideration in this analysis is whether any patterns observed in one particular model are common to a number of models when variable definitions and numbers change between the models. Because of the non-parametric nature of DEA, it is not possible to test this in the usual parametric manner associated with regression analysis. For this reason, we employ a number of models to analyse the **sensitivity of DEA results** to minor and major changes in either input or output variable definitions or the total number of variables included.

To test whether the results where sensitive to changes in model specification, a range of non-parametric testing methods (including the Mann-Whitney rank test and Spearman rank correlation test) were used. By assuming that the sample was composed of a number of homogeneous groups (for instance, FP and NFP hospitals), these non-parametric techniques where used to test the consistency of differences or similarities between the sub-samples over a range of different models. Whilst these tests have low statistical power,[28] due to their non-parametric nature, they can be used to provide some indication of whether two distributions are significantly different.

Specifically, based on Valdmanis (1992), the non-parametric Mann-Whitney test was used to investigate whether the similarity between FP and NFP hospital efficiency score distributions were affected by changes in model specification. The same procedure was used to test the similarity between the efficiency score distributions of NFPr and NFPo hospital types for each model under consideration. The differences between efficiency score distributions between hospitals of differing sizes were also analysed, this time using the Kruskal-Wallis test procedure for sub-samples of small, medium and large hospitals (defined by the average number of available beds).

Another approach adopted by the authors to compare different models was a test of the difference in ranks of individual hospitals, in terms of efficiency score, between two models. If a model is 'robust', the efficiency ordering of hospitals would tend to be similar between different models. The Spearman rank correlation test was used for this purpose, to analyse whether the ordering of firms between models (specifically between model 1 and each of the other models) was significantly different.

Because of its non-stochastic nature, the DEA technique is very susceptible to outliers in the data. This is particularly the case where an observation contains inputs which are significantly smaller, or outputs which are significantly larger, than other observations employing a similar input mix or producing a similar output level.

An outliers analysis of the current dataset[29] indicates that most hospitals use inputs and produce outputs commensurate with size, so that no significant outliers were discovered. One of the important discoveries related to the input-output mix of large hospitals. A significant number of the largest hospitals in the sample (those with more than 225 beds) showed significantly higher

---

[28] The power of a test is the probability of correctly rejecting the null hypothesis when it is false. Consequently, tests of low power may fail to reject the null even though it is false. In this case, the conclusion derived from a test may depend on the way in which the null hypothesis is stated.

[29] This test identifies observations with inputs or outputs lying more than 2.5 standard deviations on either side of the sample mean. More rigorous tests of outliers are outlined in section 7. For example, Wilson (1995), introduces a modification of the DEA technique to differentiate fully efficient firms by the 'distance' from a frontier estimated on the sample excluding the fully efficient firm. A referee also pointed out the usefulness of the 'box plot' test, outlined in Hughes and Yaisawarng (1998). The authors plan to trial these, and other recent developments in resampling techniques, in future research.

than average use of most (or all) inputs, but in most cases showed above average production in only a small number of the outputs. As a result, larger hospitals may appear at the lower end of the efficiency score range in the DEA analysis, in part because the data-set fails to adequately capture the complexity and technological advancement of operations performed.

The following presents the results of the DEA method applied to the model specifications listed in table 4.1. The authors were interested in both the sensitivity of the DEA technique to changes in specification and any conclusions that could be drawn about the efficiency of private hospitals in Australia. The results of applying DEA to each model outlined in table 4.1 for the **entire sample of private hospitals** (301 hospitals) are presented in table 4.2,[30] showing the mean efficiency score and standard deviation for each model for various measures of efficiency (technical efficiency, pure technical efficiency and scale efficiency) outlined in section 2.[31]

**Table 4.2: Mean efficiency scores, by model specification**

| All hospitals | Mean efficiency score (standard error) | | |
|---|---|---|---|
| Model | Technical efficiency | Pure technical efficiency | Scale efficiency |
| 1 | 0.734 (0.167) | 0.817 (0.181) | 0.905 (0.114) |
| 2 | 0.542 (0.216) | 0.682 (0.246) | 0.824 (0.204) |
| 3 | 0.861 (0.163) | 0.898 (0.158) | 0.958 (0.073) |
| 4 | 0.853 (0.138) | 0.881 (0.137) | 0.969 (0.054) |
| 5 | 0.441 (0.217) | 0.586 (0.282) | 0.805 (0.225) |
| 6 | 0.739 (0.190) | 0.785 (0.202) | 0.944 (0.092) |
| 7 | 0.765 (0.196) | 0.811 (0.195) | 0.944 (0.096) |
| 8 | 0.695 (0.181) | 0.800 (0.192) | 0.877 (0.135) |
| 9 | 0.287 (0.113) | 0.406 (0.181) | 0.757 (0.200) |
| 10 | 0.282 (0.112) | 0.393 (0.180) | 0.766 (0.187) |
| 11 | 0.667 (0.164) | 0.694 (0.174) | 0.961 (0.060) |
| 12 | 0.727 (0.153) | 0.748 (0.154) | 0.970 (0.056) |

The results in table 4.2 indicate that:

- As expected, the inclusion of additional variables or the disagregation of existing variables (while holding the number of observations constant) has the effect of increasing efficiency scores for observations which were not previously fully efficient. This effect is seen by the difference in average score between models 3 and 4 and model 1. This effect is discussed in detail in Nunamaker (1985).[32]
- In all of the models scale efficiency is greater than 75% (and greater than 90% in over half of the models) indicating that scale inefficiency is less important than pure technical efficiency as a source of private hospital inefficiency.
- In some cases, a change in the definition of a variable (such as the definition of surgery output from inpatient days to number of procedures (model 1 to 2)) had a large effect on the average level of efficiency of the sample.
- This is also the case when the output measure is changed from inpatient days to separations (models 1 to 5), where the average technical efficiency score for the sample falls from 73.4% in model 1 to 44.1% in model 5.

---

[30] All DEA LPs are solved using routines written for the Interactive Matrix Language module of SAS V6.12.

[31] The authors undertook some two-stage type analysis, using a Tobit regression approach, where it was found that input-output mix captured most of the variation in efficiency score for a particular sample.

[32] Nunamaker (1985) shows that no firm can become 'less' efficient by the additional of a variable, so that firms which were previously fully efficient will remain fully efficient with the addition of extra variables.

- Changing the definition of output from composite bed days to total patient revenue had the effect of dramatically increasing the average efficiency score (comparing model 9 with 11 and model 10 with 12).

Two explanations of the reason for average technical efficiency falling when separations are used instead of occupied bed days have been identified. Firstly, the inpatient days measure captures the output mix of different types of hospitals better than separations (which may discriminate against hospitals which specialise in treatments requiring longer hospital residence). An alternative explanation is that by using inpatient days we are rewarding hospitals which have a relatively slow patient turnaround over hospitals which treat patients 'efficiently' and quickly. Possibly the reason for the result is a combination of the two explanations, an observation which led the authors to look at models including average inpatient days per separation or including both inpatient days and separations output variables. The results of these models (not reported here) were almost identical to those from model 1, prompting the authors to adopt inpatient days as the output measure in the preferred model. The diversified mix of hospital types, and the 'output' of each, seems to be captured best by inpatient days in this instance.[33]

In the style of Valdmanis (1992), the results from each model were analysed by considering a **number of sub-sample structures** within the sample of efficiency scores.[34] For example, similarities between the distribution of efficiency scores for FP and NFP hospital sub-samples were tested using the Mann-Whitney statistic.

In most model specifications, the differences in technical and pure technical efficiency between various hospital types were insignificant, and therefore robust to model specification changes. Some points to note are:

- For model 1, results indicate that there is no significant difference between technical or pure technical efficiency between the FP and NFP subsamples.
- However, in the case of model 5, NFP hospitals were found to be significantly more technically efficient.
- In most cases, FP hospitals had a higher level of scale efficiency than NFP hospitals, though only in the case of model 3 was the difference significant.[35]
- Models 11 and 12 gave reasonably consistent results, both in terms of average efficiencies and the relationship between the sub-samples. However, the results were quite different from those obtained using model 1 (more disaggregated) and models 9 and 10 (using composite bed days as the sole output measure).

A similar analysis was done for the distribution of efficiency scores for **NFPr and NFPo** private hospital sub-samples within the NFP sub-sample of the previous analysis, based on results obtained from a frontier estimated on the whole sample.

The mean technical efficiency and technical efficiency distributions for different types of hospitals were not robust to even minor changes in model specification. However, the results from pure technical and scale efficiency appear to be more consistent over each model. In all cases, NFPr hospitals had a significantly higher level of pure technical efficiency and NFPo hospitals a significantly higher level of scale efficiency.

---

[33] As has been pointed out by a referee, this is despite the high correlation between the occupied bed day and separations measures. An explanation is that while the two measures are highly correlated, there is much greater diversity in the separations measure than the bed days measure.

[34] Copies of the tables relating to the sub-sample analyses can be obtained from the authors.

[35] This result might be explained by considering the different reasons for establishing each type of hospital. FP hospitals are established to earn a profit for owners, and will tend to be of a size which generates the highest profit (which should coincide with the size at which technical, and scale, efficiency, is greatest). On the other hand, NFP hospitals were generally established for charitable means, at a size which tended to provide the best service to the hospital's target 'clients'.

Similar to the analysis of profit status models, using revenue as an output measure gave very different results to models using occupied bed days.

The different general operating characteristics of the two types of NFP hospitals might be an explanation for these phenomena. As a group, NFPr hospitals could be characterised as large, metropolitan hospitals concerned primarily with the level of service or providing access to health care. On the other hand, NFPo hospitals comprise bush nursing, community and memorial hospitals and tend to be smaller in size. Because of this characteristic, NFPo hospitals may be more closely related to FP hospitals, in that they may operate at a more 'economically' feasible scale than the larger hospitals in the sample.

Another approach in examining the sensitivity of DEA results to model specification is to look at the breakdown of hospital efficiency score distributions by **size**. To proxy size, the number of available beds for each hospital was used, with the following definitions for three size categories: small (fewer than 25 beds), medium (25 to 100 beds), and large (more than 100 beds).

Mean efficiency scores by type (with standard deviation) were calculated and a Kruskal-Wallis test statistic for the hypothesis that the three efficiency score distributions are not significantly different from each other derived.

The hypothesis of similarity is rejected in all cases for the pure technical and scale efficiency cases at anything higher than the 1% significance level. However, in the case of technical efficiency scores some points to note are:

- Most of the models show the order of average technical efficiency as small> medium>large, and this the case for the preferred model (model 1).
- In all cases, the scale efficiency of large hospitals was significantly less than the scale efficiency for medium and small hospitals.
- These results appear to be the least sensitive to changes in model specification of the results presented so far.
- Including fewer labour variables in model 6 (compared with model 1), completely reversed the ordering indicated by model 1, with large hospitals showing greater average technical efficiency than medium and small hospitals.
- Unlike the results obtained using models with occupied bed days as the measure of output, where average efficiency appears to decrease with size, the results for models using revenue show average efficiency increasing with size for both technical and pure technical efficiency.

A possible reason for the ordering of mean efficiencies for most models could be that the input-output set does not adequately allow for the more complicated treatments that may occur in larger hospitals. For one of the models, this theory was tested by analysing efficiency scores in a two-stage, regression-based approach including a measure of treatment technology. In summary, it appears as if the main cause of inefficiency in small and medium hospitals is pure technical inefficiency, whereas in larger hospitals the major cause is scale inefficiency.

Overall, the results, in terms of average technical and pure technical efficiency, appear to be quite sensitive to the choice of model. This is particularly so when comparing the average efficiency scores for various sub-samples of hospitals (by ownership type and size). While general patterns can be identified from the results not all models follow the identified patterns. The sensitivity analysis indicates that the method is not robust to model specification changes, even when they are quite minor such as a definitional change to a single variable, implying that caution must be used when choosing a model for this type of analysis. When results are sensitive to variable set changes, an analyst must be able to determine what effect the particular variable set has on the results obtained and how results may differ when a different set is used in the analysis.

Continuing the analysis of the differences between various model specifications, consider the **correlations between the rankings** of individual hospitals (in terms of technical and pure technical efficiency) between different models. The Spearman rank correlation test was applied to compare the ranking of individual observations between each of the 12 models in table 4.1, based on a frontier estimated on the pooled sample using both constant- and variable-returns formulations.

Analysing the sensitivity of this comparison method to model specification in terms of the ordering of individual observations, the following conclusions were drawn:

- The test statistic is significant in all but one of the comparisons, indicating that the hypothesis that the ranks of observations between models are not correlated cannot be accepted (the exception occurs in the comparison between models 4 and 5).
- In most cases the test statistic is greater than 0.70, indicating a correlation reasonably close to monotonically increasing for the inter-model comparisons.
- Model 5 appears to have low correlation with model 1, indicating that the choice of outputs measured by inpatient days or separations is a major determinant of individual efficiency ranking.
- Models 11 and 12 have a very low correlation with model 1 (around 30%), indicating that the use of revenue as an output indicator significantly alters the patterns efficiency from models using occupied bed days (either in composite or disaggregated form).

These results indicate that the DEA method can be sensitive, in terms of the ordering of individual scores between models, to changes in model specification.

In terms of both average scores and rankings, the method cannot be considered robust to changes in model specification. For this reason, the results of any DEA study must be explained not only in the context of the sample being studied, but also the coverage of the variables used to measure the input and output of each observation.

## 4.1 Distribution of technical efficiency scores

Figure 4.1 illustrates the empirical distribution of technical efficiency scores obtained from model 1 in table 4.1. The graph indicates the frequency of efficiency score observations for a bandwidth of 0.025. This division of categories was chosen to illustrate the major features of the distribution.

An important feature of the distribution is the bimodal character. Whilst the distribution is generally bell-shaped (one of the distributional shapes we could expect to observe), the graph indicates a large spike occurring at $\theta=1$. For such a large number of fully efficient observations, a distribution showing exponential decay away from $\theta=1$ might be expected.[36] However, there are far fewer observations showing efficient scores between 0.9 (90%) and 0.99 (99%) than would be expected for this type of distribution.

In order to validate this major feature of the distribution, figure 4.1 also shows a smoothed density estimate of the empirical distribution of efficiency scores. This was obtained using a kernel density estimator of the form

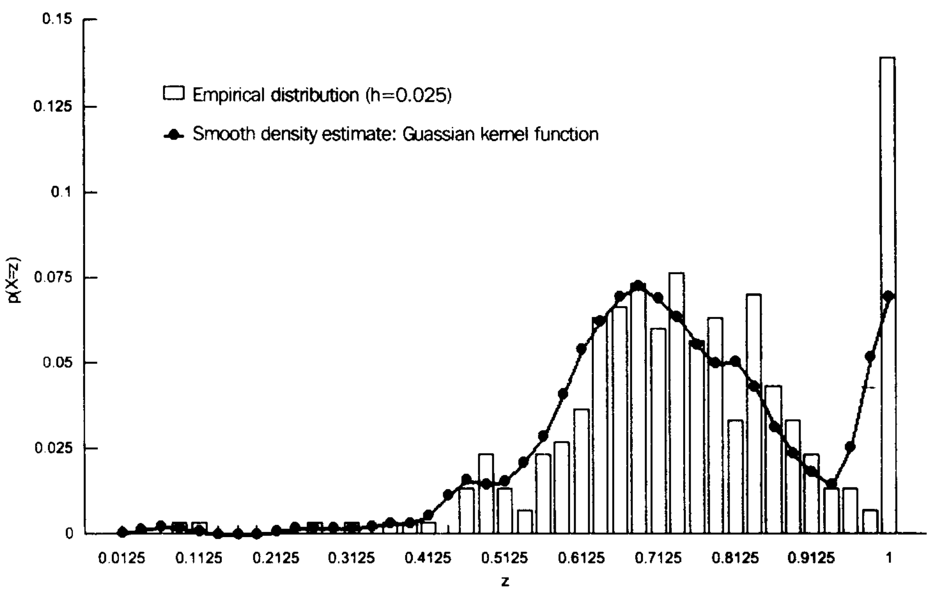
$$\hat{f}(z) = \frac{1}{hn}\sum_{i=1}^{n} K\left(\frac{z-x_i}{h}\right),$$

[36] Note that when the width of the histogram divisions increases (e.g. to 0.05 and 0.1), the efficiency score distributions more closely resemble a bell-curve. In this case, the large number of fully efficient observations is a function of the number of variables included in the model. Models with fewer variables (such as models 9 and 10 from table 4.1) show far fewer fully efficient observations and hence show distributions which more closely resemble a bell-curve distribution. Gstach (1995) mentions a number of studies in which the bimodal nature of the DEA score distribution is found, confirming a 'natural bimodality'.

where $h$ is the bandwidth, $n$ is the number of observations, $x_1, \ldots, x_n$ is a univariate sample of observations on a continuous random variable $X$ with probability density function $f(\cdot)$, and $K(u)$ is the kernel function, with the property $\int_{-\infty}^{+\infty} K(t)dt = 1$. The function $\hat{f}(z)$ is an approximation of the probability density function, based on the frequency of observations in the univariate sample.

Using the Gaussian kernel function, $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$, the smoothed density estimate for the empirical distribution of efficiency scores is $P(X = z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{z - x_i}{h})^2}$

**Figure 4.1: Efficiency score distribution and smooth density estimate: model 1**




## 4.2 Analysis of economies of scale

Using a comparison of the VRS and NIRS efficiency scores allows information to be inferred about the nature of **economies of scale** for each hospital, the overall sample and the various sub-samples of hospitals. Comparing the results of CRS, VRS and NIRS LPs for models 1 and 11 allows an observation to be characterised as operating with either constant, increasing or decreasing returns to scale. The results of this analysis are shown below in table 4.3, which presents the count of observations (for both the population and sub-samples) operating under constant, increasing or decreasing returns to scale for the two models.

The results indicate that a majority of hospitals appear to operate in areas of decreasing returns to scale, and this is particularly true for hospitals in the medium and large size groups. We note that the number of hospitals operating with constant returns may also depend on the chosen model specification, since hospitals which are categorised as operating under CRS are those which appear fully efficient using the CRS model formulation. Models with fewer variables (inputs and outputs) will tend to show fewer hospitals operating under constant returns, due to the nature of the DEA technique (see Nunamaker 1985 for a discussion).

**Table 4.3: Economies of scale (number of firms), by hospital type and model**

| Sample population | Number of Obs | Model 1 | | | Model 11 | | |
|---|---|---|---|---|---|---|---|
| | | Constant returns | Increasing returns | Decreasing returns | Constant returns | Increasing returns | Decreasing returns |
| **All hospitals** | 301 | 40 | 49 | 212 | 12 | 175 | 114 |
| **FP hospitals** | 155 | 21 | 19 | 115 | 9 | 81 | 65 |
| **NFP hospitals** | 146 | 19 | 30 | 97 | 3 | 94 | 49 |
| **Small hospitals (less than 25 beds)** | 65 | 21 | 26 | 18 | 3 | 61 | 1 |
| **Medium hospitals (25 to 100 beds)** | 180 | 17 | 23 | 140 | 7 | 112 | 61 |
| **Large hospitals (more than 100 beds)** | 56 | 2 | | 54 | 2 | 2 | 52 |

Another method of analysing scale and returns to scale, due to Banker, Conrad and Strauss (1986), is the concept of **most productive scale size** (MPSS). Using the results from the constant returns DEA LP and the number of beds for each observation, the MPSS for observation $i$ is defined as

$$MPSS_i = (\theta_{CRS} / \textstyle\sum_j \lambda_j) x_i$$

where $\theta_{CRS}$ is the efficiency score for observation $i$, $\lambda_j$ are the linear combination parameters for an observation from the DEA LP given by equation (1) in section 2.1 and $x_i$ the number of beds associated with observation $i$.

If the MPSS measure is very much smaller than the actual beds measure, Banker, Conrad and Strauss (1986) interpret this as inferring decreasing returns, and similarly for MPSS much larger than the number of beds inferring increasing returns to scale. Table 4.4 gives the results of this analysis applied to the overall sample and the various hospital type and size sub-samples, showing the mean actual size and mean MPSS for each of the profit and size groups. The table also gives the results of Mann-Whitney (and Kruskal-Wallis in the case of the size comparison) tests on the similarity of the means (and distributions) of actual beds, MPSS and ratio of beds to MPSS for related hospital groups.

**Table 4.4: Comparison of mean MPSS (number of beds), by hospital type**

| Measure vs population | Overall | Profit | NFP | NFPr | NFPo | Small | Medium | Large |
|---|---|---|---|---|---|---|---|---|
| Mean actual beds (SD) | 69.6 (62.9) | 64.0 (24.3) | 75.5 (76.5) | 118.3 (85.6) | 36.1 (35.9) | 14.4 (13.2) | 56.6 (26.2) | 175.3 (23.4) |
| Mean MPSS (SD) | 22.9 (18.4) | 24.3 (19.1) | 21.4 (17.5) | 21.3 (17.3) | 21.5 (17.8) | 13.2 (8.8) | 26.2 (17.6) | 23.4 (24.4) |
| Ratio (mean MPSS to mean beds) | 0.33 | 0.38 | 0.28 | 0.18 | 0.6 | 0.92 | 0.46 | 0.13 |
| Test statistic (similarity of beds)(a) | | -0.84 | | -6.91 | | | (b)229.57 | |
| Test statistic (similarity of MPSS) | | -1.61 | | -0.11 | | | 35.1 | |
| Test statistic (similarity of ratio) | | -0.27 | | 5.04 | | | 94.9 | |

(a) Tests are Mann-Whitney (standard normal) unless notes. The similarlity of actural beds , MPSS and size ratio is tested for FP vs NFP, NFPo vs NFPr and size.
(b) Tests for the similarlity of size sub-samples are Kruskal-Wallis tests ($\chi^2$ with 2 degrees of freedom, 5% critical value = 6.00).

A number of observations can be made from these results:

- Mean MPSS for the sample is 22.9 beds, compared with 69.9 as the mean number of beds for the sample. This indicates that, on average, hospitals operate in excess of the optimal scale, and could be characterised by decreasing returns to scale, mirroring the results presented in table 4.3.
- There are no significant differences between the actual size, MPSS and size ratio for for-profit and NFP hospitals sub-sample (religious or charitable against other).
- Whilst there are significant differences (as expected) for actual size and size ratio between NFPr and NFPo, the test for similarity of the MPSS for these groups is not rejected.
- In terms of size, tests for all three measures indicate significant differences. In fact, the MPSS for medium hospitals is larger than that for large hospitals, indicating that decreasing returns to scale prevails to a major degree in the latter group.
- In unreported Mann-Whitney tests, comparing actual beds against MPSS for each sub-sample, the hypothesis of similarity is rejected for all but the small hospitals sub-sample. In all of the rejected tests, MPSS was significantly smaller than actual size (again reflecting the predominance of hospitals operating under decreasing returns to scale).

## 4.3 Decomposition of efficiency scores

Congestion occurs when a firm is unable to dispose of unwanted inputs costlessly, creating a situation of negative marginal returns to inputs. As discussed previously, removing the usual assumption of strong disposability, we can further decompose technical efficiency scores into 'pure' technical efficiency, scale efficiency and congestion effects. Table 4.5 presents the results of this decomposition for model 1 from table 4.1.

**Table 4.5: Decompositions of estimated efficiency scores, model 1**

| Population(a) | Technical efficiency | Pure TE under weak disposability | Congestion efficiency | Scale efficiency | Allocative efficiency |
|---|---|---|---|---|---|
| Overall | 0.734 (0.167) | 0.892 (0.161) | 0.920 (0.128) | 0.906 (0.114) | 0.657 (0.214) |
| For profit | 0.740 (0.158) | 0.893 (0.157) | 0.919 (0.104) | 0.914 (0.104) | 0.659 (0.209) |
| NFPr | 0.719 (0.152) | 0.935 (0.121) | 0.925 (0.124) | 0.835 (0.137) | 0.703 (0.206) |
| NFPo | 0.738 (0.196) | 0.853 (0.192) | 0.917 (0.144) | 0.954 (0.076) | 0.610 (0.222) |
| CRS | 0.791 (0.164) | 0.915 (0.133) | 0.931 (0.115) | 0.933 (0.093) | 0.696 (0.207) |
| DRS | 0.705 (0.102) | 0.944 (0.116) | 0.932 (0.093) | 0.813 (0.115) | 0.730 (0.184) |
| IRS | 0.645 (0.175) | 0.798 (0.207) | 0.886 (0.170) | 0.936 (0.109) | 0.509 (0.180) |

(a) The table shows the average score and (standard deviation) for each population.

The results indicate that congestion, or the inability to reduce unwanted inputs costlessly, plays the smallest role in determining overall technical efficiency. On average, hospitals which have congestion inefficiency could reduce their inputs by a further 8.0% compared with strong disposability.[37] Overall, hospitals which are not fully efficient could reduce their use of inputs by 26.6% compared with the most efficient hospitals in the sample. Scale inefficiency accounts for 9.4% of the inefficiency, with the remainder accounted for by 'pure' technical inefficiency.[38]

The last column of table 4.5 shows average allocative efficiency for the sample of various hospital types. Whereas the first four columns are concerned with notions of technical efficiency (how well an observation uses its inputs to produce output) allocative efficiency is concerned with

[37] This reduction is indicated by an average congestion efficiency of 0.92 (or 92%) for the overall sample from table 4.5.
[38] In this context, pure technical inefficiency accounts for inefficiency due to any factors other than input disposability and scale.

how closely a firm's input usage corresponds to the cost-minimising vector of inputs, given its output and a set of input prices. Before presenting the results of this analysis we note they can be very unreliable due to the lack of readily definable input prices for the some of the input variables. In fact, one of the appealing features of the DEA model is that input prices are not needed to calculate firm level efficiency.

## 4.4 Input slacks

Input slacks occur in DEA analysis when the projection of an inefficient observation onto the efficient plane occurs in such a manner that the further reduction of one or more inputs is possible. Considering only two dimensions, and referring to figure 2.2 in section 2, input slacks (for inputs 1 and 2 respectively) would occur if an observation was projected onto the frontier regions ***fs'*** or ***as***. The treatment of such firms raises important points about DEA-measured efficiency and the reporting of efficiency scores.[39]

Looking at model 1 (from table 4.1), we present an aggregate analysis of the amount of input slack estimatedfor individual hospitals in the sample. Table 4.6 reports the number of observations with slacks for each of the six input variables in the model, as well as the average amount of each input slack and the level of total input slacks for those observations as a percentage of the total input use by those hospitals reporting slacks for each input type.

**Table 4.6: Analysis of input slacks (model 1, sample size=301): residual method**

| Input | Input description (measure) | No. of observations with input slack | Average input slack | Total slack as a % of total input use |
|---|---|---|---|---|
| 1 | SMO (FTE) | 77 | 2.77 | 65.16 |
| 2 | VMO ($'000) | 56 | 1.93 | 51.05 |
| 3 | Nursing staff (FTE) | 87 | 16.38 | 14.01 |
| 4 | Other staff (FTE) | 128 | 19.78 | 24.97 |
| 5 | Available beds (no.) | 29 | 3.71 | 12.19 |
| 6 | Non-labour costs ($'000) | 49 | 4.05 | 12.87 |

The results in table 4.6 indicate that hospitals with a positive slack could use, on average, 2.77 fewer SMO FTE, $193,000 fewer in VMO contracts, 16.38 fewer nursing FTEs, 19.78 fewer other staff FTEs, 3.71 fewer available beds and $405,000 fewer in non-labour costs. Input slacks range from 12.2 to 65.2% of total input usage. Some other observations from these results are:

- Of the 225 observations for which at least one input slack is estimated (74.75% of the sample), 128 (57%) have slacks for input 4 (other staff), whilst only 29 (12.9%) have slacks for input 5 (available beds).
- While slacks for input 4 are the most common, input slacks for inputs 1 and 2 (SMOs and VMO contracts respectively) are by far the largest as a percentage of total input use by hospitals with such a slack. In the case of SMOs, input slacks are 65.2% of total input usage for the 77 hospitals with slacks for input 1. This indicates that hospitals with slacks in SMO or VMO could further reduce the use of these inputs by a considerable amount, compared with hospitals with slacks in other input types.
- The input slack for available beds is the least significant, both in terms of frequency and magnitude, with the average slack being 3.7 beds compared with an average available beds measure for the sample of 69.6.

[39] Is it sufficient to simply 'project' a firm onto the efficient frontier, or, for such observations, should the amount of any additional input reduction (input slack) be reported in addition to the standard DEA efficiency score? More advanced methods for the treatment of slacks, and incorporation into efficiency, are being developed, but have not been usedin this analysis.

The method outlined above to deal with input slacks is called the one-stage, or residual slack, method. Ali and Seifort (1993) describe a method to deal with slacks using the output of the CRS DEA LP as the input into a second-stage LP. Solving this LP from each observation gives the maximum sum of input and output slacks required to move an inefficient frontier point to an efficient frontier point.

The results obtained from this analysis (not reported here) are almost identical to the results obtained from the residual analysis of slacks (reported in table 4.6), with 225 observations having at least one input slack, and the number of observations, average slacks and percentage of input use differing only marginally from the results on table 4.6.

Table 4.7, below, presents results for a one-stage method analysis of slacks for model 11 in table 4.1.

**Table 4.7: Analysis of input slacks (model 11, sample size=300): one-stage method**

| Input | Input description (measure) | No. of observations with input slack | Average input slack | Total slack as a % of total input use |
|---|---|---|---|---|
| 1 | Available beds (number) | 74 | 3.73 | 10.47 |
| 2 | Total staff (FTE) | 19 | 51.83 | 12.74 |
| 3 | Non-labour costs ($) | 1 | 11.51 | 5.15 |

Some points to note from the results in table 4.7 are:

- Of the 93 observations with slacks, 74 have slacks for beds, with only one observation showing an input slack for materials.
- Slacks account for only 5 to 12% of input use, in contrast to the results from model 1 (which slacks accounted for as much as 60% of input use for one input).

This treatment of slacks should be treated with some caution, particularly since the methods discussed here are not invariant to the units of measurement for each input, and, in the case of the two-stage method, the approach of maximisation may be incompatible with the minimisation approach of the first stage. Recently, a multistage method for the treatment of slacks has been developed (Ali and Seifort 1993), and it is planned to trial this method and contrast the results with those presented above in future analysis.

The analysis indicates the input slacks account for a significant proportion of input usage, particularly for the input of labour, and especially medical officers and other staff. For this reason the reporting of efficiency scores alone may not sufficiently characterise the nature of efficiency within the sample.

## 4.5 Panel data and temporal efficiency analysis

Another important area of analysis concerns the movement in the productivity and efficiency of hospitals over time. The Malmqvist productivity index was used to decompose inter-period productivity changes into efficiency and technological/technical changes, following the method developed by Fare, Grosskopf and Roos (1995). Using a method introduced by Balk and Althin (1996), the indices were transformed into a transitive form to better allow intertemporal comparisons. To extend the data, three additional time periods (1991–92 to 1993–94) were added to the dataset for models 1, 9 and 11 from table 4.1. The results of this efficiency analysis are given in table 4.8.

**Table 4.8: Malmqvist (transitive) productivity indices, 1991–92 to 1994–95**

| | Model | 1991–92 – 1992–93 | 1992–93 – 1993–94 | 1993–94 – 1994–95 | 1991–92 – 1994–95 |
|---|---|---|---|---|---|
| Efficiency index | 1 | 99.17 | 98.51 | 101.48 | |
| Technical index | | 91.46 | 103.99 | 98.21 | |
| Productivity index (Caves/Balk) | | 90.7 | 102.44 | 99.66 | 92.60 |
| Efficiency index | 9 | 107.75 | 152.56 | 109.94 | |
| Technical index | | 84.82 | 99.52 | 104.08 | |
| Productivity index (Caves/Balk) | | 91.46 | 151.83 | 114.43 | 158.90 |
| Efficiency index | 11 | 94.88 | 121.25 | 87.22 | |
| Technical index | | 82.35 | 77.75 | 108.01 | |
| Productivity index (Caves/Balk) | | 78.13 | 94.27 | 94.2 | 69.38 |

Overall the change in productivity for model 1, using the properties of transitive indices, can be calculated as

$$M_{14} = M_{12} \times M_{23} \times M_{34} = 0.907 \times 1.0244 \times 0.9966 = 0.9260$$

indicating a 7.4% increase in productivity in the sector over the four-year period. Whilst efficiency has remained relatively constant throughout the period, a large rise in 'technology' in the 1991–92 to 1992–93 period is largely responsible for the increase in overall productivity during the period. The authors acknowledge that the time period used is much too short to draw any firm conclusions about the movements in productivity in the sector. In particular, there is no information about the cause of the large change in the technology index which seems to drive the overall movements for the three periods.

Whilst the magnitudes of the indices vary quite considerably over the three models examined, the movements from year to year are generally consistent. Generally, there appears to be a pattern of a rise in productivity, driven mainly by a large rise in technical progress, in 1991–92 to 1992–93, followed by a decrease in productivity in the next period and a small rise again in the last period.

The results of DEA applied to these models will be used in the future as a starting point for a comparison of the SFA and DEA techniques. In this comparison, the authors are interested in the way in which these techniques order similar observations, and the similarities and differences in the ordering (i.e. the correlation of the ranks).

At the present time, however, these results again indicate the lack of robustness of the technique to changes in model specification and size and variable definitions, even when the changes are relatively minor. This is well demonstrated by the dramatic effects of a change from measuring output in terms of occupied bed days to inpatient revenue, as occurs between models 1, 9 and 11. The results of DEA applied to a sample cannot be interpreted independently of the characteristics of the sample, both in terms of the number of variables employed (relative to the number of observations) and the specific definitions used for each variable.

## 4.6 Conclusions

The purposes of this study were twofold; firstly to evaluate the robustness of a productivity analysis technique in the light of different model specifications, and secondly to draw some conclusions about the nature and pattern of efficiency within the Australian private hospital industry. Using the results presented in the previous section, a number of important observations can be made about the application and operation of the DEA methodology:

- The results presented for a range of model (input-output) specifications are not particularly robust to specification changes, where even minor variable definitional changes can produce different results.
- The comparison of mean efficiency by major ownership type (FP or NFP) showed a wide range of results from significant differences in either direction to insignificant differences.
- The comparison of rank correlations for each model with model 1 indicated that all were positive and significantly different from zero, with correlation coefficients ranging from 0.49 to 0.95.
- The lack of robustness is perhaps not surprising given the large sample size (301 observations) and the relatively small number of variables (a maximum of 16) when compared with previous studies of this type.
- Directions for future research (discussed in section 7) include implementing recent developments in detecting influential outliers in DEA analysis (for example, Wilson 1995), and applying a range of resampling techniques (including jack-knifing and bootstrapping) to develop most statistically robust measures of estimated frontiers.
- To conclude, it appears as if DEA results are as much driven by the specific data used in the models, both in nature and sample size, as the actual nature of the hospitals from which the data are gathered. While the method is very useful in analysing firm level efficiency without the need to impose a pre-defined functional form for 'production', care must be taken to analyse the results in conjunction with the data used in the study and the relative sizes of the sample and the variable set.

# 5 Parametric analysis

## 5.1 Production function estimation

A number of models were estimated where inputs and outputs were represented by a variety of variables, see table 5.1. These models were estimated on both the cross-section and the panel of data. Each model is estimated using both Cobb-Douglas and Translog functional forms.

**Table 5.1: Model specifications**

| Variables(a)(b) | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| **Inputs** | | | | |
| Beds | X | X | X | X |
| Capital stock | | | | |
| Materials I (non-labour costs) | | X | | X |
| Materials II (including VMOs) | X | | X | |
| Total staff I (total FTE) | X | | X | |
| Total staff II (labour costs) | | X | | X |
| **Outputs** | | | | |
| Revenue | | | X | X |
| Composite output I (occupied bed days) | X | X | | |

(a) X indicates that the variable in present in the model.
(b) Models 1 through 4 correspond with models 9 through 12 in section 4.

### Cross-section results

OLS and MLE results for models 1 and 3 obtained by estimating a Cobb-Douglas production function on the full sample of 300 hospitals[40] are presented in table 5.2. The MLE estimates relate to the stochastic frontier production function. The results of estimating models 2 and 4 rarely differed from models 1 and 3 respectively and therefore are only reported where they did differ or to indicate the effect of variations in defining the input set.

**Table 5.2: Cobb-Douglas production function estimation results, models 1 and 3 (n=300)**

| | Model 1 | | | | Model 3 | | | |
|---|---|---|---|---|---|---|---|---|
| **Coefficients** | OLS | SEs | MLE | SEs | OLS | SEs | MLE | SEs |
| Constant | 6.31 | 0.44 | 7.15 | 0.35 | 6.57 | 0.34 | 7.20 | 0.25 |
| Beds | 0.27 | 0.08 | 0.22 | 0.06 | 0.09 | 0.06 | 0.12 | 0.05 |
| Materials | –0.09 | 0.05 | –0.11 | 0.04 | 0.42 | 0.04 | 0.39 | 0.02 |
| Total staff | 0.77 | 0.07 | 0.78 | 0.06 | 0.57 | 0.06 | 0.56 | 0.05 |
| **Diagnostics** | | | | | | | | |
| $s_u/s_v$ | | | 2.41 | 0.34 | | | 3.73 | 0.53 |
| $\sqrt{(s^2_u+s^2_v)}$(a) | | | 0.54 | 0.01 | | | 0.42 | 0.01 |
| Adj $R^2$ | 0.88 | | | | 0.95 | | | |
| Log likelihood | –136 | | –111 | | –58 | | –10 | |

(a) This term represents the square root of the sum of variances of the stochastic and efficiency error terms.

[40] The sample used for this analysis consists of all acute care hospitals with non-zero values for each of the specified inputs and outputs.

All inputs and outputs were transformed into logarithmic form for estimation, so that the estimated parameters represent the input elasticities for output. All stochastic models seem to be significant improvements on their OLS counterparts given what appears to be highly significant stochastic parameters (calculated using the coefficients and their respective standard errors). One-sided likelihood ratio test statistics,[41] comparing the results of each regression, were calculated using the log likelihood of each regression. For example, the estimated test statistic for MLE versus OLS in model 1 is 50, with a critical value for the test of 2.71 at a 5% significance level. The relevant critical value of likelihood ratio test statistics for other stochastic frontier models will differ depending on the number of additional parameters required for estimation.[42]

The OLS estimates of the production function are useful for examining the presence of heteroskedasticity and non-normality in the errors. A Breusch-Pagan test for heteroskedasticity was calculated for models 1 and 3, test statistics were 192 and 34.2 respectively. The critical value for this test is 7.81 thus indicating the presence of substantial heteroskedasticity. While White's method was used to correct standard errors for heteroskedasticity with OLS estimation,[43] the authors were unclear on how to correct standard error estimates in the case of the MLE estimates. These results do indicate that the standard errors of these regressions should be treated with caution. A Jarque-Bera test for non-normality of the errors was also conducted producing test statistics 25.2 and 24.1 for models 1 and 3 respectively. The critical value of this test is 5.99 thus again indicating that these regression results contain substantial statistical weaknesses.

Examining the results of MLE estimates the following features were noted: Model 1 (which uses occupied bed days as a measure of output) produced significant and appropriately signed coefficients on the variables representing capital and labour. However, the coefficient on the materials variables was insignificant and negative. Model 3 (which uses revenue as a measure of output) produced correctly signed coefficients for all inputs. The negative materials output elasticity in model 1 may relate to the inability of the occupied bed day measure of output to capture appropriate differences in the severity of cases.[44] That is, those hospitals which treat high severity cases and use high levels of material inputs will appear to have a similar level of output to hospitals which do not treat these cases and use far fewer material inputs.

---

[41] As discussed in section 2.2, direct estimation of the production function will not provide consistent parameter estimates under the assumption of cost minimisation or profit maximisation. Thomas (1985, p. 225) discusses a method of obtaining consistent (but not unbiased) parameter estimates, based on a method originally proposed by Klein (1953).

The Thomas/Klein method involves estimating the output elasticities with respect to each input by relating them to factor shares, and is valid only if marginal products equal factor prices (e.g. under perfectly competitive factor markets). To calculate mean factor shares, and therefore the consistent parameter estimates, total cost or price data is required for the inputs.

The labour share is estimated using data on total labour expenditure divided by a measure of the value of output (total patient revenue). To test the sensitivity to the adopted output measure an alternative labour share measure was calculated using total recurrent expenditure as the denominator. The estimated output elasticity with respect to labour is 0.63 under the first method (revenue) and 0.62 under the second method (expenditure). The same methodology was used to estimate the output elasticity with respect to intermediate inputs. The output elasticity with respect to the first intermediate inputs measure was estimated to be 0.29 under both approaches. These estimates are entirely feasible, but do not accord with the parameter estimates obtained in the output regressions which were generally negative. Ideally this method would also be used to estimate the elasticity of output with respect to capital. However, information on the total value of capital inputs is not available and nor is a price of capital. We have assumed the quantity of capital inputs moves with the stock of available beds in the production function estimation. One option is to use depreciation charges and interest expenses to approximate the value of capital inputs. After excluding hospitals with zero depreciation the parameter estimate was calculated in a similar fashion to that outlined above for the other inputs. However, the estimate obtained of 0.06 is probably too low, and a consequence of the inadequate measure adopted to represent the value of capital inputs.

[42] See Coelli (1993) for a full discussion of the properties of this test and associated critical values.

[43] An alternative would be to use weighted least squares estimation to correct for heteroskedasticity.

[44] Output elasticity refers to the marginal percentage change in output resulting from a 1% change in one of the inputs, with the other inputs held constant. For a Cobb-Douglas mode, this is simply the coefficient for each of the input variables in the equation.

This issue, along with the difficulties in interpreting revenue as an output measure, is discussed more fully in the concluding paragraphs of this section.

Cobb-Douglas models which imposed CRS were also estimated. Likelihood ratio tests which compare these restricted models with the unrestricted models were then calculated. In all models the restriction of CRS is rejected. In models 1 and 2 the sum of coefficients was less than 1 indicating decreasing returns to scale whilst it was greater than 1 in models 3 and 4 indicating increasing returns to scale.

The results of estimating models 1 and 3 using a Translog production function are presented in table 5.3. Care should be taken in interpreting these results due to the high degree of multicollinearity between the input variables (see Appendix 1).

**Table 5.3: Translog production function estimation, models 1 and 3 (n=300)**

| | Model 1 | | | | Model 3 | | | |
|---|---|---|---|---|---|---|---|---|
| **Coefficients** | OLS | SEs | MLE | SEs | OLS | SEs | MLE | SEs |
| Constant | 27.72 | 6.77 | 32.42 | 4.45 | 15.74 | 5.19 | 11.41 | 4.20 |
| Beds | 0.31 | 1.36 | 0.49 | 1.29 | 2.53 | 1.04 | 2.10 | 0.90 |
| Materials | –4.48 | 1.41 | –5.31 | 0.91 | –1.92 | 1.08 | –0.82 | 0.85 |
| Labour | 5.16 | 1.72 | 5.83 | 1.34 | 1.89 | 1.32 | 0.89 | 1.25 |
| Beds squared | –0.36 | 0.34 | –0.20 | 0.26 | –0.34 | 0.26 | 0.08 | 0.19 |
| Labour squared | 0.05 | 0.24 | 0.20 | 0.18 | –0.03 | 0.19 | –0.04 | 0.17 |
| Materials squared | 0.45 | 0.15 | 0.54 | 0.09 | 0.27 | 0.11 | –0.15 | 0.08 |
| Beds x labour(a) | 0.39 | 0.22 | 0.31 | 0.17 | 0.38 | 0.17 | 0.13 | 0.10 |
| Beds x materials(a) | –0.01 | 0.15 | –0.05 | 0.13 | –0.19 | 0.11 | –0.20 | 0.10 |
| Labour x materials(a) | –0.44 | 0.18 | –0.51 | 0.14 | –0.19 | 0.14 | –0.04 | 0.13 |
| **Diagnostics** | | | | | | | | |
| σu/σv | | | 2.73 | 0.44 | | | 5.06 | 1.01 |
| $\sqrt{(\sigma^2_u+\sigma^2_v)}$ | | | 0.53 | 0.01 | | | 0.42 | 0.01 |
| Adj $R^2$ | 0.88 | | | | 0.95 | | | |
| Log likelihood | –127 | | –99 | | –48 | | 4.50 | |

(a) These terms represents the combinations of ineractions between each of the three inputs in the translog functional form.

The coefficients of the Translog cannot be easily directly interpreted, though an initial inspection of the coefficients seems to indicate that the materials variables may be having a negative impact on production in model 1. The insignificance of many of the Translog coefficients may be a result of multicollinearity problems. The OLS estimates were used to calculate Breusch-Pagan and Jarque-Bera test statistics and as in the Cobb-Douglas models substantial heteroskedasticity and non-normality was present. Therefore the same caveats apply in interpreting the significance of estimates.

Output elasticities with respect to inputs are presented in table 5.4; these elasticities are quite close to those estimated in the Cobb-Douglas models (the input coefficients from table 5.2).

**Table 5.4: Elasticity of output with respect to inputs, translog production function**

| Variables | Model 1 | Model 3 |
|---|---|---|
| Beds | 0.32 | 0.12 |
| Labour | 0.72 | 0.59 |
| Materials | –0.10 | 0.35 |

Likelihood ratio tests were also conducted which compared the Translog form to the Cobb-Douglas functional form, by considering the Cobb-Douglas as a restricted version of the Translog. The results of these tests indicate that the more flexible Translog form is preferred to the Cobb-Douglas.

Table 5.5 compares the mean efficiency scores calculated from the MLE estimates for the whole population with mean efficiency scores for selected sub-populations. The results for efficiency scores for all models are presented to enable differences in efficiency scores to be examined in the light of minor changes in the definitions of input variables. Mean efficiency scores did not vary greatly between models though they tended to increase when revenue was used as the measure of output. They varied very little between the same models using different functional forms.

The pattern of efficiency scores across models is consistent between groups, that is, models which typically have higher efficiency scores do so for all sub-populations. Some observations to note from the results presented in table 5.5 are:

- Eefficiency scores for particular models tended to exhibit more variation between sub-populations, particularly when hospitals were classified according to size.
- Across all models, medium and large (by bed size) hospitals tended to have higher mean efficiencies.
- FP hospitals tended to have higher mean efficiencies for models 3 and 4 which use revenue as a proxy for output, suggesting that the assumption of fixed prices across hospitals may not hold. Either the use of revenue as a proxy for output is not appropriate or differences in prices charged reflect variations in the quality of outputs, implying that revenue is an accurate measure of output.

**Table 5.5: Mean efficiency scores, by model: Cobb-Douglas and translog functional forms**

| | Cobb-Douglas | | | | Translog | | | |
|---|---|---|---|---|---|---|---|---|
| Sample or sub-sample | Model 1 | Model 2 | Model 3 | Model 4 | Model 1 | Model 2 | Model 3 | Model 4 |
| Population | 0.71 | 0.72 | 0.76 | 0.79 | 0.71 | 0.72 | 0.76 | 0.79 |
| NFP hospitals | 0.70 | 0.72 | 0.72 | 0.75 | 0.70 | 0.72 | 0.72 | 0.76 |
| FP hospitals | 0.72 | 0.73 | 0.80 | 0.82 | 0.72 | 0.72 | 0.79 | 0.82 |
| Hospitals with less than 25 beds | 0.68 | 0.70 | 0.73 | 0.74 | 0.68 | 0.69 | 0.74 | 0.76 |
| Hospitals with 25 to 100 beds | 0.70 | 0.72 | 0.78 | 0.80 | 0.71 | 0.72 | 0.77 | 0.80 |
| Hospitals with more than 100 beds | 0.75 | 0.77 | 0.73 | 0.77 | 0.75 | 0.76 | 0.76 | 0.80 |

Mann-Whintey tests were conducted to see if the difference in mean efficiency score between groups were significant for each type of model. Significant differences were found between models using revenue as output, supporting earlier results.

Kruskal-Wallis test statistics were calculated for size groupings, with differences in mean efficiency for size groupings only present in Cobb-Douglas models. This may be an indication of the restrictiveness of the assumptions of non-varying returns to scale and constant substitution embedded in these models.

Table 5.6 reports the correlations between the ranks of firm efficiency scores between models for the full population. The correlation coefficients suggest that changes in the definition of output produce changes in the ranks of individual efficiency scores, though the correlation is always positive and significant. The relative efficiency of firms for particular models changes little between Cobb-Douglas and Translog specifications.

**Table 5.6: Correlations of the ranks of efficiency scores, by model**

| | Cobb-Douglas | | | | Translog | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 1 | Model 2 | Model 3 | Model 4 |
| Cobb-Douglas, model 1 | 1.00 | | | | | | | |
| Cobb-Douglas, model 2 | 0.89 | 1.00 | | | | | | |
| Cobb-Douglas, model 3 | 0.34 | 0.12 | 1.00 | | | | | |
| Cobb-Douglas, model 4 | 0.26 | 0.20 | 0.88 | 1.00 | | | | |
| Translog, model 1 | 0.94 | 0.85 | 0.34 | 0.27 | 1.00 | | | |
| Translog, model 2 | 0.87 | 0.96 | 0.15 | 0.22 | 0.89 | 1.00 | | |
| Translog, model 3 | 0.37 | 0.16 | 0.94 | 0.82 | 0.38 | 0.17 | 1.00 | |
| Translog, model 4 | 0.26 | 0.24 | 0.80 | 0.92 | 0.28 | 0.24 | 0.84 | 1.00 |

Three preliminary conclusions can be drawn from estimating production functions on the 1994–95 cross-section. Firstly, the more flexible Translog production function seems to be the more appropriate functional form. Secondly, the hypothesis of CRS is rejected for the Cobb-Douglas model. Thirdly, differences in the inputs set are not impacting greatly on the results; however, differences in the output measure adopted produce different results both in terms of the underlying production technology and for the relative efficiency scores of firms.

One aspect of estimation which was further explored was whether the population of hospitals was suitably homogenous, that is, whether the input and outputs used were able to fully capture variation in individual production behaviour. To do this models were estimated on sub-populations classified by hospital type, bed size and level of technology. Differences between hospitals were also suggested by the differences in the mean efficiencies when hospitals were disaggregated into various groups, where these differences could be a true reflection of behaviour or simply capturing firms which have different production technologies.

Statistical tests (F–tests) were calculated to test the equality of coefficients on samples disaggregated by hospital type and size. In both Cobb-Douglas and Translog functional forms the null hypothesis of the equality of coefficients was accepted for model 1 and rejected for model 3. Size disaggregation produced similar results to hospital type. The difference in these results might not lie in production technologies but in the assumptions about output measurement. Therefore perhaps a better indicator to test differences in subsets might be one which relates more directly to technology.

Statistical tests were also calculated based on the hypothesis that the population could be divided according to the level of technology inherent in the operation of an individual hospital, against the alternative that the population could be considered homogeneous. This indicator, referred to as the 'high tech' indicator, was calculated as the percentage of occupied bed days which were classified as surgical bed days (SBD). Four groups were derived representing hospitals with less than 25% of SBDs, 25 to 50%, 50 to 75% and greater than 75%. F tests calculated across these groupings rejected the null hypothesis of equality of coefficients thus indicating that different technologies were likely to be present in the population of hospitals. A model was also estimated were these 'high tech' indicators were incorporated into regressions as dummy variables. The results of these regressions are presented in table 5.7.

An interesting result which arose out of these regressions was the differences in the sign on the 'high tech' dummies between models with different output types, where the excluded dummy was 'greater than 75% of SBDs'. In models which use occupied bed days as measure of output the sign on the dummy variables was positive suggesting that output increases with an increase in technology. However, the output measure based on revenue suggests that output decreases with an increase in technology as the signs on the 'high tech' dummies was negative (but admittedly insignificant).

**Table 5.7: Estimated regressions using 'High Tech' indicators, models 1 and 3**

| | Cobb-Douglas | | | | Translog | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1 | | Model 3 | | Model 1 | | Model 3 | |
| **Coefficients** | MLE | SEs | MLE | SEs | MLE | SEs | MLE | SEs |
| Constant | 5.02 | 0.50 | 7.37 | 0.35 | 29.55 | 4.35 | 11.36 | 4.36 |
| Beds | 0.23 | 0.06 | 0.13 | 0.05 | 0.55 | 1.08 | 1.92 | 0.96 |
| Materials | 0.05 | 0.04 | 0.38 | 0.03 | –4.95 | 0.89 | –0.80 | 0.88 |
| Labour | 0.69 | 0.06 | 0.56 | 0.06 | 5.41 | 1.13 | 1.03 | 1.25 |
| Beds squared | | | | | –0.06 | 0.19 | 0.03 | 0.20 |
| Labour squared | | | | | 0.27 | 0.15 | –0.04 | 0.17 |
| Materials squared | | | | | 0.51 | 0.09 | 0.15 | 0.09 |
| Beds x labour | | | | | 0.18 | 0.12 | 0.14 | 0.10 |
| Beds x materials | | | | | –0.05 | 0.11 | –0.18 | 0.11 |
| Labour x materials | | | | | –0.47 | 0.12 | –0.06 | 0.13 |
| High tech 1 (<25% of SBD's are surgical) | 0.50 | 0.09 | -0.06 | 0.06 | 0.47 | 0.08 | –0.02 | 0.08 |
| High tech 2 (25 to 50% of SBD are surgical) | 0.13 | 0.08 | -0.09 | 0.06 | 0.12 | 0.09 | –0.06 | 0.07 |
| High tech 3 (50 to 75% of SBD are surgical) | 0.08 | 0.08 | -0.05 | 0.06 | 0.08 | 0.08 | –0.01 | 0.07 |
| **Diagnostics** | | | | | | | | |
| $\sigma u/\sigma v$ | 2.47 | 0.30 | 3.60 | 0.55 | 2.99 | 0.53 | 4.78 | 0.99 |
| $\sqrt{(\sigma^2 u+\sigma^2 v)}$ | 0.49 | 0.01 | 0.42 | 0.01 | 0.49 | 0.01 | 0.41 | 0.01 |
| Log likelihood | –80.00 | | –8.00 | | –69.00 | | 5.00 | |

### Panel results

Cobb-Douglas and translog production functions were estimated for all models presented in table 5.1 on the balanced panel of data which contained 280 hospitals. This balanced panel was preferred to an unbalanced panel to enhance the comparisons with the DEA results presented in section 4. However, estimates were also derived for the unbalanced panel to check for the consistency of results between the two datasets. The dollar-based materials and revenue measures were deflated so as to convert them to real or volume-based measures allowing comparisons over time. A brief discussion of the results follows; further information and tabulations of model results can be obtained from the authors upon request.

The focus was on a stochastic frontier model which assumes time invariant inefficiencies. This was done for two reasons, firstly because the length of the panel is short and secondly because we hoped not to confound the time trend capturing productivity change with that capturing efficiency change. Some time varying models were estimated, with the results of these models not differing substantially from the time invariant models.

Insignificant efficiency time trend coefficients on most models also indicated that there had been little variation in efficiency scores over time and that perhaps a more simplified (time invariant) model might apply.[45]

For the Cobb-Douglas production frontiers, MLE was a significant improvement over OLS for all models. The coefficients on the inputs in each model are broadly consistent with those obtained in the cross-section except for the coefficient on materials which was positive for model 1 though statistically insignificant. The coefficient on the 'time' variable which gives an indication of productivity change is close to zero for all models indicating little or no productivity change in this period. This result is not unexpected given the relatively short time period over which these models have been estimated. In all models the hypothesis of CRS was rejected. The sum of coefficients points to DRS for model 1 and IRS for model 3 as is the case for the cross-section results.

The results of the Translog estimation tend to mirror those estimated on the cross-section. Likelihood ratio test statistics strongly favour the estimation of a frontier-based model over an OLS estimated model. The insignificant time trend coefficients point to little or no productivity change over this period. Output elasticities are similar to those obtained from cross-sectional analysis. Likelihood ratio tests were conducted to compare the Translog and Cobb-Douglas functional forms. In all models the null hypothesis that there is no difference between the restricted (Cobb-Douglas) and unrestricted Translog models was rejected.

Mean efficiency scores were calculated for all models and the same general conclusions could be drawn as from the cross-sectional analysis, that is, models using revenue as the output measure tend to have higher mean efficiency scores and larger hospitals for model 1 tended to have higher efficiency scores. All models exhibited differences in mean efficiency when disaggregated according to size.

The production functions estimated assume that productivity change is Hicks neutral, that is, productivity change shifts the production frontier out in a parallel manner. However, it is quite possible that productivity change will favour one input over another, a situation that is often called biased technological change. It is possible to test for Hicks neutrality by including interactive terms between the time trend and each of the three inputs in the production function, and testing for their joint significance using an F test. The null hypothesis of Hicks neutral technological change is rejected for model 1 and accepted for model 3. An implication of this is that the estimated coefficients differ significantly across time for model 1 and so we probably should not be pooling the four time periods together, but rather estimating each individual time period separately.[46]

**Summary of production function estimation**

Information on the structure of this industry did not vary significantly between that obtained from the cross-section of data and that obtained from the panel though this is not unexpected given the relatively short time period of the panel. Estimating Cobb-Douglas and Translog production functions does not produce vastly different coefficients (or elasticities with respect to output) though the more flexible Translog function is usually the preferred functional form. When different input sets (measures of inputs) are used there is also little difference in results. However, different output proxies do produce very different results in terms of output elasticities and correlations between the ranks of efficiency scores between different models.

The question of which output proxy, revenue or occupied bed days, is the better measure has not been answered by this analysis. The correct signing of all input coefficients in models which were estimated using revenue as a measure of output and the negative sign on the materials

[45] Lovell (1996, p. 334) points out that it may be difficult to separately identify neutral efficiency change common to all firms and neutral productivity change common to all firms.

[46] Due to potentially severe multicollinearity problems we did not test this hypothesis for the Translog functional form.

coefficient on models using occupied bed days as measure of output could suggest that revenue is a more appropriate measure. However, there are alternative explanations for why we obtained a negative coefficient on the materials input when using occupied bed day which relate to the measurement of the materials input, in particular, the fact that it is measured as a value and not a volume, in contrast to output. It can be shown that if volume discounts are available in the purchase of inputs and a simple markup model for pricing is assumed we can produce similar results to those obtained when using the two different output measures. This suggests that the assumption that prices are fixed across hospitals is violated and that the occupied bed day measure of output is more appropriate. This analysis has as yet not resolved these issues.[47]

[47] The choice between revenue and OBS as a proxy for output, where output is assumed to be truly measured by case weighted separations, will also depend on the inter-hospital variation in costs associated with the non-bed day component on clinical costs.

## 5.2 Cost function estimation

This section discusses the results of estimating Cobb-Douglas and Translog cost functions. Given the similarities in results between production function estimation of models using different input sets and because of the limitations imposed by the data in obtaining true volume measures of inputs from which average input prices could be calculated, two cost function models with two different output proxies were estimated (see table 5.8).

Cost functions are often estimated so that multiple outputs can be incorporated into the frontier estimation framework. Unfortunately this paper could not take advantage of this feature of the cost function as for many firms certain outputs were not produced and thus their output could not be logged in preparation for estimation. Battese (1996) discusses a dummy variable method for dealing with zeros in inputs and this method could possibly be extended to zeros in measured outputs, however this technique has not been applied to cost functions for the analysis in this paper.[48]

Despite the limitations of incorporating additional information about outputs the cost function still has a number of advantages over the production function including the ability to obtain estimates of technical and allocative efficiency and to better represent firm behaviour if hospitals are cost minimisers. Breyer (1987) discusses the types of cost functions that have usually been estimated on hospitals, noting though the two strands of analysis that have developed in the literature both contain inadequacies.

This paper is primarily interested in calculating changes in productivity and technical efficiency and therefore has its basis in production economics. Thus the cost functions estimated in this paper are designed to reveal as much information as possible about the structure of production and changes in technical efficiency and productivity over time. Skinner(1994) discusses the conditions under which stochastic frontier cost functions can produce misleading results, for example, when the symmetric error term is skewed. Future iterations of this paper will examine whether these conditions are present in this analysis.

**Table 5.8: Cost function model specifications**

| Variables | Model 1 | Model 3 |
|---|---|---|
| **Input prices** | | |
| Beds | X | X |
| Materials II (including VMOs) | X | X |
| Total staff II (labour costs) | X | X |
| **Outputs** | | |
| Revenue | | X |
| Composite output I (occupied bed days) | X | |

Estimates of the Cobb-Douglas cost function are presented in table 5.9. The two MLE estimates in the table represent an unrestricted model and a model which imposes linear homogeneity in prices.

The results of model 1 (which uses occupied bed days as a measure of output) were as expected, though the insignificant coefficient on the price of labour is surprising. The coefficient on output suggests that CRS prevails in this model.

[48] Other means of addressing this problem include using an alternative functional form (e.g. quadratic functions), adopting the Box-Cox transformation as in Caves, Christensen and Trethaway (1980), or splitting the sample into different hospital types, according to the outputs produced.

One-sided likelihood ratio tests whilst significant, given a critical value of 2.71, are lower or suggest a smaller improvement over the OLS estimated models than that suggested by the production function models or cost function model 3.

The results obtained from estimating model 3 (which uses revenue as a measure of output) are more disturbing, particularly given the negative coefficient on price of labour and what looks like a rather unlikely scenario that these price coefficients sum to one (i.e. linear homogeneity in prices). In fact the negative coefficient violates the assumption of a non-decreasing effect on costs in prices though in these model linear homogeneity has not been *a priori* imposed.[49]

It is also interesting to note that, when estimating cost functions, model 3 suggests decreasing returns to scale and model 1 increasing returns to scale, the opposite result to that obtained from production function estimation using corresponding measures of output.

**Table 5.9: Cobb-Douglas cost function estimation (n=280)**

| | Model 1 | | | | | | Model 3 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Variables** | OLS | SEs | MLE | SEs | MLE (LH)(a) | SEs | OLS | SEs | MLE | SEs | MLE (LH) | SEs |
| Constant | 1.97 | 0.77 | 1.47 | 0.78 | −1.48 | 0.14 | 2.80 | 0.96 | 1.88 | 0.86 | −8.10 | 0.26 |
| Output | 1.00 | 0.01 | 1.01 | 0.01 | 1.01 | 0.01 | 0.88 | 0.01 | 0.91 | 0.01 | 0.86 | 0.01 |
| Capital price | 0.05 | 0.01 | 0.05 | 0.01 | 0.06 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.09 | 0.01 |
| Labour price | 0.08 | 0.07 | 0.10 | 0.08 | | | −0.16 | 0.09 | −0.12 | 0.08 | | |
| Materials price | 0.57 | 0.02 | 0.57 | 0.01 | 0.56 | 0.01 | 0.08 | 0.02 | 0.05 | 0.01 | 0.07 | 0.02 |
| **Diagnostics** | | | | | | | | | | | | |
| su/sv | | | 1.35 | 0.22 | 1.52 | 0.23 | | | 3.60 | 0.58 | 2.99 | 0.39 |
| $\sqrt{(s^2u+s^2v)}$ | | | 0.21 | 0.02 | 0.22 | 0.02 | | | 0.30 | 0.01 | 0.37 | 0.01 |
| Adj $R^2$ | 0.98 | | | | | | 0.97 | | | | | |
| Log likelihood | 112.00 | | 110.00 | | 106.00 | | 48.00 | | 86.00 | | 15.00 | |

(a) Linear homogeneity in prices has been imposed in estimating this function.

Likelihood ratio tests were conducted to test the restrictions of linear homogeneity in prices and CRS (assuming linear homogeneity in prices). Linear homogeneity in prices is routinely imposed when estimating cost functions in order for the cost function to be theoretically sensible and from which information on the production technology can be derived. Whilst it makes sense to impose the restriction of linear homogeneity in prices from a theoretical perspective, this restriction is not supported by the data. The hypothesis of CRS (tested against linear homogeneity) was accepted for model 1 but rejected for model 3.

Translog cost functions were also estimated and results of these estimations are presented in table 5.10. As in the production function the likelihood of multicollinearity makes it difficult to interpret the significance of the coefficients. The one-sided likelihood ratio tests are also more difficult to interpret as the number of stochastic frontier parameters being estimated has increased. In this model the more general truncated normal distribution with a non-zero mean had to be assumed for the inefficiency error term in order to get the model to converge upon an appropriate solution.

Likelihood ratio test statistics were calculated to test linear homogeneity in prices, and were rejected for all models. CRS (when linear homogeneity is assumed) was also tested, and the hypothesis accepted for model 1 and rejected for model 3.

[49] The rejection of homogeneity and the incorrect signs could occur for many reasons: cost minimisation behaviour does not exist; it does exist but is on an inter-temporal planning horizon; or it does exist but is subject to certain constraints; or there are measurement problems with some prices.

The Translog functional form is also preferred to the Cobb-Douglas functional form.[50] Tests were also performed to test for input separability, and rejected for both models.

**Table 5.10: Translog cost function estimation**

| Variables | Model 1(a) | | | | | | Model 3 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OLS | SEs | MLE | SEs | MLE (LH)(b) | SEs | OLS | SEs | MLE | SEs | MLE (LH) | SEs |
| Constant | −13.4 | 24.81 | −13.47 | 34.62 | −1.26 | 1.00 | 18.26 | 32.95 | 18.08 | 25.26 | −9.36 | 3.34 |
| Output | −0.79 | 0.90 | −0.79 | 1.17 | 1.06 | 0.15 | 1.26 | 1.05 | 1.57 | 0.83 | 0.75 | 0.30 |
| Capital price | −0.36 | 0.80 | −0.37 | 0.91 | −0.15 | 0.09 | −1.45 | 1.06 | −1.58 | 0.77 | −0.14 | 0.19 |
| Labour price | 2.75 | 4.63 | 2.75 | 6.56 | | | -2.03 | 6.13 | −2.28 | 4.75 | | |
| Materials price | 5.31 | 1.26 | 5.31 | 1.39 | 0.74 | 0.17 | -0.85 | 1.77 | −1.19 | 1.26 | −0.59 | 0.45 |
| Output squared | −0.01 | 0.02 | −0.01 | 0.01 | −0.01 | 0.01 | 0.04 | 0.01 | 0.03 | 0.01 | 0.03 | 0.01 |
| Capital price sq. | 0.03 | 0.01 | 0.03 | 0.01 | 0.03 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 |
| Labour price sq. | −0.20 | 0.45 | −0.20 | 0.64 | | | 0.18 | 0.59 | 0.23 | 0.46 | | |
| Material price sq. | 0.02 | 0.03 | 0.02 | 0.03 | 0.02 | 0.03 | −0.03 | 0.04 | −0.07 | 0.03 | 0.00 | 0.03 |
| Capital x labour(c) | 0.01 | 0.08 | 0.01 | 0.09 | | | 0.12 | 0.10 | 0.12 | 0.08 | | |
| Capital x materials(c) | −0.02 | 0.01 | −0.02 | 0.02 | −0.02 | 0.02 | 0.00 | 0.02 | 0.01 | 0.02 | −0.02 | 0.03 |
| Labour x materials(c) | −0.43 | 0.12 | −0.43 | 0.13 | | | 0.10 | 0.17 | 0.11 | 0.12 | | |
| Output x capital(c) | 0.03 | 0.01 | 0.03 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| Output x labour(c) | 0.16 | 0.08 | 0.16 | 0.11 | | | −0.10 | 0.10 | −0.12 | 0.08 | | |
| Output x materials(c) | −0.01 | 0.02 | −0.01 | 0.02 | −0.01 | 0.02 | 0.00 | 0.02 | 0.02 | 0.02 | 0.04 | 0.03 |
| **Diagnostics** | | | | | | | | | | | | |
| Ø/$\sigma$u(d) | | | −0.07 | 10.26 | −0.62 | 2.61 | | | −3.71 | 2.45 | −0.96 | 0.89 |
| $\sigma$u/$\sigma$v | | | 0.69 | 0.72 | 1.23 | 0.28 | | | 10.74 | 2.52 | 3.12 | 0.43 |
| $\sqrt{(\sigma^2 u+\sigma^2 v)}$ | | | 0.16 | 0.10 | 0.21 | 0.05 | | | 0.72 | 0.18 | 0.43 | 0.05 |
| Adj $R^2$ | 0.98 | | | | | | 0.97 | | | | | |
| Log likelihood | 145.00 | | 144.00 | | 127.00 | | 66.00 | | 135.00 | | 38.00 | |

(a) The estimation results presented in this table use a truncated normal distribution for the efficiency error term.
(b) Linear homogeneity in prices has been imposed in estimating the frontier cost function.
(c) These terms represent the combinations of interactions between the variables (both output and inputs).
(d) P represents the mean estimated error of the truncated normal distribution.

Mean cost efficiency scores for the various sub-populations were calculated for both the Cobb-Douglas and Translog functional forms and are presented in table 5.11. NFP hospitals appear to be less cost efficient than FP hospitals, whilst medium size hospitals appear to be the most cost efficient in the size groupings.

[50] A test which was applied by Zuckerman, Hadley and Iezzoni (1994) but which as yet has not be applied in the same manner in this paper is whether output is a truly exogenous variable.

**Table 5.11: Mean cost efficiency scores, by model and sub-population**

| Sample or sub-sample | Cobb-Douglas Model 1 | Cobb-Douglas Model 3 | Translog Model 1 | Translog Model 3 |
|---|---|---|---|---|
| Whole population | 1.16 | 1.30 | 1.04 | 1.21 |
| NFP hospitals | 1.19 | 1.49 | 1.05 | 1.34 |
| FP hospitals | 1.14 | 1.32 | 1.04 | 1.24 |
| Hospital with less than 25 beds | 1.19 | 1.61 | 1.05 | 1.43 |
| Hospital with 25 beds to 100 beds | 1.16 | 1.33 | 1.04 | 1.23 |
| Hospitals with more than 100 beds | 1.16 | 1.43 | 1.04 | 1.32 |

Mann-Whitney and Kruskal-Wallis test statistics were derived and the results supported earlier comments regarding which hospitals are more cost-efficient, though these results are not entirely consistent with those obtained from the production function. The mean efficiencies of different size groups produced statistically different mean efficiencies when estimating a production function and this was not the case for cost function estimation.

The correlation coefficients in table 5.12 indicate that the ranks of efficiency scores for models using different proxies for output are only mildly positively correlated whilst functional form seems to make little difference (though more than production functions) to the relative ranks of hospitals.

**Table 5.12: Correlation of ranks of hospitals, by efficiency score and model**

| Time invariant | Cobb-Douglas Model 1 | Cobb-Douglas Model 3 | Translog Model 1 | Translog Model 3 |
|---|---|---|---|---|
| Cobb-Douglas, model 1 | 1.00 | | | |
| Cobb-Douglas, model 3 | 0.36 | 1.00 | | |
| Translog, model 1 | 0.92 | 0.33 | 1.00 | |
| Translog, model 3 | 0.33 | 0.92 | 0.38 | 1.00 |

Cost function models were also estimated using the 'high tech' indicator dummies constructed for the production function estimation. The 'high tech' dummy variables were usually significant contributors to the regression, indicating that costs were in part explained by the additional variables, as shown is table 5.13.

**Table 5.13: 'High tech' regressions**

| Variables | Model 1 OLS | SEs | Model 3 MLE | SEs |
|---|---|---|---|---|
| Constant | 1.61 | 0.79 | 1.58 | 0.91 |
| Output | 1.00 | 0.01 | 0.92 | 0.01 |
| Capital price | 0.05 | 0.01 | 0.05 | 0.01 |
| Labour price | 0.12 | 0.08 | –0.12 | 0.09 |
| Materials price | 0.53 | 0.02 | 0.08 | 0.02 |
| High tech group 1 | –0.12 | 0.04 | 0.10 | 0.05 |
| High tech group 2 | –0.06 | 0.04 | 0.07 | 0.05 |
| High tech group 3 | –0.02 | 0.03 | 0.04 | 0.04 |
| Diagnostics | | | | |
| $\sigma u/\sigma v$ | 1.48 | 0.22 | 3.31 | 0.55 |
| $\sqrt{(\sigma^2 u+\sigma^2 v)}$ | 0.21 | 0.02 | 0.29 | 0.01 |
| Log likelihood | 119.00 | | 88.00 | |

The differences in sign on the 'high tech' dummies between models is another example of the impact of using different measures of output.

A possible extension to this analysis involves using systems estimation to estimate the translog cost function jointly with the cost share equations. Preliminary results for the average cost function (rather than the frontier) suggest that systems estimation may go a long way to overcoming multicollinearity problems and substantially improve the efficiency of estimates. For example, standard errors were markedly lower than in equivalent models estimated using single equation OLS estimation, and consequently the results of hypothesis tests were far more conclusive. However, extending systems estimation to the case of a stochastic frontier translog cost function is more complicated, and has not yet been undertaken.

**Estimation using a panel of data**

Cost functions were estimated on the panel of data and as in the production function time invariant versions of the stochastic frontier cost function were estimated.

The estimated coefficients for the Cobb-Douglas cost function were similar to those obtained in the cross-section. Interestingly the coefficient on output fell in both models, pointing to decreasing returns to scale in both models (previously it was just model 3). Both models produced insignificant coefficients on the time trend. The one-sided likelihood ratio test produced large test statistics for both models pointing to the superiority of the stochastic frontier models over the OLS models. This was not the case in the cross-section where model 1 did not appear to be as large an improvement over the OLS model. Once again OLS estimates for model 3 produced disturbing results with the coefficient on the price of labour being negative. Linear homogeneity was rejected in both models, as was CRS.

Mean efficiencies were calculated and, as for the cross-section results, FP hospitals and hospitals with 25 to 100 beds appeared to be the most cost efficient groups.

### Summary of cost function estimation

As for the production function estimation, models using different measures of output produce quite different results in terms of implied productivity change over time, efficiency ranks of firms and estimated coefficients on price variables. One overriding feature of all models was their relatively poor performance in terms of supporting the standard assumptions underlying well-behaved cost functions, for example, linear homogeneity in prices was not accepted as a restriction in any of the models.

### Comparing production and cost function estimation

The performance of both cost and production functions was poor in terms of estimating theoretically defensible models. In production function estimation, we found, in the cross-section, that models using occupied bed days as a measure of output produced negative elasticities on the materials variables; whilst in estimating cost functions we found that models using revenue as a measure of output produced a negative coefficient on the price labour (when linear homogeneity was not imposed). These results clearly indicate that all the models are not effectively capturing the economic behaviour of these hospitals. It may be the case that standard economic explanations of behaviour upon which our models are based are inadequate representations of these hospitals' behaviour.

One comparison that can be performed between Cobb-Douglas cost and production functions is to compare the derived input coefficients, representing input elasticities.

**Table 5.14: Comparison of Cobb-Douglas elasticities**

| | Production functions | | Cost functions | |
|---|---|---|---|---|
| Input type | Model 1 | Model 3 | Model 1 | Model 3 |
| Beds | 0.22 | 0.12 | 0.06 | 0.10 |
| Materials | -0.11 | 0.39 | 0.56 | 0.08 |
| Labour | 0.78 | 0.56 | 0.38 | 0.98 |

Direct and indirect estimation of the production function produce very different elasticities, both in terms of size and, in some cases, sign. The most appropriate mechanism by which to estimate the production function depends on the quality of the underlying data and the economic behaviour of hospitals.

### Comparing and decomposing efficiency scores

To further the comparison between efficiency scores produced by the cost functions and production functions we intended to apply the Kopp and Diewert technique for decomposing cost efficiencies. As yet we have not successfully implemented the technique.

Table 5.15 shows the correlations of individual ranks for hospitals corresponding production and cost model estimations (comparing individual cost efficiencies and technical efficiencies). The comparison between efficiency scores is based on the 1994–95 data upon which cost functions were estimated (280 hospitals).

The production and cost function efficiency scores are well correlated between like models pointing at least to consistent estimation of efficiency rankings between the two approaches.

**Table 5.15 : Correlations of efficiency score rankings, by model: cost functions**

| | Production functions | | | |
|---|---|---|---|---|
| | Cobb-Douglas | | Translog | |
| | Model 1 | Model 3 | Model 1 | Model 3 |
| Cobb-Douglas, model 1 | 0.71 | 0.23 | 0.72 | 0.26 |
| Cobb-Douglas, model 3 | 0.21 | 0.85 | 0.20 | 0.84 |
| Translog, model 1 | 0.70 | 0.23 | 0.70 | 0.26 |
| Translog, model 3 | 0.23 | 0.79 | 0.23 | 0.86 |

# 6 Comparison of results

One important extension of the analysis is to make informative comparisons between results from the estimation techniques.

The important area of comparison is between the DEA technique and SFA estimation. Both of these results produce technical efficiency scores and, although the scores cannot be directly compared in terms of their *levels*,[51] judgements can be made about the *rankings* of individual hospitals. Information that the two techniques provide on scale and the estimates of aggregate productivity growth can also be examined.

Two methods of comparing the results of efficiency scores are rank correlation analysis and frequency tables:

- The former involves ranking observations by DEA and SFA technical efficiency and using a test statistic (based on the squared differences in ranks for each observation) to calculate the sample correlation between the two sets of ranks. This will give a measure of how closely matched two sets of results are, with a high correlation indicating that the two techniques tend to rank observations in a similar order.

- A frequency table provides information about the frequency in which high, medium or low efficiency observations in one results set are ranked as high, medium or low efficiency in the other results set. In fact, it is possible to derive a correlation coefficient based on the frequencies in each cell of the frequency table. Banker, Conrad and Strauss (1986) suggest a number of different comparison proxies for these frequency tables, including not only DEA technical and pure technical efficiency scores and SFA technical efficiencies but also capacity utilisation.[52] The important feature of the table is the strength of the diagonal elements compared to the off-diagonal elements, particularly the off-diagonal corners representing the extreme combinations of high and low ranked observations in each model. These cells should contain few (or zero) observations for a pair of positively correlated samples.

As an illustration, table 6.1 presents the rank correlations and frequency tables comparing a set of DEA pure technical efficiency scores with a set of results from an SFA frontier in which CRS is rejected. The results, estimated using DEA model 1 (in table 4.1) in VRS form and stochastic frontier model 1 in Translog form (from table 5.1), indicate a strong correlation in ranks between the techniques (63.5%), with the frequency table being fairly strongly diagonal. On the same table, the results of a comparison with Translog model 3 are also reported. The correlation in this case is not as strong (47%), indicating that the results of DEA model are more comparable with stochastic frontier model 1. This is to be expected since the definition of output between DEA model 1 and Translog model 1 is the same (patient numbers as opposed to patient revenue in model 3).

Similar results are presented for the comparison between DEA models 9 and 11 with Translog models 1 and 3 (tables 6.2 and 6.3 respectively). The highest correlations occur between DEA model 9 and Translog model 1 (correlation of 73.6%) and DEA model 11 and Translog model 3 (correlation of 75.5%). This is not unexpected since these sets of models use the same input-output set to generate the respective sets of results. The correlations between the alternative combinations (DEA model 9 and Translog model 3 and DEA model 11 and Translog model 1) are both very low, being around 25%. These results are reported in tables 6.2 and 6.3.

---

[51] DEA efficiency scores are calculated relative to the sample and have little meaning when comparing results between DEA studies. On the other hand, the efficiency rankings of observations in DEA studies contain useful information which can be used for comparison purposes.

[52] Capacity utilisation is defined as the ratio of inpatient days (occupied bed days) to total available beds per year.

**Table 6.1: Comparing DEA and SFA results (frequency table), DEA PTE vs Translog SFA**

| Frequency<br>DEA PTE scores, Model 1 | Translog scores *vs* DEA PTE scores (a)(b)(c)<br>Translog TE scores Model 1 \| Model 3<br>Group 1 (Rank: 1–91) | | Group 2 (Rank: 92–176) | | Group 3 (Rank: 177–243) | | Group 4 (Rank: 244–300) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| Translog model # | 1 | 3 | 1 | 3 | 1 | 3 | 1 | 3 | |
| Group 1 (PTE=1) | 59 | 51 | 20 | 22 | 6 | 9 | 6 | 9 | 91 |
| Group 2 (PTE: 0.85–1) | 24 | 30 | 40 | 27 | 16 | 17 | 5 | 11 | 85 |
| Group 3 (PTE 0.65–0.85) | 7 | 7 | 18 | 25 | 33 | 25 | 9 | 10 | 67 |
| Group 4 (PTE<0.65) | 1 | 3 | 7 | 11 | 12 | 16 | 37 | 27 | 57 |
| Total | 91 | | 85 | | 67 | | 57 | | 300 |

(a) Spearman sample rank correlations(st. normal deviate): *vs* Model 1 = 0.652(11.28); *vs* Model 3 = 0.459(7.94)
(b) $\chi^2$ statistic($p$-value): *vs* model 1 = 188.72(0.001); *vs* model 3 = 89.96 (0.001)
(c) Pearson (table) correlations: *vs* model 1 = 0.625; *vs* model 3 = 0.471

**Table 6.2: Comparing DEA and SFA results (frequency table), DEA PTE vs Translog SFA**

| Frequency<br>DEA PTE scores, Model 9 | Translog scores *vs* DEA PTE scores (a)(b)(c)<br>Translog TE scores Model 1 \| Model 3<br>Group 1 (Rank: 1–75) | | Group 2 (Rank: 76–150) | | Group 3 (Rank: 151–225) | | Group 4 (Rank: 226–300) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| Translog model # | 1 | 3 | 1 | 3 | 1 | 3 | 1 | 3 | |
| Group 1 (Rank 1–75) | 47 | 31 | 25 | 20 | 2 | 16 | 1 | 8 | 75 |
| Group 2 (Rank 76–150) | 20 | 23 | 26 | 18 | 27 | 19 | 2 | 15 | 75 |
| Group 3 (Rank 151–225) | 8 | 10 | 21 | 19 | 28 | 19 | 18 | 27 | 75 |
| Group 4 (Rank 226–300) | 0 | 11 | 3 | 18 | 18 | 21 | 54 | 25 | 75 |
| Total | 75 | | 75 | | 75 | | 75 | | 300 |

(a) Spearman sample rank correlations(*st. normal deviate*): *vs* Model 1 = 0.793(13.71); *vs* Model 3 = 0.323(5.6)
(b) $\chi^2$ statistics($p$-value): *vs* Model 1 = 207.2 (0.001); *vs* Model 3 = 29.71 (0.285)
(c) Pearson (table) correlations: *vs* Model 1 = 0.736; *vs* Model 3 = 0.285

**Table 6.3: Comparing DEA and SFA results (frequency table), DEA PTE vs Translog SFA**

| Frequency<br>DEA PTE scores, Model 11 | Translog scores *vs* DEA PTE scores (a)(b)(c)<br>Translog TE scores Model 1 \| Model 3<br>Group 1 (Rank: 1–75) | | Group 2 (Rank: 76–150) | | Group 3 (Rank: 151–225) | | Group 4 (Rank: 226–300) | | Total |
|---|---|---|---|---|---|---|---|---|---|
| Translog model # | 1 | 3 | 1 | 3 | 1 | 3 | 1 | 3 | |
| Group 1 (Rank 1–75) | 27 | 46 | 28 | 21 | 12 | 8 | 8 | 0 | 75 |
| Group 2 (Rank 76–150) | 20 | 23 | 22 | 31 | 12 | 21 | 12 | 1 | 75 |
| Group 3 (Rank 151–225) | 13 | 5 | 12 | 21 | 24 | 35 | 26 | 14 | 75 |
| Group 4 (Rank 226–300) | 15 | 1 | 13 | 2 | 18 | 12 | 29 | 60 | 75 |
| Total | 75 | | 75 | | 75 | | 75 | | 300 |

(a) Spearman sample rank correlations(*st. normal deviate*): *vs* Model 1 = 0.334(5.78); *vs* Model 3 = 0.80(13.83)
(b) $\chi^2$ statistics($p$-value): *vs* Model 1 = 36.75 (0.001); *vs* Model 3 = 241.33 (0.001)
(c) Pearson (table) correlations: *vs* Model 1 = 0.291; *vs* Model 3 = 0.755

Indirect comparisons such as those proposed by Banker, Conrad and Strauss(1986), for example using capacity utilisation, provide very similar results to those from a direct comparison and are not reported in this paper.

The results presented in tables 6.1 to 6.3 indicate that efficiency scores generated from the DEA and stochastic frontier approaches, using the same input-output set for each model, are very comparable in terms of the tendency for similar firms to be ranked highly or lowly in each model.

The results also indicate that changing input and/or output definitions (or even including more variables in the model) very quickly reduces the correlations for individual rankings between the techniques, to the point where the hypothesis of some correlation between the efficiency rankings can be rejected in some cases.

In terms of the changes in efficiency and productivity for hospitals in the dataset over the four years of the sample, the techniques give very different observations. This is primarily due to the vastly different methodologies used to define and measure productivity growth for the parametric and non-parametric techniques. For instance, the results of the Malmqvist productivity analysis for the four-year period indicate a significant increase in overall productivity (comprising increases in efficiency and technology). The results indicate that most of the gains are made in the first period (1991–92 to 1992–93) as a result of a large technical change in the period. On the other hand, the results obtained from a parametric type analysis of productivity trend (using a panel of data and including a time trend and possibly time-related errors) indicate that the changes in productivity over the period are insignificant.

It is also useful to note that the results obtained by changing variables sets are very similar to those described above. In the case of DEA, whilst the levels of the indices calculated vary quite considerably as a result of changes in variable sets and definitions, the overall movements in the indices remain the same, indicating that an underlying trend appears to be reflected by the data, independent of the exact model used.

It may also be the case that the period being studied is too short to draw any meaningful conclusions about overall trends in productivity, particularly in the case of the parametric stochastic frontier technique.

# 7 Conclusions and possible extensions

The purpose of this study was to examine the performance of different techniques for measuring productivity and to gain a better understanding of the Australian health sector through understanding the operations of private hospitals. Both the parametric and non-parametric techniques provided insights into this sector but perhaps more importantly for this study demonstrated their own strengths and weaknesses in measuring efficiency and productivity

An important conclusion from the DEA analysis was that hospital efficiency scores were not robust to changes in the sets of inputs and outputs. While this was expected, we were surprised to find that sometimes even small changes in input sets can produce very different results, particularly when outputs are disaggregated. Given that the DEA results are sensitive to the choice of inputs and outputs, it is necessary to provide sound reasons for nominating one model rather than another and (as far as possible) to explain just how different model specifications can lead to different conclusions. Overall, technical efficiency appeared to be only marginally influenced by factors such as hospital type (profit-making status) or scale, even though the majority of hospitals appear to operate under decreasing returns to scale.

Despite their immaturity, the parametric analyses have also produced a number of interesting results. Perhaps the most emphatic is that modelling with the two different measures of output (occupied bed days and deflated revenue) will produce very different results in terms of both the structure of production and the relative efficiencies of hospitals. Preliminary OLS and SFA analysis on sub-samples (characterised by size, profit-making status and degree of high technology) point to the likelihood that individual hospitals' activities vary substantially; it is unclear whether the full population of hospitals are so disparate that effective modelling may not be possible, and the population may have to be dissected into sub-industries. However, one of the purposes of this exercise was to obtain aggregate results for the private hospitals segment of the health industry and it is difficult to serve this purpose by breaking it into sub-populations.

It is also clear that the dataset on which these models have been estimated (rich though it is) is not rich enough to effectively characterise the industry using standard economic models. As well, the difficulties in incorporating all the types of variables in the dataset into the various techniques have also contributed to the ambiguity of some findings. In particular, analysis of the production structure and efficiency of Australian private hospitals would benefit from improved capital stock estimates (e.g. through collection of asset data) and a more detailed disaggregation of hospital activity (e.g. linking of DRG data to the PHEC dataset).

## 7.1 Extensions

To improve the analyses, possible extensions are suggested below. Any comment on their worth would be appreciated:

- resolve issues in decomposing cost inefficiencies and apply techniques for measuring allocative and technical efficiency in SFA analysis;
- attempt to quality-adjust our output measures;
- incorporate directly into the DEA analysis measures of quality from both a cross-section and time perspective; and
- extend the DEA technique to allow for a stochastic element in the data. For example, the authors are in the process of applying methods to identify influential outliers in DEA (using a modifications suggested by Wilson 1995 and Lovell, Walters &Wood 1993), as well as implementing a number of DEA resampling techniques (including Ferrier & Hirschberg 1997 and Atkinson & Wilson 1995).

# Appendix 1: Variable descriptive statistics

Table A1.1 presents means and standard errors for the variables defined in table 1 for the entire population, encompassing all private acute care hospitals in Australia for the 1994–5 financial year.

**Table A1.1: Descriptive statistics, inputs and outputs: all models**

| Variable | Mean | Standard deviation |
|---|---|---|
| SMO (FTE) | 1.14 | 2.70 |
| VMO ($) | 79 845.30 | 287 745.81 |
| Nursing staff (FTE) | 65.06 | 86.89 |
| Other staff (FTE) | 50.56 | 72.25 |
| Total staff (FTE) | 116.76 | 161.06 |
| Beds (no.) | 69.57 | 62.85 |
| Materials ($) | 2 732 616.72 | 3 952 263.10 |
| Total admissions (no.) | 5 778.79 | 8 720.70 |
| Total labour costs ($) | 4 780 787.56 | 6 546 420.19 |
| Acute care inpatient days | 6 682.00 | 6 757.65 |
| Accident and emergency treatments | 454.82 | 1 969.77 |
| Non-inpatient occasions of service | 2 974.50 | 14 356.80 |
| Nursing home type days | 650.91 | 3 970.48 |
| Surgical procedures | 3 561.34 | 4 412.73 |
| Surgical inpatient days | 9 476.38 | 13 112.79 |
| Surgeries | 4 134.21 | 5 241.58 |
| Composite inpatient separations | 4 336.25 | 5 012.96 |
| Composite output I | 17 561.52 | 19 597.34 |

Table A1.2 presents correlation coefficients for some of the key variables used in parametric analysis.

**Table A1.2: Correlation coefficients, key variables: SFA models**

| Variable | CAP | LC | LQ | M1 | M2 | OUT | REV |
|---|---|---|---|---|---|---|---|
| Capital (CAP) | 1.00 | | | | | | |
| Labour costs (LC) | 0.93 | 1.00 | | | | | |
| Labour quantity (LQ) | 0.93 | 0.99 | 1.00 | | | | |
| Intermediate inputs 1 (M1) | 0.89 | 0.97 | 0.96 | 1.00 | | | |
| Intermediate inputs 2 (M2) | 0.90 | 0.97 | 0.96 | 0.99 | 1.00 | | |
| Output (OUT) | 0.96 | 0.95 | 0.95 | 0.91 | 0.92 | 1.00 | |
| Patient Revenue (REV) | 0.93 | 0.99 | 0.98 | 0.98 | 0.98 | 0.95 | 1.00 |

# Appendix 2: Capital stock estimation

### The problem

For the unit record and aggregate analyses of private hospital productivity, it would be useful to derive or obtain an alternative estimate of capital input or capital stock. However, the dataset contains just the following relevant data items for each private hospital for each of the four years 1991–92 through 1994–95:

- number of beds
- interest payments
- depreciation
- gross and net capital expenditure dissected by asset type (land and buildings, computer equipment/installations, major medical equipment, plant and other equipment, intangible assets, other capital expenditure)

Net capital expenditure is equal to gross capital expenditure *less* the trade-in values of replaced items and receipts for sales of replaced items.

As discussed in section 3, it is possible to use the number of beds as a proxy measures for capital input. This note discusses the possibility of constructing a capital input measure from the data on depreciation and capital expenditure.

### The basic idea

For the moment, ignore the distinction between gross and net capital expenditure and between unit record and aggregate data. Also assume that there is no change in the prices of capital.

Let

$I_n$ be capital expenditure in year $n$; we have values for the years 1991–92 (n=1) through 1994–95 (n=4)

$D_n$ be depreciation in year $n$; we have values for the years 1991–92 through 1994–95

$K_n$ be the capital stock in year $n$; we want to estimate values for the years 1991–92 through 1994–95

The variables $I_n$ and $D_n$ are linked through:

- the 'prehistory' of capital expenditure – that is, the values of capital expenditure in years before 1991–92, and
- the depreciation profile – that is, the depreciation method (e.g. straight-line or diminishing balance) and the depreciation rate (or, equivalently, asset life).

Broadly, the method used here is to:

- postulate a plausible range of prehistories for capital expenditure and depreciation profiles;
- search that range for 'feasible combinations' that make the observed values of $I_n$ and $D_n$ consistent with one another; and
- generate an estimate of $K_n$ for each feasible combination.

If the values we obtain for $K_n$ are much the same across all feasible combinations, then we may use these values as our measure of capital stock.

The following is an example of the basic procedure to calculate aggregate estimates, assuming straight line depreciation, no price deflation and equal annual amounts of net investment in prehistory years. Table A2.1 presents the data used to construct the estimates for hospital capital stock.

**Table A2.1: Depreciation data**

| Year | Depreciation | Net investment |
|---|---|---|
| 1991–92 | 85 513 555.00 | 207 818 965.00 |
| 1992–93 | 97 101 844.00 | 254 039 354.00 |
| 1993–94 | 115 023 504.00 | 362 104 902.00 |
| 1994–95 | 138 702 997.00 | 343 601 893.00 |
| Mean | | 291 891 279.00 |

Let

$d$ be the annual depreciation rate

$K_0$ be the base period capital stock (in this instance, the stock at 30 June 1991, which reflects the prehistory of investments made before 1991–92)

$D_0$ be the (constant) annual amount of depreciation on the base period capital stock

Then the following relationships hold:

$$K_1 = K_0 - D_0 + I_1 \times (1-d/2)$$

$$D_1 = D_0 + I_1 \times d/2$$

$$K_4 = K_3 - D_0 - I_1 \times d - I_2 \times d - I_3 \times d + I_4 \times (1-d/2)$$

$$D_4 = D_0 + I_1 \times d + I_2 \times d + I_3 \times d + I_4 \times d/2$$

Because the values for the $I_n$ and $D_n$ (for n = 1,...4) are known, these equations all take the form:

$$y_n = D_0 + x_n \times d$$

and we can estimate $D_0$ and $d$ by regression.

For the sample data, we obtain the regression estimates:

annual depreciation rate: 78 022 034.00

depreciation proportion: 0.0598

Next, assume that the investments that contributed to the base period capital stock occurred in equal annual amounts over some (unknown) time horizon. For a given time horizon, $H$, the amount of annual investment can be then be calculated and the relations above can be used to calculate the capital stock in the years 1991–92 through 1994–95.

The choice of time horizon $H$ is always likely to be somewhat arbitrary. The choice can be made a little less arbitrary by imposing an additional assumption, namely that the annual amounts of investment in the prehistory period were *of the same order* as at the beginning of the period 1991–92 through 1994–95, that is, around $200 million. That assumption bolts down the value of $H$ to around six years and the index of capital stock to:

| 1991–92 | 1992–93 | 1993–94 | 1994–95 |
|---|---|---|---|
| 1.00 | 1.13 | 1.34 | 1.51 |

Indexes of capital stock were calculated for time horizons between 6 and 12 years, showing a remarkable similarity and minimal spread across the different horizons . The index of capital stock appears to increase very marginally with the time horizon and distance from the base year (1991–92 in this case).

### Variations on the basic procedure for aggregate estimates

A number of variations to the basic aggregation procedure outlined above are possible. These include:

- Other models of depreciation. For example, diminishing balance could be used instead of straight-line depreciation.
- Allowing for changes in the price of capital. The basic procedure assumes constant-price data on depreciation, investment and capital stock. If suitable price indexes for capital existed, they could be introduced into the calculations.
- Allowing for retirements and disposals. The private hospitals dataset shows both net and gross investment. For the calculations above, net investment could also have been used. But it would be possible to use gross investment and to subtract the (implied) retirements and disposals from the capital stock.

### Estimating capital stock for individual private hospitals

In principle, the procedure applied above to aggregate depreciation and investment data could be applied to the corresponding data for individual private hospitals. But that would entail running several hundred regressions (or other calculations); moreover, the authors are not confident of the accuracy of the data, especially the data on depreciation.

A less effortful procedure would be to carry results of whole-population or grouped regressions over to the construction of capital stock estimates for individual hospitals. What would that entail? Recall that the regression gives estimates of two 'parameters':

annual depreciation rate

annual amount of depreciation on prehistoric capital stock

While the depreciation rate might be carried over from the aggregate to the unit record analysis, the amount of depreciation could not.

### Reusing the aggregate depreciation rate

Assuming that the average rate of depreciation for the whole population applies to each individual hospital, the following relations can be used to estimate the annual amount of depreciation on prehistoric capital stock for each hospital:

$$D_1 = D_0 + I_1 \times d/2$$

$$D_2 = D_0 + I_1 \times d + I_2 \times d/2$$

$$D_3 = D_0 + I_1 \times d + I_2 \times d + I_3 \times d/2$$

$$D_4 = D_0 + I_1 \times d + I_2 \times d + I_3 \times d + I_4 \times d/2$$

where the $I_n$ and $D_n$ (for n = 1,..4) and $d$ are known.

### Estimating grouped depreciation rates

Assuming the same deprecation rate for all private hospitals may be implausible – some may have a much larger ratio of buildings to equipment in their capital mix, for example.

A middle ground between using the population-wide depreciation rate (which is implausible) and estimating a depreciation rate for each individual hospital (which is impractical) is to group the private hospitals into classes that may have broadly the same depreciation rate, calculate the total depreciation and investment figures for each group, run a regression on each set of totals and use for each hospital the depreciation rate estimated for its group.

# Bibliography


Ainger, D. J., Lovell, C. A. K. & Schmidt, P. 1977, 'Formulation and Estimation of Stochastic Frontier Production Models', *Journal of Econometrics*, vol. 6, pp. 21–37.

Ali, A. I. & Seifort, L. M. 1993, 'The mathematical programming approach to efficiency analysis', in *The Measurement of Productive Efficiency: Techniques and Applications*, eds. H. O. Fried, C. A. Knox Lovell & S. S. Schmidt, 1993, Oxford University Press, New York.

Atkison, S. E. & Wilson, P. W. 1995, 'Comparing mean efficiency and productivity scores from a small sample: A bootstrap methodology', *Journal of Productivity Analysis*, vol. **6**, pp. 132–152.

Australian Bureau of Statistics, *The Australian Consumer Price Index: Concept, Sources and Methods, 1987*, Cat. no. 6461.0, ABS, Caberra.

Australian Institute of Health and Welfare 1995, *Health Expenditure Bulletin*, Issue 12.

Balk, B. M. & Althin, R. 1996, 'A new, transitive productivity index', *The Journal of Productivity Analysis*, vol. **7**, pp. 19–27.

Banker, R. D., Charnes, A. & Cooper, W. W. 1984, 'Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis', *Management Science*, vol. **30**, pp. 1,078–1,092.

Banker, R. D., Conrad, R. & Strauss, R. 1986, 'A comparative application of data envelopment analysis and translog methods: an illustrative study of hospital production', *Management Science*, vol. **32**, pp. 30–44.

Battese, G.E. 1996, 'On the Estimation of Production Functions Involving Explanatory Variables which have Zero Variables', *Working Papers in Econometrics and Applied Statistics,* no. 86.

Battese, G.E. & Coelli, T.J. 1992, 'Frontier production functions, technical efficiency and panel data', *Journal of Productivity Analysis*, vol. 3, pp. 153–169.

Breyer, F. 1987, 'The specification of a hospital cost function - a comment on the recent literature', *Journal of Health Economics*, vol. 6, pp.147–157.

Butler, J.R.G. 1995, *Hospital Cost Analysis*, Kluwer Academic Publishers, Netherlands.

Caves, D. W., Christensen, L. R. & Diewert, W. E. 1982, 'Multilateral comparisons of output, input and productivity using superlative index numbers', *Economic Journal*, vol. **92**, pp. 73–86.

Caves, D.W., Christensen, L. & Tretheway, M. 1980, 'Flexible Cost functions for multiproduct firms', *The Review of Economics and Statistics*, vol. **62**, pp. 477–481.

Charnes, A., Cooper, W. W. & Rhodes, E. 1978, 'Measuring the efficiency of decision making units', *European Journal of Operations Research*, vol. **3**, pp. 392–44.

Christensen, L. & Greene, W. 1976, 'Economies of scale in US Electric power generation', *Journal of Political Economy*, vol. **84**, pp. 655–676.

Coelli, T. 1993, 'Finite Sample Properties of Stochastic Frontiers Estimators and Associated Test Statistics', *Working Papers in Econometrics and Applied Statistics,* no. 73.

Coelli, T. 1995, 'Recent Developments in Frontier Modelling and Efficiency Measurement', *Australian Journal of Agricultural Economics*, vol. **39**, pp. 219–245.

Commonwealth Department of Health and Family Services 1996, *Australian Casemix Report on Hospital Activity 1994–95*, AGPS, Canberra.

Conrad, R. & Strauss, R. 1983, 'A multiple output multiple input model of the hospital industry in North Carolina', *Applied Economics,* vol. **15**, pp. 341–352.

Cowing, T. & Hotmann, A. 1983, 'Multiproduct short-run hospital cost functions: empirical evidence and policy implications from cross section data', *Southern Economic Journal*, vol. **49**, pp. 637–651.

Cromwell J. & Pope, C. 1989, 'Trends in hospital labor and total factor productivity, 1981–86', *Health Care Financing Review*, Summer 1989, vol. **10**, pp. 39–50.

Fare, R., Grosskopf, S. & Logan, J. 1985, 'The relative performance of publicly-owned and privately-owned electric utilities', *Journal of Public Economics*, vol. **26**, pp. 89–106.

Fare, R., Grosskopf, S., Norris, M. & Zhang, Z. 1994, 'Productivity growth, technical progress and efficiency change in industrialised countries', *American Economic Review*, vol. **84**, pp. 66–83.

Fare, R., Grosskopf, S., & Roos, P. 1995, 'Productivity and quality changes in Swedish pharmacies', *International Journal of Production Economics*, vol. **39**, pp. 137–147.

Fare, R., Grosskopf, S. & Valdmanis, V. 1989, 'Capacity, competition and efficiency in hospitals: a non-parametric approach', *The Journal of Productivity Analysis*, vol. **1**, pp. 123–139.

Fare, R. & Primont, D. 1996, 'The Opportunity Cost of Duality', *The Journal of Productivity Analysis*, vol. **7**, pp. 213–224.

Farrell, M. J. 1957, 'The measurement of productive efficiency', *Journal of the Royal Statistical Society*, vol. **120**, pp. 253–282.

Ferrier, G. D. & Hirschberg, J. G. 1997, 'Bootstrapping confidence intervals for linear programming efficiency scores', *Journal of Productivity Analysis*, vol. **8**, pp. 19–33.

Ferrier, G. D. & Lovell, C. A. Knox. 1990, 'Measuring Cost Efficiency in Banking: Econometric and linear programming evidence', *Journal of Econometrics*, vol. **46**, pp. 229–245.

Fried, H. O., Lovell, C. A. Knox. & Schmidt, S. S. 1993, *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, New York.

Greene, W. 1991, *Econometric Analysis*, Macmillan Publishing Company, New York.

Greene, W. 1993, 'The econometric approach to efficiency analysis', in *The Measurement of Productive Efficiency*, eds. H. O. Fried, C. A. Knox Lovell &S. S. Schmidt, 1993, Oxford University Press, New York.

Gregan, T. & Bruce, R. 1997, 'Technical efficiency in the hospitals of Victoria', in *Data Envelopment Analysis: A Technique for Measuring the Efficiency of Government Service Delivery*, Steering Committee for the Review of Commonwealth/State Service Provision (1997), Chapter 5.1, AGPS, Canberra.

Grosskopf, S. & Valdmanis, V. 1987, 'Measuring hospital performance: A non-parametric approach', *Journal of Health Economics*, vol. **6**, pp. 89–107.

Gstach, D. 1995, 'Comparing structural efficiency of unbalanced subsamples: A resampling adaption of Data Envelopment Analysis', *Empirical Economics*, vol. **20**, pp. 531–542.

Hughes, A. & Yaisawarng, S. 1998, 'Sensitivity and Dimensionality Tests of DEA Efficiency Scores', paper presented at the University of New England, Armidale, NSW.

Klein, L. R. 1953, *An Introduction to Econometrics*, Pretence-Hall, Englewood Cliffs, N. J.

Kohler, H. 1985, *Statistics for Business and Economics*, Scott, Foresman and Company, Illinios.

Kopp, R. & Diewert, W. 1982, 'The decomposition of frontier cost function deviations into measures of technical and allocative efficiency', *Journal of Econometrics*, vol. 19, pp. 319–331.

Ling, L. (Qld health) 1997, 'A new synthesis of hospital cost analysis', Paper presented to the 19th Australian Conference of Health Economists, 10–11 July, 1997, Melbourne.

Lovell, C. A. Knox 1996, 'Applying Efficiency Measurement Techniques to the Measurement of Productivity Change', vol. 7, pp. 329–340.

Lovell, C. A. Knox, Walters, L. C. & Wood, L. L. 1993 'Stratified models of education production using DEA and regression analysis' in *Data Envelopment Analysis: Theory, Methods and Applications*, eds. A. Charnes, W. W. Cooper, Y. A. Lewin & L. M. Seiford, 1993, Quorum Books, New York.

McCarty, Therese A. and Yaisawarng, S. 1993, 'Technical efficiency in New Jersey school districts', in *The Measurement of Productive Efficiency: Techniques and Applications*, eds. H. O. Fried, C. A. Knox Lovell & S. S. Schmidt, 1993, Oxford University Press, New York.

Maddala, G. S. 1983, *Limited-dependent and qualitative variables in econometrics*,: Cambridge University Press, Cambridge.

Malmqvist, S. 1953, 'Index numbers and indifference curves', *Trabajos de Estatistica*, vol. **4**, pp. 209–242.

Meussen, W. & van den Brock, J. 1977, 'Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error', *International Economic Review*, vol. 18, pp. 435–444.

Newhouse, J. 1994, 'Fronteir estimation: How useful a tool for health economics?', *Journal of Health Economics*, vol. 13, pp. 317–322.

Nunamaker, Thomas R. 1985, 'Using data envelopment analysis to measure the efficiency of non-profit organisations: A critical evaluation', *Journal of Managerial and Decision Economics*, vol. **6**, pp. 50–58.

Ozxan, Y. 1992, 'Sensitivity analysis of hospital efficiency under alternative output/input definitions', *Knowledge and Policy*, vol. **5**, pp. 1–29.

Register, C. A., & Bruning, E. R. 1987, 'Profit incentives and technical efficiency in the production of hospital care', *Southern Economic Journal*, vol. **53**, pp. 899–914.

Roos, P. 1997, 'Measurement of productivity in hospital services using Malmqvist index approaches', paper presented at the *CSLS Preconference on Service Sector Productivity and the Productivity Paradox*, Ottawa, Canada, February 14–15.

Seifort, L. M. & Thrall, R. M. 1990, 'Recent developments in DEA: The mathematical approach to frontier analysis', *Journal of Econometrics*, vol. **46**, pp. 7–38.

Skinner, J. 1994, 'What do stochastic fronteir functions tell us about inefficiency?', *Journal of Health Economics*, vol **13**, pp. 317–322.

Steering Committee on National Performance Monitoring of Government Trading Enterprises 1992, *Measuring the Total Factor Productivity of Government Trading Enterprises*, AGPS, Canberra.

Thomas, R. 1985, *Introductory Econometrics; Theory and Applications*, Longman Group, London.

Valdmanis, V. 1992, 'Sensitivity analysis for DEA models: an empirical example using public vs. NFP hospitals', *Journal of Public Economics*, vol. **48**, pp. 185–205.

Vitaliano, D. F. & Toren, M. 1994, 'Cost and efficiency in nursing homes: a stochastic frontier approach', *Journal of Health Economics*, vol. 13, pp. 281–300.

Wilson, P. W. 1995, 'Detecting Influential Observations in Deterministic Nonparametric Frontier Models with Multiple Outputs, *Journal of Business and Economic Statistics*, vol. 11, pp. 319–323.

Zuckerman, S., Hadley, J. & Iezzoni, L. 1994, 'Measuring Hospital efficiency with Frontier Cost Functions', *Journal of Health Economics*, vol. **13**, pp. 255–280.

## PREVIOUS ISSUES

## Working Papers in Econometrics and Applied Statistics

| *Working Paper No.* | *Author(s)* | *Title* |
|---|---|---|
| 93/1 (Sep 1993) | Derek Burnell & Peter Rossiter | Investment Expectations: A unit record analysis of data from the Private New Capital Expenditure Survey |
| 93/2 (Dec 1993) | Andrew Sutcliffe | X11 Time Series Decomposition and Sampling Errors |
| 94/1 (Sep 1994) | Steven Kennedy & Peter Rossiter | Australian Labour Force projections 1995 to 2011: Methodology and Preliminary Analysis |
| 94/2 (Oct 1994) | Cynthia Kim, Gerard Salou & Peter Rossiter | Balanced Australian National Accounts |
| 96/1 (May 1996) | Leanne Johnson | Choosing a Price Index Formula: A survey of the literature with an application to price indexes for the tradable and non-tradable sectors |
| 97/1 (Dec 1997) | Genevieve Knight & Leanne Johnson | Tradables: Developing Output and Price Measures for Australia's Tradable and Non-tradable sectors. |
| 98/1 (Mar 1998) | Philip A. Bell & Anthony M. Carolan | Trend Estimation for Small Areas from a continuing survey with controlled sample overlap |
| 98/2 (Apr 1998) | Philip A. Bell | Using State Space Models and Composite Estimation to measure the effects of telephone interviewing on labour force estimates |

## SELF-HELP ACCESS TO STATISTICS

*PHONE* Call 1900 986 400 for the latest statistics on CPI, Labour Force, Earnings, National Accounts, Balance of Payments and other topics. (Call cost is 75c per minute)

*INTERNET* http://www.abs.gov.au

*LIBRARY* A range of ABS publications is available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require.

## WHY NOT SUBSCRIBE?

*PHONE* +61 1300 366 323

*FAX* +61 3 9615 7848

## CONTACTING THE ABS

ABS provides a range of services, including: a telephone inquiry service; information consultancy tailored to your needs; survey, sample and questionnaire design; survey evaluation and methodological reviews; and statistical training.

*INQUIRIES*

| | *By phone* | *By fax* |
|---|---|---|
| Canberra | 02 6252 6627 | 02 6253 1404 |
| Sydney | 02 9268 4611 | 02 9268 4668 |
| Melbourne | 03 9615 7755 | 03 9615 7798 |
| Brisbane | 07 3222 6351 | 07 3222 6283 |
| Perth | 08 9360 5140 | 08 9360 5955 |
| Adelaide | 08 8237 7400 | 08 8237 7566 |
| Hobart | 03 6222 5800 | 03 6222 5995 |
| Darwin | 08 8943 2111 | 08 8981 1218 |

*POST* Client Services, ABS, PO Box 10, Belconnen, ACT 2616

*EMAIL* client.services@abs.gov.au

Working Papers in Econometrics and Applied Statistics No. 98/3: Comparing Techniques for Measuring the Efficiency and Productivity of Australian Private Hospitals

2135100006985
ISBN 0 642 25731 0

RRP $20.00