Applied Economics, 2002, **34**, 1109–1119

Relative performance evaluation of the English acute hospital sector

ANDREW STREET* and ROWENA JACOBS

Centre for Health Economics, University of York, YO10 5DD, UK

Relative performance evaluation has been suggested as a means to overcome information asymmetry between regulators and organizations when assessing efficiency. By comparing similar organizations the relationship between costs and effort can be better isolated. The English Department of Health (DoH) has undertaken relative performance evaluation in comparing the unit costs of acute hospitals using ordinary least squares (OLS) methods. After adjusting for exogenous influences in costs, residual unexplained cost differences are deemed to represent inefficiency. This paper questions the official interpretation of the OLS residuals. The OLS model is re-estimated to calculate confidence intervals around the residuals and as a stochastic cost frontier (SCF). It is concluded that English acute hospitals exhibit less in efficiency than is implied by official estimates.

I. INTRODUCTION

Promoting public sector efficiency remains an important concern for many governments. Lacking competitive pressures, traditionally it has been held that the public sector has little inherent incentive to pursue efficient behaviour. In recent years, governments have sought to encourage efficiency by simulating the effects of competition. For instance, in the UK, an 'internal market' was introduced in the National Health Service (NHS) in 1991. Under this arrangement, separate institutions are assigned managerial responsibility for the functions of supply and demand. The production of services becomes the sole concern of providers, such as hospitals and nursing homes. Demand for services is expressed by budget holding health authorities and general practitioners, commissioned to secure health improvements for their resident populations by buying health services. The mainstay of the internal market is the process of competitive tendering, or contracting, designed to promote competition among providers. The contractual process is supposed to force providers continually to seek to improve their productive processes, the more efficient winning more contracts by submitting bids of lower cost and/or higher quality than their competitors.

However, it is now no longer widely held that the competitive pressures of the internal market are sufficient to encourage efficiency gains among hospitals, most of which enjoy local monopoly power. Instead, the rhetoric of 'competition' has been replaced by a desire to encourage 'co-operation' in the health care sector (NHS Executive, 1997). Co-operative behaviour may encourage hospitals to avoid duplication of services and to secure cost reductions through exploitation of economies of scale and scope but, equally, co-operation could lead to collusion and inefficient behaviour (Goddard and Mannion, 1998). Recognizing this possibility, the English Department of Health (DoH) has adopted a more direct mechanism to promote efficiency gains. The official method is to undertake relative performance evaluation, whereby each hospital's costs are compared to those observed in other hospitals. This policy accords with the theory of yardstick competition (Schleifer, 1985) or of benchmarking (Dopuch and Gupta, 1997). Indeed, the government first announced the policy as a benchmarking exercise (NHS Executive, 1997). Yardstick competition or benchmarking is of particular relevance in contexts where regulators are poorly informed about the specific conditions facing each organization. Theory suggests that comparison across organiza-

* Corresponding author: email ads6@york.ac.uk

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0003684011008810 0

tions will lead to more accurate assessment of the relationship between observed costs and effort. Information asymmetry can be reduced by observing costs in similar organizations. If organizations are heterogeneous, it is argued that multiple regression techniques can be used to control for exogenous influences in the operating conditions facing each organization (Schleifer, 1985).

This type of approach has been adopted in England, where the unit costs of acute hospitals are compared after correcting for exogenous influences using Ordinary Least Squares (OLS) methods. This paper assesses the econometric approach adopted, placing particular attention on the residual from the OLS model. The residual measures the difference between observed and expected costs and is interpreted by the DoH as evidence of inefficiency. This official interpretation is challenged in Section II, firstly, by considering the uncertainty that pertains to econometric estimation and, secondly, by reformulating the average cost function as a cost frontier. The implications of these challenges to the official interpretation of the residual are illustrated in section III, where the OLS model is re- estimated. This section reports confidence intervals around the OLS residuals and the results of using maximum likelihood techniques to estimate a stochastic cost frontier (SCF). The results are discussed in Section IV.

II. USING RESIDUALS AS INEFFICIENCY ESTIMATES

In general terms, the econometric model used by the DoH is defined as:

$$y_i = \alpha + \mathbf{x}_i\beta + \varepsilon_i \quad i = 1, \ldots, N \tag{1}$$

where y_i is the (total or unit) cost of production of the ith hospital in either linear or logarithmic form; α is a constant; $\mathbf{x}_i$ is a vector of (transformations of the) input prices and output of the ith hospital; β a vector of unknown parameters; and ε_i is the error term. Under this formulation the error term is interpreted as representing inefficiency. In other words, the difference (residual) between observed unit costs and those predicted by the model is said to be due to inefficient behaviour (Feldstein, 1967). A hospital with a residual of zero is interpreted as displaying average efficiency, while those with negative (positive) residuals are deemed of above (below) average efficiency.

This interpretation is opposite to that drawn when observing firms in a competitive environment. In this context, firms are forced to operate at minimum cost levels and standard OLS residuals would be interpreted as being due solely to random influences or measurement error. Public hospitals in England do not face substantial competitive pressures to operate at cost minimizing levels and their ability to reduce costs is compromised by a variety of political constraints that limit opportunities for input substitution, restructuring and investment. Limited competitive pressure and unmeasured political constraints will be captured in the residual term. However, inefficiency may not be the sole explanation if at least some of the difference between observed and predicted costs is due to random 'noise' arising from measurement error, omitted variables or unobservable heterogeneity.

In this paper two approaches have been adopted to explore whether the OLS residual should be interpreted as being indicative of inefficiency alone. The first approach applies statistical techniques to assess uncertainty surrounding the measurement of those factors deemed important influences in explaining unit costs. The second approach uses maximum likelihood techniques to partition the residual into inefficiency and error components.

Statistical analysis

One of the complaints about using the OLS residuals as inefficiency estimates is that they fail to provide any indication of the precision of the estimate. This contrasts with the publication of clinical data in England where, for each dimension of clinical performance, data at hospital level are accompanied by 95% confidence intervals (NHS Executive, 1999a). It is less conventional to report confidence intervals around residuals from a least squares regression, particularly when performed on non-hierarchical cross-sectional data, as is the case for the DoH model.[1] Nevertheless, it is possible to compute the variance around the residual for each hospital (Maddala, 1988). Consider a model with no constant term:

$$\mathbf{y} = \mathbf{x}_i\beta + \mathbf{e} \tag{2}$$

where $\mathbf{y}$ is the vector of n observations on the dependent variable; $\mathbf{x}_i$ is a vector of k explanatory variables; β is a vector of unknown parameters; and $\mathbf{e}$ is a vector of unobservable errors, with zero mean and constant variance.

The model yields predictions ($\hat{\mathbf{y}}$) from the least squares regression

$$\hat{\mathbf{y}} = \mathbf{x}_i\hat{\beta} \tag{3}$$

where $\hat{\beta}$ is a vector of estimated coefficients. $\hat{\mathbf{y}}$ has a variance due to the variance of the estimated coefficients. This reflects the prediction error due to uncertainty about what the parameter estimates (β) actually are. This uncertainty can be expressed as:

[1] Techniques are available for calculating confidence intervals when data are hierarchical or when there are repeat observations over time (Goldstein and Spiegelhalter, 1996; Marshall and Spiegelhalter, 1998).

$$\text{var}(\hat{\mathbf{y}}) = \text{var}(\mathbf{x}_i\hat{\beta}) = s^2 h_i \quad (4)$$

where s^2 is the mean square error of the regression and $h_i = \mathbf{x}_i(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i^T$, $\mathbf{X}$ is a $n \times k$ matrix of observations on the independent variable, $\mathbf{X}^T$ is the transpose of this matrix, and $(\mathbf{X}^T\mathbf{X})^{-1}$ is the inverse of their product. h_i is a diagonal element of the projection matrix and is commonly termed the 'hat matrix'. As an example, in the single variable case, this gives an indication of the proximity of the ith observation to the mean of the other observations divided by the sum of the deviation for all observations from the mean.

From 4 the standard error of the prediction can be derived:

$$s_{p_i} = s\sqrt{h_i} \quad (5)$$

Although the errors ($\mathbf{e}$) are unobservable, the model yields residuals, (ε), such that $\mathbf{y} = \hat{\mathbf{y}} + \varepsilon$. Moreover, $\text{var}(\mathbf{y}) = \text{var}(\hat{\mathbf{y}}) + \text{var}(\varepsilon) = s^2 h_i + s^2$. It is possible to decompose s^2 into prediction and residual variances:

$$s^2 = s^2 h_i + s^2(1 - h_i) \quad (6)$$

where $s^2(1 - h_i)$ is the residual variance. Thus the standard error of the residual is:

$$s_{r_i} = s\sqrt{(1 - h_i)} \quad (7)$$

Confidence intervals can be calculated once this residual variance is known.

Stochastic cost frontier analysis

A more conventional approach to the problem of interpreting the regression residual as indicating inefficiency alone is to ascertain whether the OLS residual can be partitioned into two components, one of which represents inefficiency and the other random noise. Evidence of skewness in the OLS residuals is taken to suggest inefficiency within the sample (Schmidt and Lin, 1984). If residuals are distributed symmetrically around the mean all cross-sample variation is interpreted as being attributable to statistical noise (Wagstaff, 1989).

Stochastic cost frontier analysis has been developed as a means to explore the extent of inefficiency. The approach decomposes the residual term from the estimated cost function into two parts with zero covariance:

$$\varepsilon_i = v_i + u_i \quad \text{cov}(v_i, u_i) = 0 \quad i = 1, \ldots, N \quad (8)$$

The dual specification of the residual is defended on the grounds that each component reflects an economically distinct disturbance (Aigner *et al*., 1977). v_i can be interpreted as representing stochastic (random) events not under control of the firm, such as climatic conditions, luck (Aigner *et al*., 1977), random equipment failure (Greene, 1993), or errors in identifying or measuring independent variables (Timmer, 1971). u is a non-negative error term accounting for the cost of inefficiency in production. Under this specification the u_i defines how far the ith hospital operates above the cost frontier. If allocative efficiency is assumed, the u_i are closely related to the cost of technical inefficiency. Technical inefficiency may arise from managerial slack (X-inefficiency), outmoded equipment, or inadequate capital stock. If the assumption of allocative efficiency is not made, the u_i in a cost function incorporates both technical and allocative inefficiencies.[2]

The stochastic cost function is then written as (Coelli, 1996):

$$y_j = \alpha + \mathbf{x}_i\beta + (v_i + u_i) \quad i = 1, \ldots, N \quad (9)$$

In estimating this function, it is necessary to specify the distributional characteristics of the two components of the residual. Commonly, v_i is assumed to be independent and identically distributed (i.i.d.) with zero mean and variance σ_v^2, hence $v_i \sim N(0, \sigma_v^2)$. Various distributions have been proposed for u_i, which must be observed indirectly since the residual $(y_i - \alpha - \mathbf{x}_i\beta)$ estimates ε_i, not u_i. The entire $\varepsilon_i (= v_i + u_i)$ can be estimated easily for each observation but for a long time an unresolved problem was how to separate it into its two components, v_i and u_i. Estimation of stochastic production and cost functions were accomplished by several groups of authors (Aigner *et al*., 1977; Schmidt and Lovell, 1979), but because they were unable to solve the decomposition problem, none were able to obtain estimates of efficiency for each firm in the sample.

A solution to this problem was first proposed by Jondrow *et al*. (1982) by considering the expected value of u_i, conditional on $(v_i + u_i)$.[3] They specified the functional form of the distribution of the one-sided inefficiency component and derived the conditional distribution $(u_i | v_i + u_i)$. The resulting residuals are decomposed to estimate the efficiency for each observation. Jondrow *et al*. (1982) proposed a model, assuming a half-normal distribution for u_i, in which the conditional expectation for inefficiency was defined as follows:

[2] Any failure in optimization, whether technical or allocative, will show up as higher cost. The computation is dependent on the inputs chosen and whether they are allocatively efficient. Thus, a producer may be operating technically efficiently by a production function, but show up as inefficient with respect to a cost function. Therefore it has been argued that the interpretation of the one-sided error on the cost side as a measure of technical inefficiency is only appropriate if the measure is defined in terms of costs, rather than output. This implies that efficiency is best measured by costs rather than outputs (Greene, 1993). Consequently, inefficiency is often interpreted as 'cost inefficiency', the total of both technical and allocative inefficiency.

[3] Their solution was for the production function formulation of a stochastic frontier model with an error term $(v_i - u_i)$.

$$E[u_i|\varepsilon_i] = \frac{\sigma\lambda}{(1+\lambda^2)}\left[\frac{\phi(\varepsilon_i\lambda/\sigma)}{\Phi(-\varepsilon_i\lambda/\sigma)} - \frac{\varepsilon_i\lambda}{\sigma}\right] \tag{10}$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$. $\lambda = \sigma_u/\sigma_v$ and captures inefficiency. Where $\lambda = 0$, every observation would lie on the frontier (Greene, 1993). $\phi(.)$ and $\Phi(.)$ are, respectively, the probability density function and cumulative distribution function of the standard normal distribution.

Truncated normal and exponential distributions have also been proposed. The truncated-normal model is a more general form of the half-normal, where u is distributed with a modal value of μ (Stevenson, 1980). The explicit form for the conditional expectation is obtained by replacing the $\varepsilon_i\lambda/\sigma$ in the half-normal model with:

$$u_i^* = \frac{\varepsilon_i\lambda}{\sigma} + \frac{\mu}{\sigma\lambda} \tag{11}$$

If μ is not significantly different from zero, the model collapses to the half-normal.

If an exponential distribution is imposed, with a density function of the general form $f(u_i) = \theta \exp^{-\theta u_i}$, the conditional expectation is expressed as (Greene, 1995):

$$E[u_i|\varepsilon_i] = (\varepsilon_i - \theta\sigma_v^2) + \frac{\sigma_v\phi[(\varepsilon_i - \theta\sigma_v^2)/\sigma_v]}{\Phi[(\varepsilon_i - \theta\sigma_v^2)/\sigma_v]} \tag{12}$$

in which θ is the distribution parameter to be estimated.

The formulations for the various error distributions are as follows (Greene, 1995):

Normal–half-normal model:

$$E[u] = \sqrt{\frac{2}{\pi}}\sigma_u \tag{13}$$

$$\mathrm{Var}[u] = \left(\frac{\pi}{2} - 1\right)\sigma_u^2 \tag{14}$$

Normal–truncated model:

$$E[u] = \mu + \sigma_u\lambda_u \tag{15}$$

$$\mathrm{Var}[u] = \sigma_u^2\left[1 - \lambda_u\left(\frac{\mu}{\sigma_u} + \lambda_u\right)\right] \tag{16}$$

Normal–exponential model:

$$E[u] = \frac{1}{\theta} \tag{17}$$

$$Var[u] = \frac{1}{\theta^2} \tag{18}$$

These formulations produce an unbiased but inconsistent estimator of u_i because, regardless of the sample size, the variance of the estimate remains non-zero (Greene, 1993). The inconsistency of the estimator u_i is unfortunate in view of the fact that the purpose of the estimation is to approximate inefficiency. However, no improvements on this measure have yet been forthcoming in the literature for single-equation cross-sectional studies. The problem can be avoided if longitudinal data are available (Greene, 1993).

This paper uses two econometric estimation packages, namely LIMDEP (Greene, 1995) for the exponential distribution and FRONTIER (Coelli, 1996) for the half-normal and truncated error distributions. Predictions of individual hospital efficiencies are computed automatically in FRONTIER. The measure of efficiency relative to the cost frontier is defined as: $EFF_i = E(y_i|u_i, \mathbf{x}_i)/E(y_i|u_i = 0, \mathbf{x}_i)$, where y_i is the cost of the ith firm. EFF_i will take a value between one and infinity and can be defined as: $(\mathbf{x}_i\beta + u_i)/(\mathbf{x}_i\beta)$. In order to scale these scores to be comparable to the cost indices, the reciprocal is obtained such that $0 < 1/EFF_i < 1$.

III. COMPARISON OF EFFICIENCY RANKINGS

Data and model

The current paper reproduces analysis undertaken for the English Department of Health, the results of which were circulated to acute hospitals in 1999 (Audit Commission and Department of Health, 1999). The first stage of the analysis involved calculating an index of actual to expected costs for each hospital (Söderlund and van der Merwe, 1999). This index was then regressed against a series of independent variables that sought to explain variations in index scores. Data were compiled for the year 1995–1996 from a variety of routine sources, including the Hospital Episodes Statistics, Trust Financial Returns, Trust Annual Accounts and the Hospitals Yearbook.

The dependent variable, the Casemix Cost Index (CCI), for hospital i is an index of actual over expected casemix weighted costs:

$$\mathrm{CCI}_i = \frac{\mathrm{C}_i}{[(\mathrm{IC} \times \mathrm{IP}_i \times \mathrm{H}_i)/\mathrm{IP} \times \mathrm{H}] + [\Sigma_j \mathrm{OP}_{ij} \times (\mathrm{OC}_j/\mathrm{OP}_j)] + [\mathrm{AE}_i \times \mathrm{AC}/\mathrm{AE}]} \tag{20}$$

where C_i is the cost of inpatient, outpatient and A&E care in hospital i; IC is the total cost of inpatient spells for all acute hospitals; IP_i is the number of inpatient (including day case) spells in hospital i; H_j is the Healthcare Resource Group (HRG) index for hospital i, this being a measure of casemix complexity (Benton *et al.*, 1998); IP is the total number of inpatient spells for all acute hospitals; H is the average casemix index for all acute hospitals; OP_{ij} is the number of first outpatient attendances across all specialties in hospital i; OC_j is the total cost of outpatient attendances for all acute hospitals in specialty j; OP_j is the number of first outpatient attendances for all acute hospitals in specialty j; AE_i is the number of first A&E attendances in hospital i; AC is the total cost of A&E attendances in all acute hospitals; and AE is the number of first A&E attendances in all acute hospitals.

The DoH efficiency estimates are derived from the following estimated equation:

$$CCI_i = \alpha + \mathbf{x}_i\beta + \varepsilon_i \quad i = 1, \ldots, 217 \qquad (21)$$

where the residual ε_i provides the DoH index of efficiency scores for all hospitals. This index was termed the Casemix Costliness Cost Index (2CCI).

Many of these variables included in the vector $\mathbf{x}$ attempt to account for the possibility that, even within an HRG, some hospitals will treat more costly patients. Hospital transfers, multiple-episode spells, and the proportion of elderly or female patients are included to account for cost differences over and above the HRG casemix adjustment. In addition, allowance is made for possible cross-subsidization between patient care and teaching or research which may not be adequately dealt with in the funding allocations, and for differences in local factor costs, assessed using the Market Forces Factor.

The explanatory variables are listed in Table 1 with basic descriptive statistics, for 217 hospitals. These data exclude outliers.[4]

Statistical analysis

Summary statistics relating to the variance around the OLS residuals from the 2CCI model are shown in Table 2, where individual hospitals are clustered into similar 'family' groups (Audit Commission and Department of Health, 1999). $\mathrm{Var}(\varepsilon_i)$ summarizes the uncertainty around the efficiency scores for each group of hospitals. It might be expected that uncertainty about the efficiency scores of (say) specialist hospitals should be greater than that for other hospitals, perhaps because HRGs fail to capture adequately their more specialized casemix. However, the data do not reflect this, with the mean standard deviation around the residuals for specialist hospitals (0.091) being less than for all hospitals (0.104). The reason is that the efficiency score is derived from the residuals, and the uncertainty around the residuals stems directly from the relationship between $\mathrm{var}(\hat{y}_i)$ and $\mathrm{var}(\hat{e}_i)$. Since $\mathrm{var}(\hat{y}_i) = s^2 h_i$ and $\mathrm{var}(\varepsilon_i) = s^2(1 - h_i)$, predicted values for observations further from the regression line will have relatively large variances and their corresponding residuals will have relatively small variances (Cook and Weisberg, 1982).

Figure 1 presents the point estimates of the efficiency scores and confidence intervals surrounding these estimates for each hospital, ordered according to the estimated efficiency score. The scale is centred at 1, this indicating the average efficiency score for all English NHS acute hospitals. The confidence interval when making comparisons across all hospitals is calculated as $\varepsilon_i \times 3.68 \times s_{r_{ij}}$. This critical level was calculated using the Šidák adjustment and preserves a 95% significance level when making multiple comparisons (Sidak, 1967). The diagram shows that con-

Table 1. *Descriptive statistics*

Variable key	*Description*	*Mean*	*Std Dev*	*Minimum*	*Maximum*
Dependent variable					
CCI	Casemix cost index	0.992	0.146	0.715	1.724
Independent variables					
TRANSIPP	Transfers into hospital per spell	0.016	0.030	0.000	0.241
TRANSOPP	Transfers out of hospital per spell	0.021	0.015	0.000	0.125
EMERGPP	Emergency admissions per spell	0.346	0.087	0.020	0.549
FCEINPP	Finished consultant episode inter-specialty transfers per spell	0.020	0.014	0.000	0.114
OPNP	Non-primary outpatient attendances per inpatient spell	2.923	0.833	0.797	7.847
EMERINDX	Standardized index of unexpected emergency admissions/total emergency admissions	0.058	0.015	0.016	0.150
EP_SPELL	Episodes per spell	1.068	0.118	0.785	1.661
HRGWTNHS	HRG weight, case mix index	93.996	20.722	72.018	242.028
PROP15U	Proportion of patients under 15 years of age	0.094	0.101	0.000	0.838
PROP60P	Proportion of patients 60 years or older	0.340	0.083	0.000	0.590
PROPFEM	Proportion of female patients	0.572	0.056	0.308	0.897
STUDENPP	Student whole time teaching equivalents per inpatient spell	0.001	0.001	0.000	0.012
RESEARPC	Percentage of total revenue spent on research	1.750	6.090	0.000	73.065
MFF_COMB	Market forces factor – weighted average of staff, land, buildings and London weighting factors	87.479	9.902	75.817	132.789

[4] Outliers were identified using the DFITS procedure, and applying a cut-off point where DFITS $> 3 \times (k/n)^{0.5}$ where k is the number of parameters estimated and n the number of observations (Söderlund and van der Merwe, 1999).

Table 2. *Residuals (ε_i) and variance ($var(\varepsilon_i)$) from the 2CCI, by family group*

2CCI		ε_i				Var(ε_i)			
Family Group	n	Mean	SD	Min	Max	Mean	SD	Min	Max
Small/medium acute hospitals, outside London	29	−0.016	0.085	−0.159	0.142	0.107	0.001	0.101	0.108
Large acute hospitals, outside London	44	−0.029	0.103	−0.254	0.222	0.107	0.001	0.101	0.108
Very large acute hospitals, outside London	41	−0.022	0.081	−0.201	0.252	0.106	0.004	0.083	0.108
Acute hospitals, London	19	−0.013	0.083	−0.181	0.176	0.104	0.008	0.073	0.107
Specialist hospitals	16	0.013	0.115	−0.167	0.275	0.091	0.011	0.056	0.104
Acute teaching hospitals	27	0.006	0.130	−0.189	0.345	0.102	0.006	0.083	0.107
Small/medium/large multi-service hospitals	25	0.033	0.106	−0.153	0.224	0.106	0.004	0.088	0.108
Very large multi-service hospitals	16	0.098	0.100	−0.032	0.270	0.107	0.001	0.105	0.108
All acute hospitals	*217*	*0.000*	*0.105*	*−0.254*	*0.345*	*0.104*	*0.006*	*0.056*	*0.108*

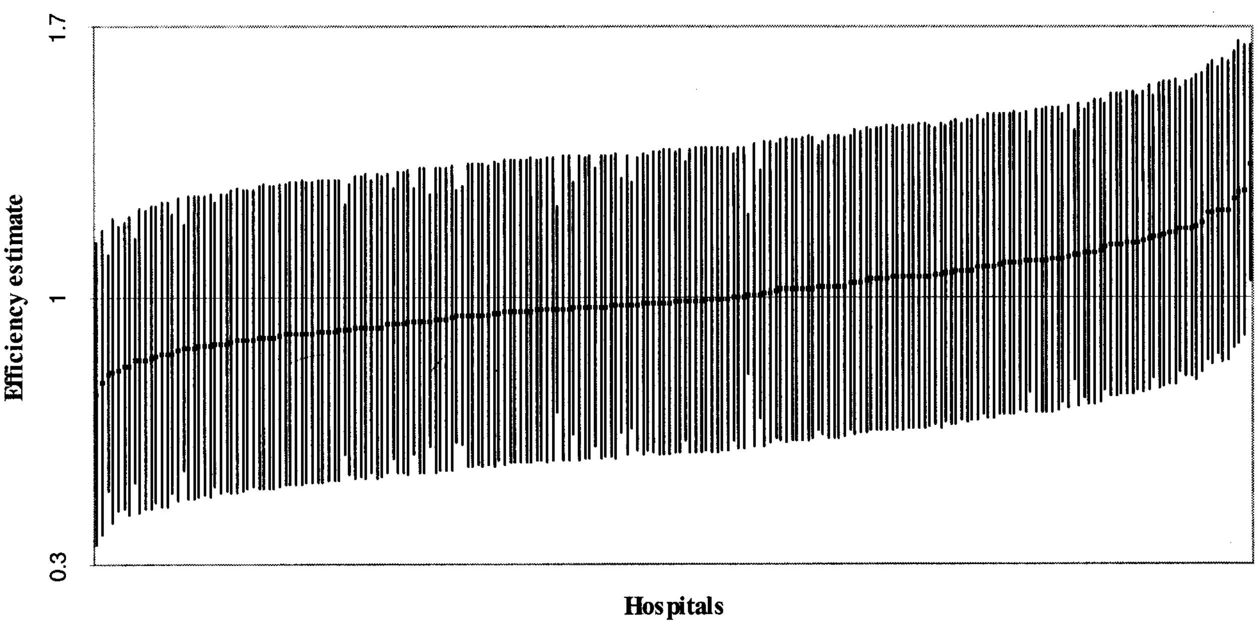

Fig. 1. *Multiple comparison 95% confidence intervals*

fidence intervals overlap along the entire series. This would suggest that no significance can be attached to the mean differences in unit costs observed among hospitals. In other words, it cannot confidently be claimed that the hospitals that appear to be the most efficient are actually any more efficient than those hospitals that appear to be the least efficient.

Stochastic frontier analysis

Table 3 presents the original OLS regression results for the 2CCI, as estimated by the DoH. The model implies a mean level of 'inefficiency' of around 74%. The coefficient of skewness, $\sqrt{b_1}$, suggests that the residuals are significantly skewed in the direction expected for a cost function. The OLS results are accompanied by the three stochastic cost frontier regression estimations, corresponding to the half-normal, truncated normal and exponential error distributions. The distribution parameters of both the half-normal and exponential models (λ and θ respectively) are significant, suggesting that these models are an improvement to OLS estimation. In contrast, the truncated-normal model yields a value for μ that is not significantly different from zero, making it equivalent to the half-normal model.[5]

[5] Moreover, μ is estimated as taking a negative value. This is contrary to the requirement for a cost function.

Table 3. *Results for alternative specifications of the 2CCI, 1995–1996 data*

Dependent variable: CCI

Model	*Ordinary Least Squares regression*			*Stochastic frontier regression – half normal error distribution*			*Stochastic frontier regression – truncated normal error distribution*			*Stochastic frontier regression – exponential error distribution*		
	Coeff	Std Err	*t*-ratio	Coeff	Std Err	*t*-ratio	Coeff	Std Err	*t*-ratio	Coeff	Std Err	*t*-ratio
INTERCEPT	0.66	0.25	2.65	0.58	0.22	2.61	0.56	0.23	2.48	0.53	0.22	2.47
TRANSIPP	1.09	0.42	2.58	1.39	0.37	3.79	1.41	0.36	3.94	1.42	0.59	2.39
TRANSOPP	−2.92	0.64	−4.58	−2.73	0.59	−4.66	−2.70	0.57	−4.73	−2.69	0.64	−4.21
EMERGPP	0.11	0.12	0.88	0.08	0.11	0.75	0.12	0.13	0.94	0.17	0.14	1.29
FCEINPP	1.06	0.58	1.82	1.37	0.53	2.59	1.35	0.52	2.61	1.32	0.71	1.85
OPNPP	0.06	0.01	5.62	0.05	0.01	5.24	0.05	0.01	4.49	0.05	0.01	4.09
EMERINDX	−0.58	0.59	−0.98	−0.56	0.45	−1.25	−0.63	0.46	−1.37	−0.71	0.81	−0.88
EP_SPELL	0.18	0.07	2.55	0.12	0.07	1.67	0.10	0.07	1.35	0.09	0.07	1.25
HRGWTNHS	0.00	0.00	2.68	0.00	0.00	2.11	0.00	0.00	2.12	0.00	0.00	2.13
PROP15U	−0.21	0.12	−1.79	−0.17	0.09	−1.92	−0.15	0.10	−1.53	−0.13	0.08	−1.58
PROP60P	−0.43	0.15	−2.97	−0.28	0.13	−2.19	−0.24	0.15	−1.63	−0.20	0.15	−1.37
PROPFEM	−0.38	0.22	−1.75	−0.26	0.19	−1.35	−0.20	0.22	−0.87	−0.12	0.20	−0.63
STUDENPP	12.42	7.72	1.61	13.58	7.69	1.77	15.16	7.26	2.09	16.39	10.26	1.60
RESEARPC	0.00	0.00	−1.17	0.00	0.00	−0.25	0.00	0.00	0.03	0.00	0.00	0.19
MFF_COMB	0.00	0.00	2.22	0.00	0.00	2.25	0.00	0.00	2.00	0.00	0.00	1.80
μ				μ is restricted to be zero			−0.28	0.84	−0.33			
θ										11.60	2.03	5.70
λ				2.74	0.46	5.96	1.63	0.74	2.19			
$p(\sqrt{b_1})$		0.000										
σ_v^2					0.0029			0.0048			0.0042	
σ_u^2					0.0219			0.0128			0.0074	
Log likelihood		184.7112			189.9315			190.1538			189.9361	
R^2		0.4965										
Adjusted R^2		0.4616										
$E[u]$					0.1181			−0.0957			0.0862	
Var$[u]$					0.0125			0.0304			0.0074	
Mean efficiency		0.7415			0.8857			0.9034			0.9196	

Ignoring the truncated-normal results, the coefficients and the significance of the independent variables are broadly similar across all specifications. This is to be expected, since both the OLS estimates (which provide the starting values for the iterations) and the maximum likelihood estimates used in the SCF regressions are consistent estimators (Greene, 1993). The main difference is that fewer variables appear significant under the exponential specification.

Table 3 indicates that the mean level of efficiency is similar across the SCF error distributions, at around 90%. This is considerably higher than was implied by the OLS model. This is expected because the OLS residual comprises both inefficiency and noise and, hence, yields lower estimates of efficiency (Schmidt, 1985).

For the SCF models, in each case the one-sided component of the error, σ_u^2, is larger than the symmetric component, σ_v^2. The ratio of the inefficiency component to total variance can be used to separate the influence of inefficiency from that of noise. If hospitals were fully efficient, all error variance would be due to noise. For the half normal model, the contribution of u_i to total variance is defined as (Greene, 1993):

$$u_i = \frac{[(\pi/2) - 1]\sigma_u^2}{\{\sigma_v^2 + [(\pi/2) - 1]\sigma_u^2\}} \tag{22}$$

where $[(\pi/2) - 1]\sigma_u^2$ is the variance of u under the half normal specification (from Equation 14). The amount of unexplained variation attributed to inefficiency in the half-normal specification is 81%, with statistical noise accounting for around 19%. The exponential model suggests that inefficiency is less prevalent, amounting to 64% of total variance.

A graphical representation of each of the distributions of efficiency scores derived from the four specifications is presented in Fig. 2. These figures are scaled such that estimated efficiency increases along the horizontal axis. Fig. 2a, derived from the OLS model, can be interpreted as suggesting that few hospitals are efficient, with inefficiency distributed not quite normally among the sample. The alternative specifications derived from the SCF models

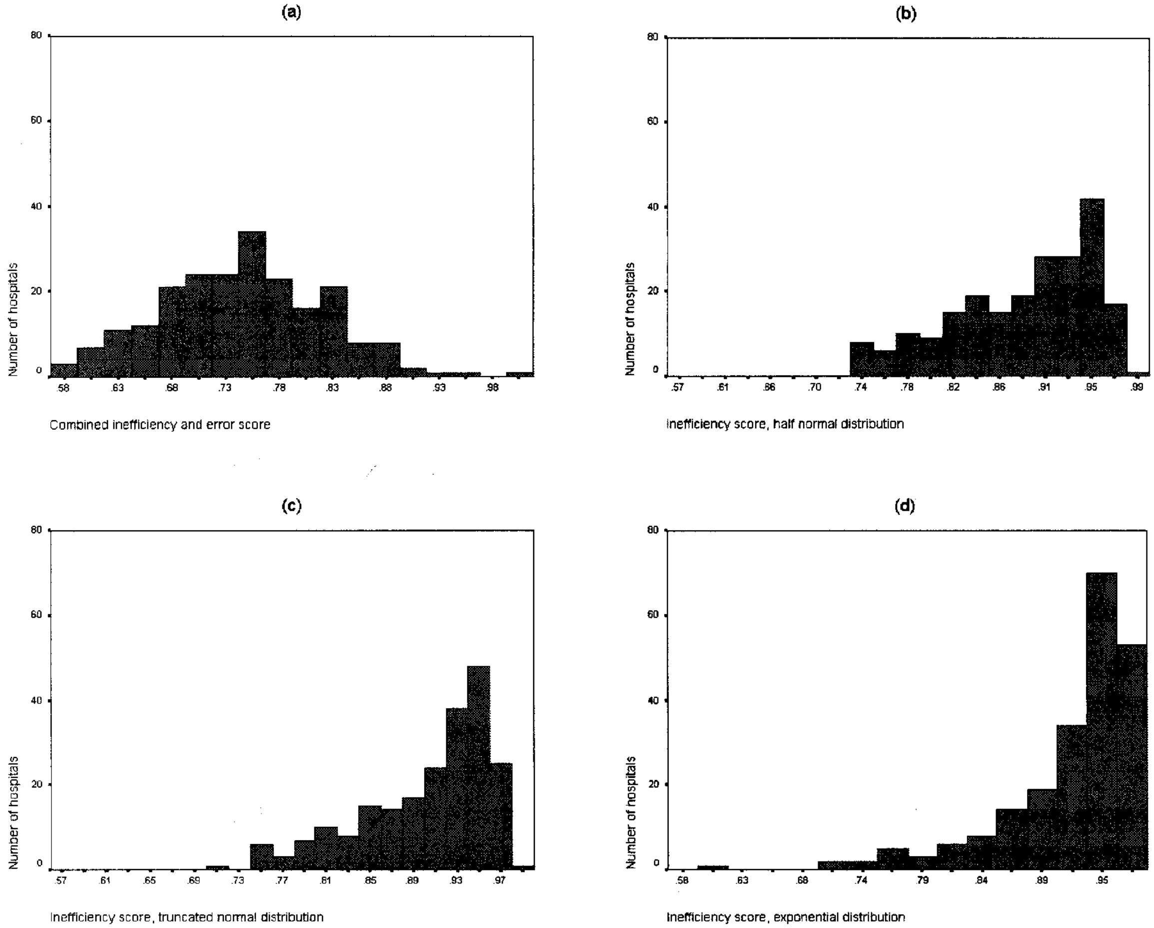

Fig. 2. *Histogams showing (a) the residual from the OLS model; (b) the inefficiency term from the half-normal model; (c) the inefficiency term from the truncated model; (d) the inefficiency term from the exponential model*

imply that most hospitals are (or are close to being) relatively efficient as they are clustered to the right hand side of the distribution. The exponential distribution (Fig. 2d), with its more pronounced negative skew, implies that inefficiency is less widespread among hospitals than is assumed under the alternative specifications. This is indicated by the variance of u being very small (0.0074) under this specification.

These graphs suggest that conclusions about relative performance may be sensitive to the measurement technique employed (Berger, 1993; Cummins and Zi, 1997). Comparison across specifications is made by considering differences in the estimated efficiency scores for each hospital and how hospitals appear relative to one another in a ranking of their performance. Table 4 shows the results for the correlation matrix of the efficiency scores and ranks from the OLS model and the SCF estimations.

There is almost perfect correlation among the estimates produced by the OLS and SCF models, with ranked scores displaying slightly higher correlations than the raw scores. Hence, conclusions about relative efficiency remain stable regardless of the error distribution that is imposed. This implies that an OLS model might be appropriate for ordering of hospitals. However, because the OLS model overestimates the prevalence of inefficiency, it should not be relied upon for target setting.

IV. DISCUSSION

The official interpretation applied to the residuals from an OLS model estimating unit costs across English acute hospitals is that unexplained cost differences are attributable to inefficient behaviour. This interpretation has been challenged in two ways. Firstly, uncertainty about residual values has been assessed by applying similar statistical criteria to those adopted by the DoH when making comparisons across other dimensions of performance (NHS Executive, 1999a; NHS Executive, 1999b). Secondly, economic methods have been used to reformulate the average cost function as a cost frontier.

Table 4. *Correlations of OLS and SCF estimates of efficiency*

	OLS	SCF-half	SCF-trunc	SCF-exp
SCORES				
OLS	1.000			
SCF-half	0.949	1.000		
SCF-trunc	0.925	0.992	1.000	
SCF-exp	0.852	0.918	0.958	1.000
RANKS				
OLS	1.000			
SCF-half	0.984	1.000		
SCF-trunc	0.980	0.998	1.000	
SCF-exp	0.974	0.981	0.987	1.000

Note: SCF-half, SCF-trunc and SCF-exp are, respectively, half normal, truncated normal and exponential error distributions of stochastic cost frontier regressions on 2CCI variables.

Both approaches imply that inefficiency among hospitals is less prevalent than suggested by the OLS model. The conclusion drawn from the statistical analysis is that, at 95% confidence intervals, the observed differences in 'efficiency' among hospitals are not statistically significant, suggesting that hospitals might be equally (in)efficient – or, at least, that the model has been unable to determine where differences in efficiency might lie.

The SCF approach implies that the mean level of efficiency is around 90%, compared to only 74% as implied by the OLS model. However, in contrast to the statistical analysis, the SCF approach suggests that there is variation in efficiency among hospitals. This variation is evident in the tail of less efficient hospitals shown in Figs 2(b) and 2(d). It may be possible to reconcile the different conclusions deriving from the statistical and SCF approaches. One way to achieve reconciliation would be to require a less stringent 'standard of proof' when undertaking statistical comparisons. Reducing the required significance level to less than the (arbitrary) 95% level would result in shorter confidence intervals, some of which may no longer overlap. Then it may be concluded that differential efficiency probably exists – although the probability is lower than it is when a 95% confidence interval is imposed.

Another way to move towards comparability of the results produced by the statistical and SCF approaches is to question the interpretation commonly placed on each component of the partitioned error term in stochastic frontier models. Traditionally, the decision about whether to estimate a stochastic frontier model on cross-sectional data hinges on whether or not the OLS residual is skewed (Schmidt and Lin, 1984; Skinner, 1994). The absence of a skewed residual, according to this convention, indicates that there is no cross sample variation in efficiency. But in drawing conclusions about relative efficiency, this econometric approach relies on two strong assumptions. These are, firstly, that the econometric model is correctly specified and, secondly, that neither inefficiency nor noise are subject to spillover across the partitioned error term.

While the importance of model specification is widely appreciated, the latter assumption is less frequently challenged, despite the distributional assumptions being arbitrary and not easily defended (Schmidt, 1985). In most cross sectional estimations of stochastic frontier models it is assumed that, as in OLS models, v_i is independently and identically distributed as $N(0, \sigma_v^2)$. Kopp and Mullahy (1990) describe a method of moments based approach in which the assumption of normality is relaxed in favour of a partially distribution-free approach where $f_v(v_i)$ is assumed symmetric around zero. Unfortunately, generality comes at the expense of an inability to estimate u_i (Greene, 1993), so this distribution-free method has not been widely applied.

Interpreting the non-negative term as comprising inefficiency alone is, perhaps, more contentious as it raises the possibility that some organizations will be stigmatized as being less efficient than they actually are. A non-negative term may be observed for a variety of reasons other than inefficiency. Arguing that the existence of X-inefficiency could not be determined in the absence of a theory of error, Stigler suggested that supposedly 'inefficient' behaviour might be observed because of an incorrectly specified objective function, a failure to account for all relevant inputs, and a lack of recognition of the constraints on the production process (Stigler, 1976). These factors may explain one-sided disturbances in the stochastic frontier framework (Dopuch and Gupta, 1997). Unobserved characteristics of the acute hospitals may contribute to a significant non-negative term. Possible explanations include the following:

(1) Excess expenditures may be incurred in hospitals that maintain spare capacity in order to accommodate emergency admissions, the arrival of which is subject to an unpredictable daily process (Bagust *et al.*, 1999). If it is considered important that hospitals

do not turn emergency patients away because beds are unavailable, there is an argument for incorporating this explicitly in the hospital's objective function.

(2) Excess expenditures can be due to periodic capital investment or renovation programmes, the costs of which are not spread accurately over subsequent years (Skinner, 1994).

(3) Hospitals face diverse constraints on their operating process. For example, those hospitals operating in environments where community and primary care is underdeveloped will be more constrained in their ability to discharge patients to more appropriate settings.

(4) Coding practices may be less accurate in particular types of hospitals. Hospitals with a more varied and complex casemix may find that the complexity of their activity is under-reported because medical records personnel code imprecisely the less common diagnoses or procedures.

(5) Accounting practices may vary, if hospitals with a more diverse set of activities – such as those engaged in teaching and research – are able to exercise discretion about what costs to attribute to patient care services.

As these examples suggest, the problems of interpretation are intrinsically bound up with model specification and accurate observation of all relevant data (Dopuch and Gupta, 1997). Some of these limitations might be resolved by analysis of panel data, comprising observations of the same organizations over time. Panel data estimation has advantages over cross-sectional analysis, in that repeat observations make it possible to control for unobservable organization-specific attributes and, by avoiding the need to partition inefficiency and noise, the approach provides consistent estimation of the inefficiency term (Linna, 1998).

However, panel data estimation has three shortcomings. Firstly, it is necessary to specify a trend describing how organizational efficiency changes (or remains constant) over time. This requirement would appear to preclude the possibility that changes in relative efficiency across time periods are not subject to a consistent process. Yet if regulators set efficiency targets based on performance relative to peers rather than performance in the previous period, efficiency changes over time might not follow a clear pattern. As much of the rationale for studying efficiency stems from a desire to ascertain whether inefficient organizations improve their performance in response to government targets, the assumption that efficiency is independent of official intervention is not particularly attractive (Wagstaff, 1989).

Secondly, panel data estimation is often out of step with common policy requirements for annual performance targets. Unless policy makers can be encouraged to take a longer term perspective on assessing organizational performance, the demand for cross-sectional analyses is likely to persist.

A final problem with longitudinal analysis is that it requires consistency in data definitions and counting practices across years, or some means of dealing with inconsistencies. This is a particular problem in the NHS. Hospitals have revised continually their approaches to counting activity in order to realize annual productivity targets (Radical Health Statistics Group 1992; Radical Health Statistics Group, 1995). Moreover, the unit of analysis has not remained constant over time, as a result of hospital mergers. Nevertheless, panel data estimation of NHS hospital costs may be a fertile subject of future research, particularly if policy makers can be convinced to take a longer term view when monitoring the performance of public sector organizations.

In conclusion, this paper has challenged the interpretation placed on the residuals from a cross-sectional OLS model designed to compare the relative performance of acute hospitals in England. The residual, on which efficiency estimates have been based, may not merit the interpretation that it has been accorded by the English Department of Health. The analysis suggests that English acute hospitals exhibit less inefficiency than is implied by official estimates and that OLS residuals should not be relied upon for target setting.

ACKNOWLEDGEMENTS

The authors are grateful to the Department of Health, the Medical Research Council and Northern and Yorkshire Region for financial support. The authors would like to thank Diane Dawson, Mark Dushieko, Maria Goddard, Antonio Giuffrida, Andrew Jones, Peter Smith, colleagues at the University of York Seminars on Health Econometrics, and participants at the Health Economics Study Group, Newcastle, 5–7 January 2000 and at the ASSET conference, Lisbon, 3–4 November 2000 for helpful comments. The authors also gratefully acknowledge Neil Söderlund who did much of the initial work on which this paper is based. All omissions and errors are the sole responsibility of the authors.

REFERENCES

Aigner, D., Lovell, C. A. K. and Schmidt, P. (1977) Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics*, **6**, 21–37.

Audit Commission and Department of Health (1999) *NHS Trust Profiles Handbook – 1997/98*, London, Audit Commission.

Bagust, A., Place, M. and Posnett, J. W. (1999) Dynamics of bed use in accommodating emergency admissions: stochastic simulation model, *British Medical Journal*, **319**, 155–8.

Benton, P. L., Antony, P., Evans, H., Light, S. M., Mountney, L. M. and Sanderson, H. F. (1998) The development of healthcare resource groups, version 3, *Journal of Public Health Medicine*, **20**, 351–8.

Berger, A. N. (1993) 'Distribution-free' estimates of efficiency in the US banking industry and tests of the standard distributional assumptions, *Journal of Productivity Analysis*, **4**, 261–92.

Coelli, T. (1996), *A guide to FRONTIER version 4.1: a computer program for stochastic frontier production and cost function estimation*, Armidale: Centre for Efficiency and Productivity Analysis, University of New England.

Cook, R. D. and Weisberg, S. (1982) *Residuals and Influence in Regression*, Chapman and Hall, London.

Cummins, J. D. and Zi, H. (1997) *Measuring cost efficiency in the US life insurance industry: econometric and mathematical programming approaches*, The Wharton School, University of Pennsylvania, Philadelphia.

Dopuch, N. and Gupta, M. (1997) Estimation of benchmark performance standards: an application to public school expenditures, *Journal of Accounting and Economics*, **23**, 147–61.

Feldstein, M. S. (1967) *Economic Analysis for Health Service Efficiency: Econometric Studies of the British National Health Service*, Amsterdam, North-Holland, North Holland Publishing Company.

Goddard, M. and Mannion, R. (1998) From competition to co-operation: new economic relationships in the National Health Service, *Health Economics*, **7**, 105–19.

Goldstein, H. and Spiegelhalter, D. J. (1996) League tables and their limitations: statistical issues in comparisons of institutional performance, *Journal of the Royal Statistical Society*, **A159**, 385–443.

Greene, W. H. (1993) The econometric approach to efficiency analysis, in *The measurement of productive efficiency: Techniques and applications*, (Eds) H. O. Fried, C. A. K. Lovell and S. S. Schmidt, Oxford University Press, New York.

Greene, W. H. (1995) *Limdep Version 7.0 User's Manual*, Econometric Software, Inc, Castle Hill.

Jondrow, J., Lovell, C. A. K., Materov, I. S. and Schmidt, P. (1982) On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics*, **19**, 233–8.

Kopp, R. J. and Mullahy, J. (1990) Moment-based estimation and testing of stochastic frontier models, *Journal of Econometrics*, **46**, 165–83.

Linna, M. (1998) Measuring hospital cost efficiency with panel data models, *Health Economics*, **7**, 415–27.

Maddala, G. S. (1988) *Introduction to Econometrics*. London: Collier Macmillan.

Marshall, E. C. and Spiegelhalter, D. J. (1998) Comparing institutional performance using Markov chain Monte Carlo methods in *Statistical analysis of medical data: new developments* (Ed.) Everitt, B. S.; Dunn, G., Arnold, London.

NHS Executive (1997) *The New NHS: modern, dependable*, NHS Executive, Leeds.

NHS Executive (1999a) *Quality and Performance in the NHS: clinical indicators*, NHS Executive, Leeds.

NHS Executive (1999b) *Quality and Performance in the NHS: high level performance indicators*, NHS Executive, Leeds.

Radical Health Statistics Group (1992) NHS Reforms: the first six months – proof of progress or a statistical smokescreen, *British Medical Journal*, **304**, 705–9.

Radical Health Statistics Group (1995) NHS 'indicators of success': what do they tell us?, *British Medical Journal*, **310**, 1045–50.

Schleifer, A. (1985) A theory of yardstick competition, *Rand Journal of Economics*. **16**. 319–27.

Schmidt, P. (1985) Frontier production functions, *Econometric Reviews*, **4**, 289–328.

Schmidt, P. and Lin, T. (1984) Simple tests of alternative specifications in stochastic frontier models, *Journal of Econometrics*, **24**, 349–61.

Schmidt, P. and Lovell, C. A. K. (1979) Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers, *Journal of Econometrics*, **9**, 343–66.

Sidak, Z. (1967) Rectangular confidence regions for the means of multivariate normal distributions, *Journal of the American Statistical Association*, **62**, 626–33.

Skinner, J. (1994) What do stochastic frontier cost functions tell us about inefficiency?, *Journal of Health Economics*, **13**, 323–8.

Söderlund, N. and van der Merwe, R. (1999) *Hospital benchmarking analysis and the derivation of cost indices*, Centre for Health Economics discussion paper 174, University of York, York.

Stevenson, R. F. (1980) Likelihood functions for generalized stochastic frontier estimation, *Journal of Econometrics*, **13**, 57–66.

Stigler, G. (1976) The Xistence of X-efficiency, *American Economic Review*, **66**, 213–16.

Timmer, C. P. (1971) Using a probabilistic frontier production function to measure technical efficiency, *Journal of Political Economy*. **79**, 776–94.

Wagstaff, A. (1989) Estimating efficiency in the hospital sector: a comparison of three statistical cost frontier models, *Applied Economics*, **21**, 659–72.