THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 1999/36

Measuring Performance in Primary Care: Econometric Analysis and DEA

by

Antonio Giuffrida and Hugh Gravelle

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Measuring Performance in Primary Care: Econometric Analysis and DEA

Antonio Giuffrida and Hugh Gravelle[*]

Abstract

We use data from the Health Service Indicators database to compare different methods of measuring the performance of English Family Health Services Authorities (FHSAs) in providing primary care. A variety of regression and data envelopment analysis methods are compared as summary efficiency measures of individual FHSA performance. The correlation of the rankings of FHSAs across DEA and regression methods, across two years of data and across three different specifications of the technology of primary care are examined. Efficiency scores are highly correlated within variants of the two methods, and across years for a given method. Inter method correlations are smaller and correlations across different specifications of the primary care production process are negligible and sometime negative.

---

[*] National Primary Care Research and Development Centre, Centre for Health Economics, University of York, York, YO1 5DD; emails: ag19@york.ac.uk, hg8@york.ac.uk. Support from the Department of Health to the NPCRDC is acknowledged. The views expressed are those of the authors and not necessarily those of the Department of Health. We are grateful to Martin Roland, Peter Smith, Matthew Sutton and a referee for helpful comments.

1. Introduction

In setting out its policy for the English National Health Service in the White Paper *The New NHS: Modern Dependable,* the incoming Labour government stressed the need for a new performance framework to measure progress towards its objectives (Department of Health, 1997). Subsequent NHS Executive (1998; 1999) documents suggested performance indicators to compare the quality of health care in health authorities and the extent to which health authorities are effectively delivering appropriate health care.[1]

The performance of an organisation can be described along a large number of dimensions relating to outputs produced, aspects of the quality of those outputs, and the inputs used. NHS performance indicators typically relate to one or at most two of these dimensions. Such performance indicators can be useful as a means of identifying unusual performance and stimulating questions about the reasons for it, but assessment of the overall performance of the organisation requires a means of combining the information on the different dimensions. Without a systematic method of summarising performance along all the dimensions it is not possible to assess whether unusually bad performance on one indicator is due to poor decision making or whether it is explained by the organisation having an unusually good performance on some other dimension.

Two main techniques have been used to improve on single-indicator performance measures. The econometric approach estimates a production or cost function by fitting a regression plane to the data. Data Envelopment Analysis (DEA) uses linear programming techniques to construct a frontier which envelops all the observations.

Most studies of efficiency in the production of primary care to date have used DEA rather than regression analysis.[2] The American literature on efficiency in primary health has expressed enthusiasm about the usefulness of DEA. For example, the pioneering study of Huang and McLaughlin (1989) concluded that DEA can contribute to the evaluation of rural primary health care programs. Prominent advocates of the use of DEA in evaluating physicians' performance are Chilingerian and Sherman (1996; 1997). They have explored the use of DEA to identify "best practice" primary care physicians, and have calculated the potential savings if inefficient physicians were to adopt "best practice" patterns. Similarly,

---

[1] The proposed indicators for primary care management of chronic diseases have been analysed in Giuffrida et al. (1998; 1999).

[2] Hollingsworth et al. (1999) review DEA efficiency measurements in health care generally.

Ozcan (1998) used DEA to examine primary care physicians' efficiency in the treatment of otitis media by analysing geographic variations in practice patterns, and the impact of inefficient practice patterns on the cost of treatment.

In the UK Thanassoulis et al. (1995, 1996) looked at perinatal care and stressed DEA's advantage, compared with traditional ratio analysis, of considering more than one dimension of performance. Salinas-Jimenez and Smith (1996) used DEA to compare efficiency across Family Health Service Authorities (FHSAs) which were the administrative unit for primary health care in England at the time. DEA based Malmquist indices have been constructed to examine changes in FHSA productivity over the 1990/1-1994/5 period (Giuffrida, 1999). Only small improvements in productivity over the period were found which were attributable to changes in technical and scale efficiency, with no change in the technology. Bates et al. (1996) analysed the prescribing costs of general practices in one FHSA with DEA but found that the results were strongly influenced by a few outliers. Garcia et al. (1997) used DEA to examine Spanish primary health care centres but the efficiency estimates obtained were heavily affected by small changes in the specification of the output.

There have been fewer regression studies of primary care efficiency. Defelice and Bradford (1997) applied stochastic frontier regression methods to analyse primary care physicians in the United States and found evidence that physicians working in group practice were more efficient than solo practitioners. Whynes et al. (1997) used econometric methods on the same data set of Lincolnshire general practitioners as Bates et al. (1996) and obtained more robust results indicating that GP fundholders were more efficient prescribers than non-fundholders. Giuffrida et al. (1999b) have used stochastic frontier regression to examine the determinants of administrative costs in primary care at FHSA level and to identify inefficient FHSAs.

There are no theoretical grounds for generally preferring DEA to regression analysis or vice versa. Eminent practitioners examining efficiency in different industries using different data sets have reached different conclusions about which method is to be preferred (Cubbin and Tzanidakis, 1996; Smith, 1990; Thanassoulis, 1993). Given the increased emphasis on the measurement of performance of health authorities in primary care in England and Wales (NHS Executive, 1999) it is important that the implications of using the alternative methods of measuring performance on health authority level data on primary care are investigated.

This paper is a first step in that direction. Since there appears to be no reason to prefer one method to the other on theoretical grounds we first see if the method used makes a difference either to the average efficiency of health authorities. We then examine the correlations of individual primary care health authority efficiency scores across methods to determine if the method used makes a difference when the efficiency of authorities are compared. Finally, we examine the robustness of the different methods by calculating the correlations of efficiency scores for authorities across two consecutive years and across different specifications of the production process in primary care.

We outline the two methodologies and their advantages and disadvantages in the next section. Section 3 describes the database, the alternative specifications of the primary care technology and the procedures for estimating efficiency of FHSAs. The results from using the two methodologies are compared in section 4 and section 5 summarises the lessons to be drawn.

2. An outline of the methodologies

We are concerned with cost inefficiency where the costs of producing a given output are not minimised. There are three sources of cost inefficiency. *Technical inefficiency* arises when too little output is being produced from a given bundle of inputs. The decision making unit (DMU) lies above the isoqant for its output level. There is *allocative inefficiency* when inputs are being employed in the wrong proportion, given their prices and productivity at the margin. The DMU is on its isoquant but at the wrong point. Finally, *scale inefficiency* occurs when total cost can be reduced by changing the size and number of DMUs: individual DMUs are on the wrong isoquants.

2.1 Econometric approaches to efficiency measurement

2.1.1 Deterministic cost frontier (COLS)

Efficiency can be measured by estimating a cost function in which the error term e_i is constrained to be non-negative (Aigner and Chu, 1968)

$$\ln C_i = \ln C(y_i, b) + e_i, \qquad e_i \geq 0 \tag{1}$$

where C_i is the observed cost for unit i, $\ln C(y_i, b)$ is the log cost function, y_i is a vector of outputs; and b is a vector of parameters to be estimated in the regression.

The parameters of the cost function are estimated by OLS and the intercept is shifted up until all residuals are non-negative and one is equal to zero. If $\hat{e}^{\min}$ is the value of the most negative residual, the Corrected OLS (COLS) residuals are $\hat{e}_i^{\mathrm{COLS}} = \hat{e}_i - \hat{e}^{\min}$. The efficiency scores are calculated as $\exp\left(-\hat{e}_i^{\text{COLS}}\right)$. The COLS procedure counts the most efficient unit as 100% efficient.

2.1.2 Stochastic cost frontier

In the stochastic frontier approach (Aigner et al., 1977; Meeusen and Van Den Broeck, 1977) the error in the cost function is decomposed into two terms: $e_i = u_i + v_i$. The one sided error u_i represents inefficiency and v_i is the usual statistical noise

$$\ln C_i = \ln C(y_i, b,) + u_i + v_i, \quad u_i \geq 0 \tag{2}$$

Estimation of a stochastic cost frontier is possible only if the residuals are positively skewed. The most common error specifications are the normal distribution for the statistical noise, v_i, and half normal, exponential, and truncated (at zero) normal distributions for the inefficiency term, u_i.

2.1.3 Canonical regression

Canonical regression (Vinod, 1968) is a method for efficiency measurement which does not require the aggregation of inputs into a cost measure. Vinod (1976) proved that canonical regression technique is maximum likelihood and provides consistent estimates of slope parameters. Ruggiero (1998) showed that the technique provides an alternative way to estimate technical efficiency for multiple output production processes.

Assume that the production function is separable in outputs and inputs and representable as a generalised Cobb-Douglas

$$\sum_{m=1}^{M} a_m \ln y_{im} = \sum_{n=1}^{N} b_n \ln x_{in} + \ln g_i \tag{3}$$

where y_{im} are the outputs, x_{in} are the inputs and g_i is the efficiency index ($0 < g_i \leq 1$) for DMU i. Two variables U_i and V_i are specified as log-linear combinations of the outputs and inputs

$$U_i = \sum_{m=1}^{M} a_m \ln y_{im} \quad \text{and} \quad V_i = \sum_{n=1}^{N} b_n \ln x_{in} \tag{4}$$

The weights a_m and b_n are chosen so that the correlation between U_i and V_i across DMUs is maximised

$$r^* = \max_{a_m, b_n} \text{Corr}(U_i, V_i) \tag{5}$$

leading to estimates a_m^* and b_m^*. Since

$$U_i = r^* V_i \tag{6}$$

substitution in (3) gives an estimate of the efficiency of DMU i as

$$\hat{g}_i = \exp\left[\sum_{m=1}^{M} a_m^* \ln y_{im} - \sum_{n=1}^{N} r^* b_n^* \ln x_{in}\right] \quad (7)$$

As with the COLS procedure the efficiency scores are rescaled relative to the most efficient unit.

2.2 Mathematical programming approach: DEA

Data Envelopment Analysis (DEA) has its origins in the seminal work by Charnes et al. (1978). The technical efficiency of decision making units (DMUs) is found by solving a linear programming problem for each DMU *i*

$$\begin{aligned}
&\max_{u_i, v_i} q = \sum_{m=1}^{M} u_{im} y_{im} / \sum_{n=1}^{N} v_{in} x_{in} \qquad \text{s.t.} \\
&\sum_{m=1}^{M} u_{km} y_{km} / \sum_{n=1}^{N} v_{in} x_{in} \le 1,\ k = 1,\ldots,K \\
&u_{im} > 0, v_{in} > 0, \forall m,n
\end{aligned} \quad (8)$$

Efficiency in DMU *i* is measured as the ratio of a weighted sum of outputs and to a weighted sum of inputs: $\left(\sum_{m=1}^{M} u_{im} y_{im} / \sum_{n=1}^{N} v_{in} x_{in}\right)$. The linear programming problem finds the vectors of weights u_i, v_i which maximise the efficiency score of the DMU *i*, subject to the constraint that no unit has an efficiency score greater than one at those weights. This is sometimes expressed as putting the DMU "in the best possible light". The method can also be applied when there is information on costs and outputs be treating the cost as a single "input" variable.

When constant returns to scale or constant average costs are assumed, as in (8), the efficiency frontier is, in the single output-single input case, a ray from the origin through the unit with the lowest average cost. DEA can also be applied under the assumption of variable returns to scale or non-constant average costs (Banker et al., 1984). With variable returns the cost frontier is piecewise linear and fewer DMUs will be shown to be inefficient.

2.3 Properties of regression and DEA methods

The theoretical advantages and disadvantages of the regression and DEA methods have been extensively discussed (Fried et al., 1993) and are summarised in Table 1. The DEA procedure yields a number of useful by-products. It generates the set of "peer" units with which a unit is compared. DEA identifies the output dimensions along which an inefficient unit is performing badly because it tends to assign unusually high weights to dimensions where there is poor performance in an attempt to maximise the efficiency score for the unit. DEA can also easily model multiple input, multiple output production processes. The main advantage of DEA is that it does not require specification of the functional form of the production (or cost) function.

Econometric techniques require more assumptions about the production or cost function and also about the distribution of the errors. However, it is possible to test for the validity of the assumptions and to determine whether particular variables are relevant. With canonical regression it is also possible to allow for multiple output production processes, though at the cost of imposing a highly restrictive functional form.

Since theoretical considerations are inconclusive Monte Carlo simulations have been performed to compare the success of the two approaches in estimating a known production process. Banker et al. (1993) suggested that COLS performs better than DEA when inefficiency has a half normal distribution and the sample has more than 50 observations. On the other hand DEA gives more precise estimates when inefficiency has an exponential distribution or the sample size is than 50. Smith (1997) showed that DEA produces reasonably accurate estimates with small samples and that accuracy improved with highly correlated inputs or outputs.

Gong and Sickles (1989; 1992) compared DEA and stochastic frontier regression when panel data are available. The results were again conflicting. Stochastic frontier models outperform DEA if the assumed functional form is close to the underlying technology. But as the misspecification of the functional form becomes more serious, DEA estimates become more accurate than the econometric-based estimates. Read and Thanassoulis (1995) compared DEA and stochastic frontiers methods where the assumed specification of the production function was good except at certain points where one of the output or input variables was very small or large. They found that stochastic frontier regression estimates of efficiency were worse than DEA in these regions of poor specification. Ruggiero (1998) compared canonical regression with DEA in estimating efficiency when the true production function is Cobb-Douglas. Canonical regression estimates were more highly correlated with the true efficiency irrespective of whether irrelevant input or output variables were included in the models estimated. He did not consider the relative performance of the two methods for other technologies.

Theoretical considerations and simulation studies suggest that regression methods may be superior to DEA under some circumstances and worse under others. When we are interested in measuring the performance of particular groups of DMUs it is therefore sensible to investigate the results from applying the alternative methods to the particular data set. Since we will not know the true underlying technology we can proceed by initially asking if the alternative methods yield similar results when applied to alternative specifications of the underlying production process.

If the methods produce different results we can compare their robustness in two ways (Parkin and Hollingsworth, 1997). First, since, at least in primary care, it is plausible that the underlying technology does not vary markedly from one year to the next, we can investigate the correlation of the efficiency scores of FHSAs between consecutive years. If the scores show little correlation this suggests that the method is not a robust measure of efficiency in primary care. Second, we can compare the robustness of the methods across different assumptions about the underlying technology. When the true technology is unknown and there are a number of plausible alternative specifications, risk aversion suggests that we should look more favourably on a method which yields highly correlated efficiency scores across the different assumptions about the technology.

3. Data, models and methods

3.1 Data

The unit of analysis is the Family Health Service Authority (FHSA) which was, during the period of our study, the administrative unit for primary health care in England. There were 90 FHSAs, with average populations of around 560,000 individuals served by around 290 GPs. Patients are registered with a GP whose major role is to act as a gatekeeper, referring patients to specialised services. GPs treat minor ailments, prescribe pharmaceuticals, carry out various public health tasks, and may carry out minor surgery. Most primary medical care services are provided from GP surgeries. With the exception of about 20% of prescriptions, primary care is free at point of use and financed from taxation. GPs are remunerated by a mixture of lump sums, capitation, payments for achieving certain vaccination and screening targets, and fees per item of service (mainly for out of hours visits). They employ practice nurses and other staff.

The data are derived from the Health Service Indicators data sets (NHS Executive, 1996) for the 1993-4 and 1994-5 years financial years (April 1 to 31 March).[3] Table 2 lists the

[3] 10 FHSAs in 1993/94 and 7 in 1994/95 had one or more missing values for some variables. We interpolated missing values using information from other years. Preliminary analysis showed that using the interpolated values yielded results which were very similar to those when observations with missing values were dropped.

variables and the roles they play in the econometric and DEA analyses and has summary statistics. Variables used in DEA are in absolute form while the variables in the econometric analysis are in natural logarithms.

3.2 Alternative models of performance

Our purpose is not so much to find the "best" fitting model of costs but to examine the robustness of performance scores to the model fitted, as well as to the data and the estimation methods. We specify three different models of primary care provision to obtain estimates of FHSAs' performance. The first model replicates Salinas-Jimenez and Smith (1996), which is the only previous attempt to measure FHSAs' performance using DEA. The second model uses the same cost measure but an overlapping set of output variables. The third model uses information on inputs and their prices to construct a synthetic cost measure.

3.2.1 Model 1: Costs and quality in primary care

In model 1 the cost of producing primary care is assumed to depend on a set of measures of quality and on environmental factors. Cost is defined as gross expenditure by the FHSA on General Medical Services in £000's (*EXPEND*)[4] and includes the cost of GPs and their prescribing. Higher factor costs in London were allowed for by applying wage weights for inner and outer London according to the proportion of expenditure on practice staff.

Salinas-Jimenez and Smith (1996) suggest a number of variables as indicators of primary care quality and hence as factors affecting the cost of primary care. These are: GPs (*GPs*); practices with nurses (*WITHNURSE*), GPs with less than 2,500 patients in their list (*LOWLIST*); GPs not practising single-handed (*NOTSOLO*); GPs who achieved the higher target for childhood immunisation (*IMM*); GPs who achieved the 80% target for cervical cytology testing (*CER*); and practices which satisfied the minimum standard set defined by the Statement of Fees and Allowances (*MINSTAN*).

Two other factors are also suggested to affect cost. The first is the number of patients registered with GPs in the area (*PATIENTS*) since cost is measured as total expenditure, rather than expenditure per capita. The second is the standardised mortality rate from all causes of

---

[4] *EXPEND* is defined as: gross payments in £'000s for General Medical Services, FHSA cash limited and non cash limited expenditure, excluding fundholders' drugs, inclusive of superannuation contribution but net of prescription charges in the FHSA in the financial year.

FHSA residents aged 0-64 (*SMR*). This variable is interpreted as an environmental variable which is expected to affect FHSAs' ability to deliver care of given quality.[5]

3.3 Model 2: Alternative quality specification

The second model uses the measure of cost but a different specification of the factors affecting it. Two variables are dropped. GP list size (*LOWLIST*) was removed because the model also has the number of patients and GPs entered separately. The number of GPs who were not single handed practice (*NOTSOLO*) was dropped because the evidence on the quality of care of single handed practices is ambiguous (Curtis, 1987; Roos, 1980). We included the number of practice nurses in the FHSA (*NURSES*), rather than the number of practices with a practice nurse (*WITHNURSE*), since it could be argued that this is a more sensitive measure of the use of practice nurses and hence of quality. Additional measures of primary care quality are included: the number of GPs who achieved the higher target for pre-school boosters (*BOOST*) and the number of GPs who provided minor surgery services (*MSURG*).

3.4 Model 3: Inputs, outputs and quality in primary care

There is a possible misspecification of the cost variable in the first two models. The variable *EXPEND* includes GPs' remuneration for achieving certain quality targets (for example for childhood immunisation and cervical cytology tests) which were also included among the measures of quality assumed to affect the cost of primary care. If, *ceteris paribus,* more GPs in FHSA *A* achieve the targets for cervical cytology tests than in FHSA *B*, then *A* is more efficient than *B*. However, since *EXPEND* includes the payments given to GPs for meeting targets, *A* is shown as more expensive than *B*. In DEA both FHSAs would be estimated as efficient and in regression analysis the coefficient on cervical cytology targets would be biased upward, the inefficiency of *B* would be over-estimated and that of *A* would be under-estimated.

We can avoid the problem by using a third specification in which the quality of care depends on the input vector (GPs and practice nurses). Estimating inefficiency by standard regression methods in this specification requires the aggregation of the inputs into a "synthetic" cost variable (*COST*) using their annual remuneration as weights (Medeconomics, 1995). With canonical regression the weights for the aggregation of the inputs are generated

[5] Salinas-Jmenez and Smith (1996) had data from 1991/2 and used the standardised illness ratio and the unemployment rates derived from the 1991 population census as environmental variables. Rather than use 1991

by the analysis. For DEA there are two options. The synthetic COST measure can be used as a single input and the two inputs can be included separately.

Two additional factors are allowed for. The area of the FHSA (*AREA*) is included since less densely populated FHSAs may require more GPs or practice nurses to provide primary health care. The number of patients classified as deprived (*DEPRIV*) is also included to allow for the possibility that more inputs are required to deliver primary care in deprived areas.[6]

3.5 Estimation of efficiency

A variety of regression and DEA methods were applied to the three models and two years of data to estimate the efficiency of FHSAs in producing primary care, yielding 45 different sets of efficiency scores. Table 3 shows the combinations used.[7] Five regression methods were examined: corrected ordinary least squares (COLS), stochastic frontier with half normal, exponential and truncated errors (SFN, SFE, SFT) and canonical regression (CAN). Some combinations of method and model were not estimated because they were not sensible (for example the canonical regression is pointless when there is a single dependent variable as in models 1 and 2). Other combinations could not be estimated on the data set. In SFT did not converge for models 1 and 2 and for 1994-5 the residuals were negatively skewed so that none of the stochastic frontier models could be estimated.

For DEA it is possible to estimate both constant (CRS) and variable returns to scale (VRS) and, for model 3 to estimate cost inefficiency (CE) using synthetic cost and to estimate technical efficiency (TE) using the GP and practice nurse inputs separately. Hence there are four possible types of DEA: DEA VRS CE, DEA VRS TE, DEA CRS CE, DEA CRS TE. All the DEA efficiency scores are calculated using the input orientation and show the reduction in inputs which would be possible given the outputs.

The methods were applied to the two years separately and to the pooled set of two years data. For the pooled data the stochastic frontier was estimated using a random effects panel specification with the inefficiency term assumed to have a half-normal distribution and

---

census measures, we preferred to use the SMR as it is available for 1993/4 and 1994/5.

[6] In DEA deprived patients were subtracted from the total number of patients to construct the *PATIENTS* variable to avoid double counting. In the regression analyses the percentage of the patients that were *not* classified as deprived was used to avoid taking the logarithm of zero for some FHSAs.

[7] Stochastic frontier models were estimated using the software LIMDEP 7.0 (Greene, 1995); canonical regression was estimated using STATA 6 (StataCorp, 1999); DEA efficiency scores were estimated using the Warwick Windows DEA 1.02 package (Thanassoulis and Emrouznejad, 1996).

computed using the Battese and Coelli (1988) procedure. The panel was too short to use the within fixed effects estimator. The DEA efficiency measures for the pooled data are defined as the average of the scores in the two years relative to the frontier estimated from the pooled data (Tulkens and Vanden Eeckaut, 1995).

4. Results

In this section we report the efficiency scores estimated using the alternative methods. The regressions used to generate the efficiency scores are not given here but are available on request from the authors. Model 3 regression results were usually more in line with prior expectations than models 2 and 3 but all models performed reasonably well in terms of specification tests and the significance and signs of coefficients.

4.1 Comparison of average performance scores

Table 3 reports the average and standard deviation of FHSA efficiency scores estimated by applying the different methods to the different models across the different years of data. The differences in the mean efficiencies across the COLS and stochastic frontier regression methods are not large. COLS efficiency scores are usually lower than the stochastic frontier estimates since the COLS method assumes that all but one FHSA are inefficient. The different assumptions about the distribution of the inefficiency term in the stochastic frontier methods makes very little difference to average efficiencies. Somewhat lower efficiency scores with a wider dispersion are produced by the canonical regressions.

DEA average efficiencies are similar to those from the regression based methods and are usually within the ranges of COLS and the stochastic frontier models. As anticipated, when variable returns to scale are assumed the average efficiency scores are higher and more FHSAs are efficient. Also in line with expectations (Smith, 1997), the models with larger numbers of inputs and outputs yield higher average efficiencies.

4.2 Comparison of FHSA efficiency scores across methods

The focus of the performance assessment framework in the NHS is not on the average efficiency of DMUs but the efficiency scores of individual DMUs. Hence, we wish to compare the alternative methods of estimating efficiency as procedures for identifying FHSAs which appear to be performing relatively well or badly compared to other FHSAs. We start by comparing the correlations of efficiency scores across the different methods for given models and years. The aim is to see if the method makes a difference to the ranking of FHSAs by estimated efficiency.

Tables 4 to 6 report the Spearman's pairwise rank correlation coefficients for alternative methods for the three models of primary care production and for the separate years of data.[8] The entries below the diagonal show the correlations between pairs of methods in year 1 and entries above the diagonal give the correlations from year 2 data. Thus in table 6 the correlation on year 1 data between COLS and CAN is 0.823 and the correlation on year 2 data between CAN and SFT is 0.826

The tables indicate that it does not matter greatly whether FHSAs are ranked by COLS or the three stochastic frontier methods since the pairwise correlations are over 0.98. The correlation between the canonical regression estimates (CAN) and the other econometric-based methods are somewhat lower but never less than 0.81.

Correlations between the DEA models are also quite high, between 0.74 and 0.87. Comparison of the rankings from variable and constant returns to scale DEA suggest that assumptions about returns to scale have only a modest impact.

The correlations between the rankings obtained using econometric methods and DEA are much lower: they are between 0.56 and 0.73 in the first model, and between 0.53 and 0.64 in the second model. In the third model the correlation between overall efficiency estimated by DEA and the econometric techniques is never greater than 0.54. The correlation coefficients for rankings by DEA TE and CAN are as low as 0.21.

Our correlations between regression and DEA efficiency estimates are less than those reported by De Borger and Kerstens (1996) from a study of Belgian municipalities. They found a rank correlation coefficient of 0.99 among the regression based methods and 0.82 between these techniques and DEA. By contrast Cummins and Zi (1997) analysed efficiency among USA life insurance companies and found a 0.96 rank correlation among the stochastic frontier methods and but only a 0.59 correlation between stochastic frontier and DEA.

The role of the returns to scale assumption in the DEA model has a mixed effect on the correlation between DEA and efficiency scores given by the econometric methodologies. In the first two models, the rank correlations are higher with VRS, but in the third model correlations are greater with CRS. A similar mix of results have been observed in other studies. Bryce et al. (1998) analysing the efficiency of American Health Maintenance

[8] Pairwise Pearson's correlation coefficients were almost identical.

Organisations found that econometric efficiency estimates were more highly correlated with those from DEA VRS than with DEA CRS. By contrast Linna (1998), in a study of Finnish hospitals, found that the correlation of efficiencies generated by regression and DEA was greater when the DEA model was CRS than VRS.

The conclusion we reach is that for estimating efficiency at authority level in English primary care the rankings of FHSAs are relatively insensitive to the choice amongst regression methods or amongst DEA methods for a given year of data or a given underlying specification of the technology. The choice between DEA and regression does make a considerable difference to the rankings.

4.3 Inter-temporal consistency of different methods

It seems unlikely, on a priori grounds, that there are large variations in performance at FHSA level between consecutive years. One would expect the law of (fairly) large numbers to operate to even out practice level year to year variations, given that a typical FHSA had around 100 practices and 300 GPs. A measure which suggested great year to year variation in efficiency ranking should therefore be regarded unfavourably.

Table 7 reports the correlations of rankings between years 1 and 2 for a given method and a given model. All the econometric methods show reasonably high inter-year correlations over all three models (between 0.82 and 0.87) with the canonical regression method doing best. Efficiency scores estimated by DEA are more stable than the scores estimated by the regression methods in the first model, but less stable in the other two models.

Parkin and Hollingsworth (1997) reported inter year correlations of DEA scores of between 0.53 and 0.69 in a study of Scottish hospital. These results are generally somewhat less than our DEA inter year correlations and markedly less than the inter year correlations for the regression methods.

Choice between the methods on grounds of intertemporal stability requires an assumption about which specification of the technology is better. As our account of the three models suggests, model 3 is perhaps the best specification so that regression methods, particularly CAN, are to be preferred on grounds of intertemporal stability in assessing primary care efficiency.

4.4 Inter-model consistency of different methods

Table 8 gives the pair-wise rank correlations across different models for given methods and given years of data. The correlations of rankings from models 1 and 2 are quite high for both regression and DEA methods and for both years of data. However, the rankings from models 1 and 2 are poorly correlated with those of model 3 for all methods and years. This is perhaps not surprising: models 1 and 2 used the same input and have overlapping output sets. The results indicate that relatively minor changes in the assumptions about the technology have relatively small effects on rankings. Model 3 has a very different, and arguably more sensible specification. Rankings of FHSAs by their efficiency scores are thus very sensitive to the model used, much more so than to the choice of the methodology or the year.

5. Discussion

There is increased emphasis placed on performance management in the English NHS. We argue that it is important to consider the overall efficiency of decision making units as well as comparing them on the basis of performance indicators which capture only one dimension of performance. There are no general grounds for preferring one method of measuring overall efficiency to another. Previous studies in other industries have shown that the results from comparing the different methods can differ so that it is important to compare the methods using data generated by the particular industry being examined.

One must first investigate whether the different methods yield similar results. In the case of primary care our investigation shows that the method makes a considerable difference for a given year of data and a given model. Average efficiency scores differ across the methods and the correlations of the efficiency scores of individual health authorities between DEA and regression methods were low. The correlations between variants of the regression methods or between variants of DEA were high.

When the method does make a difference to the results, as it does in primary care, one can then try to assess the robustness of the methods across years and across models of the technology. The across year correlations of authorities efficiency scores were generally higher for the regression methods than for the DEA methods. Since it seems unlikely that there were marked changes in efficiency between 1993/4 and 1994/5 the temporal robustness of the results suggest that the regression methods perform better as measures of efficiency.

There was little to choose between the alternative methods in terms of their robustness to assumptions about the underlying productions process in primary care: each method had high correlations between the rankings from models 1 and 2 but very low correlations between models 1 or 2 and model 3. None of the methods was robust across major differences in the assumed technology.

Assumptions about the underlying technology of primary care make more difference to rankings of authorities by their efficiency scores than the method used to measure efficiency. Hence the measurement of efficiency requires a means of choosing between the models of technology as well as criteria, such as intertemporal robustness, for choosing between methods of measurement.

Theory-based arguments for choosing specifications are required. Although the literature on regression and DEA methods contains much sophisticated statistical and mathematical analysis it has not been based on models of decisions which affect the efficiency of the organisations examined. The primary care data has been generated by decisions taken by patients, practices, trusts, managers and national level policy makers. In order to interpret the data and thus to decide if it is useful for measuring performance it is necessary to model the behaviour which generates the data.[9]

[9] We have made an initial attempt to model the implications of how inefficiency arises in organisations for the interpretation of the efficiency scores produced using regression methods (Giuffrida et al, 1999b).

References

Aigner DJ, Chu SF. (1968) On estimating the industry production function. *American Economic Review*, 58, 226-39.

Aigner DJ, Lovell CAK, Schmidt P. (1977) Formulation and estimation of stochastic production function models. *Journal of Econometrics*, 6, 21-37.

Banker RD, Chang H, Cooper WW. (1984) Some models for estimating technical and scale efficiency in data envelopment analysis. *Management Science*, 30, 1078-1092.

Banker RD, Gadh VM, Gorr WL. (1993) A Monte Carlo comparison of two production frontier estimation methods: Corrected ordinary least squares and data envelopment analysis. *European Journal of Operational Research*. 67, 332-343.

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

Battese GE, Coelli TJ. (1988) Prediction of firm-level technical efficiencies with a generalised frontier production function and panel data. *Journal of Econometrics,* 38, 387-399.

Bryce CL, Engberg JB, Wholey DR. (1998) *Comparing the Agreement Among Alternate Models In Evaluating HMO Efficiency*. University of Pittsburgh, mimeo.

Charnes A, Cooper WW, Rhodes E. (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429-444.

Chilingerian JA, Sherman HD. (1996) Benchmarking physician practice patterns with DEA: a multi-stage approach for cost containment. *Annals of Operational Research*, 67, 83-116.

Chilingerian JA, Sherman HD. (1997) DEA and primary care physician report cards: deriving preferred practice cones from managed care service concepts and operating strategies. *Annals of Operational Research,*73, 35-66.

Cubbin J, Tzanidakis G. (1996) *Regression Versus Data Envelopment Analysis for Efficiency Measurement: An Application to the England and Wales Water Industry*. The City University, Discussion Paper 55.

Curtis SE. (1987). The patient's view of general practice in an urban area. *Family Practice,* 4, 200-6.

Cummins DJ, Zi H. (1997) *Measuring Cost Efficiency in the U.S. Life Insurance Industry: Econometric and Mathematical Programming Approaches*. Wharton School. University of Pennsylvania. Discussion Paper 97-03.

De Borger B, Kerstens K. (1996) Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches. *Regional Science and Urban Economics*, 26, 145-170.

Defelice LC, Bradford DW. (1997) Relative inefficiency in production between solo and group practice physicians. *Health Economics*, 6, 455-465.

Department of Health (1997) *The New NHS: Modern Dependable*. HMSO, London.

Fried HO, Lovell CAK, Schmidt SS. (1993) *The measurement of productive efficiency: techniques and applications*. Oxford University Press, New York.

Garcia F, Marcuello C, Serrano D, Urbina O. (1997) *Evaluation of technical efficiency in primary health centres*. Paper presented to the EURO-INFORMS congress, Barcellona, 14-17 July.

Giuffrida A. (1999) Productivity and efficiency changes in primary care: a Malmquist index approach. *Health Care Management Science*, 2, 11-26.

Giuffrida A, Gravelle H, Roland M. (1998) *Performance Indicators for Primary Care Management in the NHS.* Centre for Health Economics Discussion Paper Series 160, University of York.

Giuffrida A, Gravelle H, Roland M. (1999) Measuring quality of care with routine data: avoiding confusion between performance indicators and health outcomes. *British Medical Journal*, 319, 94-98.

Giuffrida A, Gravelle H, Sutton M. (1999b) *Efficiency and Administrative Costs in Primary Care*. Department of Economics and Related Studies Discussion Paper Series 99/27, University of York.

Gong BH, Sickles RC. (1989) Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. Journal of Productivity Analysis, 1, 229-261.

Gong BH, Sickles RC. (1992) Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. *Journal of Econometrics*, 51, 259-284.

Greene W. (1995) *LIMDEP Version 7.0 User's Manual*. Econometric Software Inc., Plainview, NY.

Hollingsworth B, Dawson PJ, Maniadakis N. (1999) Efficiency measurement of health care: a review of non-parametric methods and applications. *Health Care Management Science*, 2, 161-172.

Huang Y-GL, McLaughlin CP. (1989) Relative efficiency in rural primary health care: an application of data envelopment analysis. *Health Services Research*, 24, 143-158.

Linna M. (1998) Measuring the hospital cost efficiency with panel data models. *Health Economics*, 7, 415-427.

Medeconomics (1995) NHS Pay Database. *Medeconomics*, 16(10), 74-87.

Meeusen, W. Van Den Broeck, J. (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18, 435-44.

NHS Executive (1996) *Health Service Indicators Handbook*. NHS Executive, Leeds.

NHS Executive (1998) The New NHS Modern and Dependable: A National Framework for Assessing Performance. HMSO, London.

NHS Executive (1999) Quality and Performance in the New NHS: High Level Performance Indicators. HMSO, London.

Ozcan YA. (1998) Physician benchmarking: Measuring variation in practice behavior in treatment of otitis media. *Health Care Management Science*, 1, 5-17.

Parkin D, Hollingsworth B. (1997) Measuring production efficiency of acute hospitals in Scotland, 1991-94: validity issues in Data Envelopment Analysis. *Applied Economics*, 29, 1425-1433.

Read L, Thanassoulis E. (1995) *A comparison of data envelopment analysis and stochastic frontiers under variation of fit*. Business School, Working Paper 234, University of Warwick.

Roos NP (1980) Impact of the organization of practice on quality of care and physician productivity. *Medical Care*, 18, 347-59

Ruggiero J. (1998) A new approach for technical efficiency estimation in multiple output production. *European Journal of Operational Research*, 111, 369-380.

Salinas-Jimenez J, Smith P. (1996) Data envelopment analysis applied to quality in primary health care. *Annals of Operational Research*, 67, 141-161.

Smith P. (1990) The use of performance indicators in the public sector. *Journal of the Royal Statistical Society*, Series A 153, 53-72.

Smith P. (1997) Model misspecification in Data Envelopment Analysis. *Annals of Operations Research*, 73, 233-252.

StataCorp. (1999) *Stata Statistical Software: Release 6.0*. Stata Corporation, College Station, TX.

Thanassoulis E. (1993) A comparison of regression analysis and data envelopment analysis as alternative methods for performance assessments. *Journal of Operational Research Society*, 44, 1129-1144.

Thanassoulis E, Boussofiane A, Dyson RG. (1995) Exploring output quality targets in the provision of prenatal care in England using DEA. *European Journal of Operational Research*, 80, 588-607.

Thanassoulis E, Boussofiane A, Dyson RG. (1996) A comparison of Data Envelopment Analysis as tools for performance assessment. *Omega, International Journal of Management Science*, 24, 229-244.

Thanassoulis E, Emrouznejad A. (1996) *Warwick Windows DEA*. Business School, University of Warwick.

Tulkens H, Vanden Eeckaut P. (1995) Non-parametric efficiency, progress and regress measures for panel data: Methodological aspects. *European Journal of Operational Research*, 80, 474-499.

Vinod HD. (1968) Econometrics of joint production. *Econometrica*, 36, 322-336.

Vinod HD. (1976) Canonical ridge and econometrics of joint production. *Journal of Econometrics*, 4, 147-166.

Whynes DK, Baines DL, Tolley KH. (1997) Prescribing costs in UK general practice: the impact of hard budget constraints. *Applied Economics*, 29, 393-399.

Table 1 Characteristics of regression and DEA

	Regression	DEA
Assumptions about production/cost frontier	Strong	None
Test assumptions about frontier	Yes	No
Assumptions about error distributions	Strong	None
Test distributional assumptions	Yes	No
Test for inclusion of variables	Yes	No
Distinguish random factors from efficiency variations	Yes	No
Allow for environmental factors	Yes	Yes
Allow for multiple outputs/multiple inputs	Only if canonical regression	Yes
Problems if multi-colinearity	Yes	No
Provide information on “peer” organisations	No	Yes
Vulnerable to small number of observations	Yes	Moderately
Vulnerable to endogeneity bias	Yes	Yes
Test for endogeneity bias	Yes	No

Table 2 Description and role of the variables

Name	Description of the variable in the analysis	Role in the Econometric analysis and DEA			Average	Min	Max	Standard deviation
		Model 1	Model 2	Model 3				
EXPEND	Gross expenditure [natural logarithm] in £000's on GMS corrected by the London rates and discounted to 1994, using the NHS pay and price deflator	Dependent Input	Dependent Input	- -	28,643 (£000's)	6,707	85,669	176
COST	Remuneration in £000's [natural logarithm] of GPs and practice nurses in the FHSA	- -	- -	Dependent Input	13,892 (£000's)	3,270	42,612	8,499
GP	Number [natural logarithm] of GPs for whom the FHSA is the responsible Committee	Independent Output	Independent Output	- Input	294	73	899	179
PATIENTS	Number [natural logarithm] of patients registered with a GP [and not classified as deprived] in the FHSA in '000	Independent Output	Independent Output	Independent Output	558 (in '000)	126	1,635	334
NURSES	Number [natural logarithm] of practice nurses (WTE)	- -	Independent Output	- Input	104	5	362	68
SMR	Number of deaths [natural logarithm of the standardised mortality rate] from all causes of FHSA residents aged 0-64	Independent Output	Independent Output	Independent Output	102	84	125	10
DEPRIVED	Number of registered [natural logarithm of the percentage of] patients living in wards which are classified as [non] deprived, based on the Jarman underprivileged area score (in '000)	- -	- -	Independent Output	10 (%)	0	77	13
AREA	Area [natural logarithm] of the FHSA in hectares	- -	- -	Independent Output	145,009 (in hectares)	3,384	831,297	185
WITHNURSE	Number of practices in the FHSA [natural logarithm of the percentage of] employing practice nurse	Independent Output	- -	- -	94 (%)	72	100	6
LOWLIST	GPs [natural logarithm of the percentage of] with less than 2,500 patients in their list	Independent Output	- -	- -	91 (%)	69	100	6
NOTSOLO	GPs [natural logarithm of the percentage of] not practising single-handed	Independent Output	- -	- -	88 (%)	65	99	8
IMM	Number of GPs [natural logarithm of the percentage of] who achieved the higher rate of payments for childhood immunisation in the FHSA	Independent Output	Independent Output	Independent Output	83 (%)	24	100	15
CER	Number of GPs [natural logarithm of the percentage of] who achieved the 80% target for cervical cytology in the FHSA	Independent Output	Independent Output	Independent Output	88 (%)	18	100	18
BOOST	GPs [natural logarithm of the percentage of] who achieved the higher rate of payments for pre-school boosters	- -	Independent Output	Independent Output	78 (%)	18	100	18
MINSTAN	Number of general practices [natural logarithm of the percentage of] which satisfied the minimum standard set out in para. 51.10 of the Statement of Fees and Allowances, excluding exemptions under para. 51.11	Independent Output	Independent Output	Independent Output	94 (%)	36	100	13
MSURG	Number of GPs [natural logarithm of the percentage of] who were on minor surgery list in the FHSA	- -	Independent Output	Independent Output	76 (%)	23	100	18

In square parentheses is the description of the variable used in the econometric analysis

Table 3 Efficiency measures: geometric mean and (standard deviation)

	Year 1			Year 2			Both years		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
COLS	90.195	91.464	86.825	89.729	90.214	88.594	-	-	-
	(3.81)	(3.89)	(4.73)	(4.12)	(4.12)	(4.45)	-	-	-
SFN	98.168	97.004	94.233	-	-	93.894	94.017	93.832	94.07
	(0.47)	(1.24)	(3.55)	-	-	(4.21)	(3.76)	(3.88)	(3.75)
SFE	98.235	97.543	97.619	-	-	87.199	-	-	-
	(0.65)	(1.53)	(1.79)	-	-	(3.47)	-	-	-
SFT	-	-	95.875	-	-	96.882	-	-	-
	-	-	(3.03)	-	-	(2.1)	-	-	-
CAN	-	-	81.316	-	-	80.101	-	-	-
	-	-	(6.61)	-	-	(6.95)	-	-	-
DEA CRS CE	91.68	94.098	97.503	91.04	93.417	97.142	90.411	92.659	96.392
	(5.49)	(4.58)	(3.18)	(5.38)	(4.8)	(2.87)	(5.32)	(4.57)	(3.08)
DEA VRS CE	95.838	96.638	98.422	95.129	95.623	98.317	94.869	95.513	97.766
	(4.47)	(3.95)	(2.57)	(4.7)	(4.43)	(2.18)	(4.55)	(4.12)	(2.44)
DEA CRS TE	-	-	98.738	-	-	98.841	-	-	98.406
	-	-	(2.12)	-	-	(1.92)	-	-	(2.07)
DEA VRS TE	-	-	99.257	-	-	99.356	-	-	99.002
	-	-	(1.57)	-	-	(1.35)	-	-	(1.55)

COLS: corrected OLS.
SFN: stochastic frontier, half normal.
SFE: stochastic frontier, exponential.
SFT: stochastic frontier, truncated at 0 normal.
CAN: canonical regression.
DEA CRS: DEA with constant return to scale.
DEA VRS: DEA with constant return to scale.
CE: cost efficiency.
TE: technical efficiency.

Table 4 Correlations among efficiency measures: Model 1

		SFN	SFE	DEA CRS	DEA VRS	***Year 2***
SFN	1	-	-	0.674	0.712	COLS
SFE	1	1	-	-	-	SFN
DEA CRS	0.562	0.562	0.562	-	-	SFE
DEA VRS	0.605	0.605	0.605	0.769	0.83	DEA CRS
Year 1	COLS	SFN	SFE	DEA CRS		

See notes in Table 3

Table 5 Correlations among efficiency measures: Model 2

		SFN	SFE	DEA CRS	DEA VRS	***Year 2***
SFN	1	-	-	0.641	0.613	COLS
SFE	1	1	-	-	-	SFN
DEA CRS	0.537	0.537	0.537	-	-	SFE
DEA VRS	0.566	0.566	0.566	0.823	0.854	DEA CRS
Year 1	COLS	SFN	SFE	DEA CRS		

See notes in Table 3

Table 6 Correlations among efficiency measures: Model 3

		SFN	SFE	SFT	CAN	DEA CRS CE	DEA VRS CE	DEA CRS TE	DEA VRS TE	***Year 2***
SFN	0.989	0.967	0.996	0.991	0.827	0.545	0.363	0.415	0.384	COLS
SFE	0.988	0.997	0.981	0.988	0.809	0.466	0.296	0.366	0.319	SFN
SFT	0.99	0.999	0.997	0.998	0.827	0.528	0.348	0.398	0.363	SFE
CAN	0.823	0.813	0.809	0.818	0.826	0.519	0.34	0.394	0.356	SFT
DEA CRS CE	0.457	0.429	0.446	0.44	0.268	0.42	0.21	0.36	0.299	CAN
DEA VRS CE	0.39	0.375	0.388	0.387	0.27	0.849	0.834	0.739	0.605	DEA CRS CE
DEA CRS TE	0.286	0.265	0.286	0.274	0.21	0.847	0.775	0.675	0.745	DEA VRS CE
DEA VRS TE	0.29	0.277	0.293	0.287	0.248	0.727	0.867	0.84	0.77	DEA CRS TE
Year 1	COLS	SFN	SFE	SFT	CAN	DEA CRS CE	DEA VRS CE	DEA CRS TE		

See notes in Table 3

Table 7 Correlations between years

Model 1		***Model 3***	
COLS	0.844	COLS	0.826
DEA CRS	0.862	SFN	0.818
DEA VRS	0.854	SFE	0.821
		SFT	0.826
Model 2		CAN	0.869
COLS	0.832	DEA CRS CE	0.713
DEA CRS	0.687.	DEA VRS CE	0.708
DEA VRS	0.762	DEA CRS TE	0.735
		DEA VRS TE	0.616

See notes in Table 3

Table 8 Correlations among alternative models

	COLS				**Stochastic frontier**		
	Model 2	Model 3	***Year 2***		Model 2	Model 3	***SFN year 1***
	0.977	-0.184	Model 1		0.91	-0.126	Model 1
Model 2	0.91	-0.191	Model 2	Model 2	0.91	-0.15	Model 2
Model 3	-0.149	-0.182		Model 3	-0.129	-0.156	
Year 1	Model 1	Model 2		***SFE year 1***	Model 1	Model 2	

	DEA CRS				**DEA VRS**		
	Model 2	Model 3 CE	***Year 2***		Model 2	Model 3 CE	***Year 2***
	0.8297	0.195	Model 1		0.875	0.286	Model 1
Model 2	0.803	0.019	Model 2	Model 2	0.816	0.266	Model 2
Model 3 CE	0.029	0.082		Model 3 OE	0.236	0.089	
Year 1	Model 1	Model 2		***Year 1***	Model 1	Model 2	

	Two years pooled				**SFN panel**		
	Model 2	Model 3 CE	***DEA VRS***		Model 2	Model 3	
	0.885	0.285	Model 1		0.982	-0.102	Model 1
Model 2	0.804	0.228	Model 2			-0.152	Model 2
Model 3 CE	0.165	0.163					
DEA CRS	Model 1	Model 2					

See notes in Table 3