Determinants of Cost efficiency of Finnish Hospitals: A Comparison of DEA and SFA

MIIKA LINNA
National Research and Development Centre for Welfare and Health, Siltasaarenkatu 18, PO BOX 220 00531 Helsinki, Finland, Phone +358 9 3967 2295, telefax +358 9 3967 2485, Internet: miika.linna@stakes.fi

UNTO HÄKKINEN
Institute of Pharmaco-Economics #710, 10665 Jasper Avenue, Edmonton, AB T5J 3S9, Canada

Address for correspondence:
Miika Linna
National Research and Development Centre for Welfare and Health
Siltasaarenkatu 18, PO BOX 220
00531 Helsinki, Finland
Phone +358 9 3967 2295, telefax +358 9 3967 2485

Internet: miika.linna@stakes.fi

Abstract

This study investigated various factors explaining the technical, allocative and cost efficiency of Finnish hospitals using parametric frontier models and nonparametric DEA models. DEA scores were correlated to the same set of variables as efficiency scores by stochastic frontier models. Specialization, specialization in expensive DRGs, sufficiently high proportion of resident physicians and increasing the relative share of physician input contributed to efficiency. In addition, some evidence of decreasing returns to scale were found.

Keywords: Hospitals; DEA; Stochastic frontier models; Cost efficiency

JEL classification: C14; C21; D24; I10

Introduction

Until 1993 Finnish hospitals were reimbursed on a cost basis, which did not provide strong incentives for cost minimization or monitoring. Subsequently, it was not surprising that several studies found evidence of significantly decreased productivity among hospitals (Pekurinen et al., 1991) and health centers (Luoma and Järviö, 1992) in the 1980s. Moreover, large variations in efficiency between different health care institutions were also identified (Linna and Häkkinen, 1996; Luoma et al., 1996). The economic depression which hit the country in the early 1990's eventually impelled the government and municipalities to react to the assumed excessive costs of health care production. Health care costs and the efficiency of health care producers became common topics of public debate.

The new state subsidy system introduced in 1993 altered the position of hospitals dramatically. All health care state subsidies (except those for research and education) were for the first time allocated to municipalities rather than providers (e.g. hospitals), as before. The new system meant that hospitals and hospital districts had to begin selling services to the municipalities, which were now the budget holders of public health care money. All these changes have increased the cost conciousness of hospitals and the demand for information for managerial decision-making purposes. Despite some attempts to quantify inefficiencies in production, a lack of the administrative personnel and relevent data has prevented hospitals from being able to identify sources of inefficiency. Until recently, the only available guidance for managerial decision making were measures such as per capita costs of hospital care in a hospital district, or average cost of patient day, discharge or visit. It is obvious that measures of this type provide little insight into the complex issue of productive efficiency and are insufficient data for operative planning.

A number of studies elsewhere have employed either DEA (Burgess and Wilson, 1996; Ferrier and Valdmanis, 1996; Magnussen, 1996; Färe et al., 1994; Grosskopf and Valdmanis, 1987;) or stochastic frontier models (Eakin and Kniesner, 1988; Zuckerman et al., 1994; Wagstaff and Lopez, 1996). Comparative studies of different approaches to efficiency analyses are still scarce (Banker et al., 1986). The theoretical literature contains several, often conflicting, explanations for efficiency disparities. However, the conclusions drawn in most of these studies may depend at least partly on the model specification or method used. Moreover, most studies of hospital efficiency can be criticized for not having measured output or even case-mix appropriately.[1] These problems provided the motivation for using several specifications, a comprehensive case-mix measure and the testing of different methods - parametric and non-parametric - for calculating individual efficiency scores.

In this paper we have evaluated the determinants of cost efficiency in somatic hospital care in Finland. Overall cost efficiency was divided into two components of allocative and technical efficiency. Technical efficiency measures how the levels of all the used inputs, given certain amount of outputs, compare with the optimal, best practice use of inputs. Allocative efficiency refers to the extent to which the input choices fail to satisfy the marginal equivalences for cost minimization. A two stage procedure was used: First, efficiency scores were calculated for each hospital, and in the next stage these scores were explained using a variety of factors expected to affect the observed inefficiencies. Nonparametric and parametric methods were used to obtain overall cost efficiency and nonparametric methods to decompose the cost efficiency measure into technical and allocative efficiency.

2. Decomposition of cost efficiency into technical and allocative efficiency

2.1. Methods and specification of the stochastic models

Parametric models are specified by a stochastic frontier cost function (Aigner et al., 1977). Cost or profit functions provide a convenient theoretical setting for decomposing cost efficiency into technical and allocative efficiency. Using cost functions, Shephard's lemma gives the input choices which are efficient and thus provide the benchmark against actual demands.

However, since there are many problems apparent in decomposition using stochastic models (Atkinson and Cornwell, 1994; Kumbhakar, 1996), the stochastic frontier estimation technique was used here only to estimate overall cost efficiency. Cost efficiency was estimated with short-run multi-product cost functions since major capital investments were excluded. A set of statistical specification tests (Linna and Häkkinen, 1996) suggested that the Box-Cox transformed frontier cost function (SFMODEL) would best describe the costs of Finnish hospitals:[2]

$$\ln \frac{C_i}{w_d} = a + \sum_{j=1}^{m} b_j y_{ij}^{(l)} + d \ln \frac{w_{oi}}{w_{di}} + u_i + n_i \tag{1}$$

where C is total costs, w_d is input prices for doctors and w_o is input prices for other staff. The Box-Cox transformation is $y^{(l)} = (y^l - 1)/l$. Linear homogeneity was preserved by leaving the total cost and price variables untransformed. In order to obtain individual efficiency measures the residual must be decomposed using the technique suggested by Jondrow et al. (1982). The conditional estimates of u_i, $E\left[u_i | v_i + u_i\right]$, must be derived to find estimates for the individual inefficiency terms.

2.2. Methods and specification of the DEA models

Most DEA models assume the production possibilities set to be convex and to exhibit constant or variable returns to scale. With DEA it is not necessary to make assumptions about the parameters and the functional form of the production correspondence. The important advantage of DEA is that it is relatively easy to use when the decision making units (DMUs) use multiple inputs to produce multiple outputs. The efficiency scores are determined by the ratio of the sum of weighted outputs to the sum of corresponding weighted inputs. The weights are determined so as to show the DMU at maximum relative efficiency (Charnes et al., 1978; Banker et al., 1984). Cost efficiency can be measured if input prices are available in addition to output and input data.

The measurement of cost efficiency is relatively straightforward using nonparametric methods (Ray and Kim, 1995; Ferrier and Valdmanis, 1996). The standard measure of cost efficiency is obtained via a two-stage process: i) estimate the minimum price-adjusted resource usage given technological constraints, and ii) compare this minimum to actual, observed costs. Let $x = (x_1,...,x_k) \in R_+^k$ denote a vector of inputs and $y = (y_1,...,y_m) \in R_+^m$ denote a vector of outputs. Formally, the cost efficiency model (DEACE1) can be specified as:

$$
\begin{aligned}
& Min_{z,x} \sum_j w_{j0} \cdot x_j \\
& s.t. \quad z \cdot Y \geq y_0, \\
& \qquad z \cdot X \leq x, \\
& \qquad z_i \geq 0 \\
& \qquad \sum_{i=1}^{n} z_i = 1
\end{aligned}
\tag{2}
$$

where Y is an $n \times m$ matrix of observed outputs for n hospitals and X is an $n \times k$ matrix of inputs for each hospital. z is a $1 \times n$ vector of intensity variables and $w = (w_1,...,w_k) \in R_+^k$

denotes input prices. The constraints of the model (2) define the input requirement set given by:

$$L(y) = \{x\colon z \cdot Y \geq y_0,\ z \cdot X \leq x,\ z_i \geq 0, \sum_{i=1}^{n} z_i = 1\} \tag{3}$$

The input requirement set specifies a convex technology with variable returns to scale (VRS), which is imposed by the constraint $\sum_{i=1}^{n} z_i = 1$. Leaving the constraint out of the model changes the technology to constant returns to scale (CRS). In this study CRS model (DEACE2) was also estimated to control any possible identification problems (Försund, 1992).

The cost minimizing set of inputs x^* (a solution to model (2)) can then be used to calculate the cost efficiency (CE) by CE = $w' \cdot x^* / w' \cdot x$, where x are actual, observed inputs used.

Another possibility for obtaining cost efficiency estimates is to measure a 'global cost efficiency', where total costs $TC = w' \cdot x$ are used as the input variable. The usual meaning of allocative inefficiency is that the input factor mix is suboptimal with respect to prevailing input prices when different sets of prices are defined exogenously for each DMU. In Finland, where wages are centrally negotiated and price variation is fairly small, this difference between the two measures of cost efficiency should be insignificant. Assuming identical input prices, cost efficiency can be calculated by solving the following linear program (DEA3):

$$\begin{aligned} & Min_{z,\, l_{CE}} \quad l_{CE} \\ & s.t. \quad z \cdot Y \geq y_0, \\ & \qquad z \cdot C \leq l_{CE} \cdot c_0, \\ & \qquad z_i \geq 0 \\ & \qquad \sum_{i=1}^{n} z_i = 1 \end{aligned} \tag{4}$$

where c is a scalar representing a cost or budget level, and C is the $n \times 1$ matrix of observed costs. Eliminating the summation constraint changes the model (DEA4) to constant returns to scale.

The decomposition into allocative and technical components can be accomplished by first solving the following linear program, which gives the input oriented technical inefficiency component:

$$
\begin{aligned}
& Min_{z,m}\ m \\
s.t. \quad & z \cdot Y \geq y_0, \\
& z \cdot X \leq m \cdot x, \\
& z_i \geq 0 \\
& \sum_{i=1}^{n} z_i = 1
\end{aligned}
\tag{6}
$$

The technical inefficiency component is given by solution $TE = m^*$. This is a reciprocal measure of the distance function by Farrell (1957) and Shepard (1970). Now it is simple to calculate the allocative efficiency by AE = CE/TE.

Again, in model (6) the summation constraint on intensity variables z imposes variable returns to scale (VRS). Eliminating the summation constraint changes the model to constant returns to scale (CRS). The scale efficiency measure SCE can be calculated as the ratio of CRS technical efficiency to VRS technical efficiency, $SCE = TE_{CRS} / TE_{VRS}$.

3. Variables and data

In this study we used total operating costs as an explanatory variable, four classes of variables as a measure of output, and three input variables with their corresponding input prices (Table 1). Ferrier and Valdmanis (1996) used aggregate number of personnel and number of beds as input variables. In this study it was possible to measure the input of two categories of personnel: physicians and other personnel. The working time for doctors also included the working time of resident physicians.

Table 1 here.

3.1. Data

We used cross-sectional data from 1994 on 48 acute care hospitals. The five university hospitals were disaggregated into main specialities operating as managerially independent units. Thus the total number of observations (or decision making units, DMUs) was 95. Private hospitals (except for one observation), military hospitals, psychiatric hospitals and psychiatric wards of acute hospitals were excluded.

Data were collected from hospital statistics published annually by Suomen Kuntaliitto (1994) and the National Discharge Registry, and were supplemented by research and teaching variables obtained via a separate questionnaire sent to the hospitals. The data were sent to the hospital administration for final checking and verification. Input prices were obtained from the wage statistics for 1994 collected by Statistics Finland.

3.2. Inpatient and outpatient services

In the measurement of hospital patient output a crucial conceptual distinction is whether the output is the actual provision of the medical treatment itself or the resulting improvement in the patients' health status using e.g. QALYs (Butler, 1995). Perhaps the most successful approach for determining the final products of hospital inpatient care is the DRG classification system, which takes into account the hospital's case-mix. Although the DRG system has been evolving for several years, DRGs still suffer from certain weaknesses; in particular, some exhibit significant heterogeneity in resource use because they fail to explicitly take account of the severity of illness of the patient (Horn et al., 1985; Averill et al., 1992). However, the effect of this bias remains unclear when the DRG system is used as a method for aggregating hospital inpatient treatments into a single output measure.

We used the Fin-DRG patient classification system, which is a Finnish version of the HCFA (Health Care Financing Administration) DRG grouping system (Virtanen et al., 1995). We weighted the DRG groups with actual average costs incurred by each episode. The DRG cost weights were based on a study using data from three Finnish university hospitals (Salonen et al., 1995). The variability in DRG groups was processed by analyzing the outliers (measured by the length of stay) separately. If the inpatient episode exceeded a certain cut-off point, the remaining patient days were inserted into a separate variable (BED-DAYS).

There is no widely accepted classification system or standard for outpatient visits. In addition, the introduction of short stay surgery has had a tremendous impact on the production of some elective operations. Thus it is possible that hospitals which record the relatively demanding short stay surgery as ordinary outpatient visits will appear inefficient. We therefore asked the hospitals if they recorded any of their short stay surgery as outpatient visits, and used explicit weighting when necessary. In this study we used two outpatient visit classes: 1) outpatient visits (VISIT) and 2) emergency visits (EMVIS).

3.3. Teaching and research output

Teaching and research may increase the costs of hospitals both directly and indirectly. Direct costs are additional investments sunk in teaching and research programs, lecture rooms, research laboratories and equipment. An indirect influence of teaching and research is the loss of labor productivity in patient care; students and research projects absorb the time of the professional personnel. The more students in relation to professionals, the more time for teaching is needed.

Medical and nursing students are also production factors. Students are, by definition, less productive; they use more time, materials and tests for the same tasks as professionals, while the salaries of postgraduate medical students are nearly as high as those of professionals. The patients who take part in clinical research projects stay longer in hospital and use more outpatient visits, tests and treatments.

In this study it was possible to describe the teaching and research output fairly accurately. As a measure of teaching activity we used the number of postgraduate medical students (RESIDENTS). This variable can be interpreted as a one-year postgraduate training output. The teaching activity with respect to junior medical students was measured by the number of teaching periods in clinical work (STUEDU). To measure the influence of nursing education we collected the number of on-the-job training weeks of nursing students in different hospitals (NURSE-EDU).

Research output (RESEARCH) was measured by collecting the bibliographic data of refereed scientific articles and medical dissertations in 1994 from all hospitals. We then weighted each article by the impact factor (SCI 1994) of the journal publishing it. Identical weighting is used in the teaching reimbursement for Finnish hospitals.

3.4. Cost variable

The net operating costs were used as an explanatory variable in the cost function models and were obtained by subtracting additional personal revenues and purchasing costs of special services (not included in the outputs) from the total operating costs of a hospital. The cost variable reflects the total costs of hospital services paid by the purchasers (municipalities, central government and patients). Net operating costs include all production-related (direct and indirect) costs of a hospital, but not capital costs such as depreciation and interest charges. However, operating costs do include the flow of new minor capital re-investments, though not the major capital investments.

3.5. Input price variables

Input price variables were constructed by using the average total working hours and average total wages paid in five employee categories: i) doctors, ii) nurses and other high-level care personnel, iii) care personnel with lower levels of education, iv) maintenance and catering personnel, and v) others (e.g. administrative staff). Because there was some variation in how the hospitals recorded their staff in the registers, we minimized this by combining the input price categories into two price variables: one for doctors and one for other employees (Table 1). The prices of materials, equipment and minor capital investments were assumed to be uniform because there were no satisfactory means to measure true prices.

4. Efficiency estimates

Models (2) and (4) - (6) were solved using linear programming routines of GAMS software. The frontier cost function model (1) and respective inefficiency measures were estimated using the FRONTIER subroutine in LIMDEP (Greene, 1993). The algorithm is a maximum likelihood estimator that uses OLS estimates as starting values.

As there is little theoretical guidance for picking the distributional form for u in SFMODEL, the only solution is to try various alternative distributions for u. Both the half-normal and exponential forms were chosen. Because the choice of the inefficiency term u did not seem to affect the results at all, only estimates by half-normal model are reported. In a cost frontier model the composed error $u + v$ should be positively skewed with a non-zero mean. The LIMDEP program automatically checks the OLS residuals for correct skewness before proceeding to the maximum likelihood estimate of the frontier. Multicollinearity was no worse than moderate. Cost functions were similar for the teaching and non-teaching hospitals according to the Chow test. The Hausman test revealed no evidence of endogeneity and the final model could be estimated without instrumental variables. The estimation results are shown in Table 2. The estimated parameters are plausible both in sign and magnitude. The estimation strategies and specification tests are reported more thoroughly in Linna and Häkkinen (1996).

Table 2 here.

The average cost efficiency level was between 0.84 and 0.92 in DEA models, and 0.86 with the stochastic frontier model (Table 3). The correlations between the parametric and nonparametric cost efficiency scores were 0.61 (DEACE1), 0.41 (DEACE2), 0.59 (DEA3), and 0.58 (DEA4).

Table 3 here.

The average level of technical inefficiency was 0.95 with VRS hypothesis and 0.91 with CRS. Allocative inefficiency added an average of 5 - 7% to hospital costs, indicating that approximately half of the cost inefficiency could be attributed to technical inefficiency, and the other half to allocative inefficiency.

63% of all the hospitals yielded the highest technical efficiency score of 1 in VRS models and 40% in CRS models. It is possible that the number of output and input variables was too large for the sample size, which may have caused problems especially with the VRS model (Smith,1997).

Additional cost did not seem to be due to scale inefficiency: the average scale efficiency was as high as 0.96. Scale efficiency is maximized in the size range between 40 and 250 beds. It is possible that the optimal scale varies according to output levels. Inspecting the intensity variables of model (6), 16.8% of the scale inefficient hospitals operated at increasing returns, and 35.8% at decreasing returns. However, even if the scale measure were unambigious, the utilization of increasing or decreasing returns may be difficult.

Following Ferrier and Valdmanis (1996), DEA results can be used to study more closely the sources of allocative inefficiency. The hospitals' input vector can be contracted radially to reach the efficiency frontier by multiplying the input vector by the efficiency score obtained from (6). On the other hand, the solution to model (2) yields the cost minimizing vector of inputs. Now it is possible to compare each (observed, and technical efficiency adjusted) element of input vector individually to cost minimising level of the corresponding input. The results are reported in Table 3 for the total sample and according to hospital type. Estimates were calculated separately for CRS and VRS models.

Table 4 here.

The results for the CRS model indicate that all types of hospital underutilized labor input of doctors by an average of 13%. Simultaneously, the labor inputs of other personnel, and materials and equipment were overutilized by approximately 6% each.

The underutilization of physician labor held across all types of hospital but the use of material and equipment was different between university and other hospitals. University hospitals seemed to overutilize the materials and equipment, while their use of non-physician labor was nearly optimal.

Central district hospitals and rural hospitals used other personnel at higher levels than optimal by 9% and 15%, respectively, but underutilized material and equipment allocation by 5% and 7%.

For VRS models the results were broadly similar, but there was less underutilization of doctor labor. However, only 10% of the hospitals in the sample had physician labor/optimal physician labor ratios larger than 1.

5. Determinants of inefficiency

5.1. Explanatory variables

In the second part of the study the estimated efficiency scores were analysed by regressing them against a set of observed characteristics of the hospitals and their environments. These included the degree of specialisation, use of modern technology (e.g. treating patients in outpatient clinics), input allocation, quality control, scale of operation and patient transfers to other care facilities. Some of the tested factors were less clearly controllable by managerial decisions. These included teaching status, provision of emergency services, primary health care organisation and, at least partly, scale of operation. The last category of explanatory variables reflected environmental characteristics mostly beyond the influence of managerial actions. Competitive pressure, financial status of the local authorities and various geo- and demographic factors were some examples of these environmental characteristics used in the analysis.

Although competition is not a prominent feature of the Finnish hospital market, the market share of hospitals was measured by an index (MARKET) relating the percentage of a hospital's inpatient admissions to the total number of admissions in the hospital district.

Competition has been found to increase specialization among hospitals (Dayhoff and Cromwell, 1993). Dayhoff and Cromwell argue that specialization is a logical result of the search for cost efficiency. A low degree of diversification may require more costly efforts by the management. Larger volumes in some procedures have also been associated with superior outcomes in several studies (Effective Health Care, 1996), although many of the latter are not of the highest quality. Furthermore, there are strong beliefs that concentrating complicated cases in one hospital decreases cost because of the opportunity of exploiting the learning curve.

The degree of specialization in the DRGs was measured by the information theory index (DRG-ITI) (Evans and Walker, 1972; Farley, 1989). Let c_{ij} $(i = 1,\ldots,N,\ j = 1,\ldots,K)$ be the number of cases belonging to hospital *i* and DRG category *j*. Let $q_{ij} = c_{ij} / \sum_j c_{ij}$ denote the proportion of DRG cases *j* to all DRG cases in hospital *i*. $p_j = \sum_i c_{ij} / \sum_{i,j} c_{ij}$ represents the proportion of all DRG cases *j* in all hospitals to all cases. Then, the information theory index for DRGs for hospital *i* can be calculated as

$$E_i = \sum_j q_{ij} \cdot \ln(\frac{q_{ij}}{p_j}) \tag{7}$$

The resulting index number E is equal to zero if no specialization occurs and the hospital DRG case proportions are equal to national proportions, and increases with the level of specialization. The proportions of expensive DRGs (DRGHIGH) of the total inpatient output were measured as the number of cases exceeding the 90th percentile (in costs) divided by total number of cases.

Because there is some evidence that the DEA method tends to give high efficiency scores for units with a specialized output structure (Nunamaker, 1985), the specialisation of a hospital's aggregate level production structure was described by a similar information theory index (OUT-ITI). The OUT-ITI is the weighted log of a hospital's product proportions (in monetary units, calculated as the produced quantity y_j multiplied by the marginal cost estimates $\hat{b}_j$ from model (1)) compared to national proportions. More formally, the OUT-ITI is calculated analogously to (7) by

$$H_i = \sum_j \hat{b}_j y_{ij} \cdot \ln\left(\frac{\hat{b}_j y_{ij}}{\hat{b}_j \bar{y}_j}\right) \tag{8}$$

There are not many possible quality or outcome indicators available from routine hospital statistics. The most common indicators have been mortality and readmission rates, both of which have serious weaknesses as outcome measures (Chambers and Clarke, 1990). However, quality here was measured by a readmission rate index (READMIS), readmission defined as the next admission of a patient within 60 days of a index discharge taking place in 1994.

Cost-shifting due to patients being referred to other care facilities was controlled via the variable TRANSF, which gives the percentage of inpatient discharges transferred to other care facilities within three days of admission. Another indicator of treatment style was the relative number of outpatient visits to all patients (OUTPAT). The variation in the hospital’s decisions about input allocations were measured by the percentage of total physician working hours to the total working hours of other personnel (DOCSHARE).

It is possible that DEA overestimates the importance of the four teaching and research output variables in the production process. We tested this assumption by including variable (RESIDENTS), which measured the ratio of resident physician working hours to the working hours of other physicians. The explanatory variable for hospital size was the number of beds (SIZE).

Exogenous factors that might affect the hospital’s efficiency included the average population weighted distance from the nearest hospital in the hospital district (DISTANCE), resources used for primary care and health education (PINVEST) and an index of the need for health services (HEALTH) measured as the proportion of the district’s residents on disability pension.

If municipalities or municipality federations receive generous subsidies, incentives for exercising cost control within hospitals may weaken (Luoma et al., 1996). The effect of exogenous economic factors is tested by the average income of the residents in the hospital’s catchment area (PINCOME) and the unemployment rate. PINCOME and the unemployment rate variable were found to be highly collinear, and therefore unemployment rate was not used in the final estimation.

Parametric efficiency scores were examined with standard OLS regression. For DEA scores, a censored Tobit model was used in the analysis (Greene, 1993). The efficiency score EFF was modified to describe the degree of inefficiency by setting INEFF = (1/EFF) -1. In this case the inefficiency scores are regressed, i.e. the negative sign of a coefficient means an association with efficiency, which allows it to be modeled by the following form:

$$
\begin{aligned}
&INEFF^* = \sum_j \beta_j \cdot x_j + \nu \\
&INEFF = 0 \qquad\qquad \text{if } INEFF^* \le 0 \\
&INEFF = INEFF^* \quad\; \text{if } INEFF^* > 0
\end{aligned}
\tag{9}
$$

where $\nu \sim N(0, \sigma^2)$, and β_j are the parameters for explanatory variables x_j.

5.2. Results

The results indicate (Table 4 and Table 5) that both the non-parametric and parametric methods gave systematically higher efficiency scores to hospitals with a specialized production structure (higher OUT-ITIs) and high share of resident physicians (RESIDENTS). OUT-ITI variable probably reflects the fact that extremal or 'corner' points (in the input-output space) of the sample tend to get higher efficiency scores only because of the scarcity of adjacent reference points. The variable perhaps catches certain inherent technical properties of the methods used and tells nothing meaningful about the economies of scope.

Table 5 here.

The nonparametric cost efficiency models seemed to be quite robust to the addition of scale or weight restrictions, giving generally the same explanations for efficiency. However, the VRS models systematically indicated SIZE to be a positive correlate for efficiency whereas CRS models could not verify this association.

According to these results, cost efficiency improved by shifting input allocation towards more physician manpower. Higher share of resident physicians and high proportion of expensive DRGs contributed to cost efficiency. Cost efficiency could not be explained by any of the exogenous factors, readmissions or transfers to other hospitals.

Table 6 here.

Technical efficiency was associated with a high proportion of expensive DRGs (DRGHIGH), and with specialization and a high share of residents in a similar manner to the cost efficiency scores.

Allocative efficiency correlated with the use of a high proportion of physician input, high proportion of residents and high level of specialisation in output selection (OUT-ITI). Scale efficiency was negatively correlated to hospital size and market share (MARKET). This indicates that in general the hospitals were large enough to achieve scale efficiency and decreasing returns started to dominate by the time a hospital reaches 250 beds in size.

6. Discussion

In this study we used parametric and nonparametric methods to analyse hospital cost efficiency. The findings indicate that the choice of modeling approach did not substantially affect the results. Using DEA models it was possible to decompose overall cost efficiency into allocative and technical components. DEA models revealed several factors contributing to technical, allocative and scale efficiency.

The level of cost inefficiency was estimated to lie between 8-15%, which suggests that improving the overall efficiency of hospitals could reduce the hospital costs by FIM 1.1 - 1.6 billion ($ 300 - 400 million). Approximately half of the observed inefficiency was due to technical and half to allocative inefficiency. Scale inefficiency was found to be a minor factor in overall inefficiency.

Specialization, physician input share and high proportion of resident physicians to all physicians were associated with cost efficiency. Somewhat unexpectedly, provision of outpatient services did not contribute to efficiency. High share of outpatient visits should in general indicate that the hospital uses modern, efficient technologies such as day-care and short stay surgery in patient care, thus contributing to lower costs. It is possible that there are differences in the case-mix and severity of outpatient visits, which gives biased estimates for cost efficiency and leads to false interpretation of the economic advantages of outpatient care.

Technical efficiency was correlated with the same factors as overall cost efficiency, except that specialization in expensive DRGs was also found to increase efficiency. Allocative inefficiency was found to result from underutilizing physician manpower and overutilizing other employees' work input. A logical explanation for this would be that cost minimization is not the only objective of a hospital; it also serves as a important employer within the community. The supply of services is not merely a health policy issue, because employees generate income tax revenues for the municipalities hosting the hospitals. These tend to become overstaffed in the pursuit of meeting an 'employment' or 'tax revenue' objective, which thus compromises allocative efficiency. The larger supply of less educated work force means it is natural for a community to pressure hospitals to hire an excessive non-physician staff. The overutilisation of non-physician manpower is therefore pronounced in hospitals which can be more clearly identified as ‘local’ in the sense that they serve well-defined populations.

The relative prices for physician working hours are low compared to those for other personnel with respect to the importance of physician input in the production process. At the prevailing wages, this would allow hospitals to approach optimum by increasing the relative share of physician working hours and decreasing the relative share of other personnel’s work input. In addition, teaching hospitals tend to overutilize material and equipment (investments) in production. The presence of technical inefficiency also suggests that some hospitals are using all inputs excessively.

Teaching and research output was measured more accurately than previous studies, which have used the number of residents or a teaching dummy as an indicator of teaching activity. High proportion of resident physicians to all physicians was found to contribute significantly to efficiency. Parametric estimates indicated that hospitals are able to produce both teaching and research output at decreasing marginal costs. This is probably due to product specific economics of scale; the production of teaching output and research articles needs ‘critical mass’ i.e. teaching and research programmes that are sufficiently large. The large research programmes generally take place in university hospitals. The main policy implication is that if, as proposed, the residency programs were shifted more to non-university hospitals, extra money would be needed due to losses in efficiency.

Scale efficiency was found to correlate negatively to size and market share variables. However, hospitals with large market shares (local monopolies) do not have incentives or even realistic opportunities to downsize organisation for the sake of scale efficiency.

The degree of output specialization was the most significant explanatory variable in the models explaining the different types of efficiencies. It is interesting to note that in a previous study the same variables (specialization, size, teaching status) explained the observed differences between parametric and nonparametric efficiency scores (Linna and Häkkinen, 1996). Since DEA is an estimation procedure which relies on extremal points it could, however, be sensitive to e.g. variable selection (Seiford and Thrall, 1990). As the number of outputs increases the ability to discriminate between the DMUs decreases. The more variables are added the greater becomes the chance that some inefficient unit will dominate in the added dimension and become efficient. It is possible that the results were affected by the relatively small sample used (or the number of output variables was too high).

In the study by Zuckerman et al. (1994) the conclusion regarding inefficiency of the hospitals did not appear to be sensitive to the specification of the output control (including quality) variables. Readmission rate, which did not turn out to be a significant correlate for efficiency in this study, was used as a control of quality differences. Despite the problems with this measure, some of the observed inefficiency may have been unmeasured output differences across hospitals.

It was reassuring to observe that cost efficiency scores by parametric frontier method were associated with the same factors as those using the DEA models. These models suggest that specialization (both in aggregate level output selection and expensive DRGs) and increasing the relative share of physician input contribute to efficiency. The results of this study indicated that even though there were some differences in the individual efficiency scores, DEA and stochastic frontier methods both provide the same determinants of efficiency.

Notes

1. For example, most studies using DRGs as a case-mix measure can only use data on Medicare patients.
2. The likelihood ratio test rejected the null hypothesis of $l = 0$ and favoured the choice of the Box-Cox specification.

References

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

Atkinson, S. and C. Cornwell. (1993). “Parametric estimation of technical and allocative inefficiency with panel data.“ *International Economic Review* 35, 231-243.

Averill, A., T. McGuire, B. Manning, D. Fowler, S. Horn, P. Dickson, M. Coye, D. Knowlton and J. Bender. (1992). “A study of the relationship between severity of ilness and hospital cost in New Jersey hospitals.“ *Health Services Research* 5, 587-604.

Banker, R., R. Conrad and R. Strauss. (1986). “A comparative application of DEA and translog methods: an illustrative study of hospital production.“ *Management Science* 1, 30-44.

Banker, R., A. Charnes and W. Cooper. (1984). “Some models for estimating technical and scale inefficiencies in DEA.“ *Management Science* 30, 1078-1092.

Burgess, J. and P. Wilson (1996). “Hospital ownership and technical inefficiency.“ *Management Science* 42, 110-123.

Butler, J. (1995). *Hospital cost analysis.* Dordrecht: Kluwer Academic Publishers.

Chambers, M. and A. Clarke. (1990). “Measuring readmission rates.“ *British Medical Journal* 17, 1134-1136.

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

Dayhoff, D.A. and J. Cromwell. (1993). “Measuring differences and similarities in hospital caseloads: a conceptual and empirical analysis.“ *Health Services Research* 28, 293-316.

Eakin, K.B. and T.J. Kniesner. (1988). “Estimating a non-minimum cost function for hospitals.“ *Southern Economic Journal* 54, 583-597.

Effective Health Care. (1996). “Hospital volume and health care outcomes, cost and patient access.“ *Effective health Care* 8, 1-16.

Evans, R.G. and H.D. Walker. (1972). “Information theory and the analysis of hospital cost structure.“ *Canadian Journal of Economics* 5, 398-418.

Farley, D. (1989). “Measuring casemix specialization and the concentration of diagnoses in hospitals using information theory.“ *Journal of Health Economics* 8, 185-207.

Farrell, M. (1957). “The measurement of productive efficiency.“ *Journal of the Royal Statistical Society* Vol. 120.

Ferrier, G. and V. Valdmanis. (1996). “Rural hospital performance and its correlates.“ *The Journal of Productivity Analysis* 7, 63-80.

Färe, R., S. Grosskopf, B. Lindgren, P. Roos. (1994). “Productivity Developments in Swedish Hospitals: A Malmquist Output Index Approach“ In Charnes A, Cooper W, Lewin A, and Seiford L (eds.), *Data Envelopment Analysis: Theory, Methodology and Applications.* Boston: Kluwer Academic Publishers.

Försund, F. (1992). “A comparison of parametric and non-parametric efficiency measures: the case of Norwegian ferries.“ *The Journal of Productivity Analysis* 3, 25-43.

Greene, W. (1993). *Econometric Analysis*, Second edition. New York: Macmillan Publishing Company.

Greene, W. (1992). *LIMDEP 6.0.* Belport, New York: Econometric software Inc.

Grosskopf, S. and V. Valdmanis. (1987). “Measuring hospital performance. A non-parametric approach.“ *Journal of Health Economics* 6, 89-107.

Horn, D., R. Horn, P. Sharkey and A. Chambers. (1985). “Interhospital differences in severity of illness.“ *The New England Journal of Medicine* July 4, 20-24.

Jondrow, J., C. Lovell, I. Materov and P. Schmidt. (1982). “On the estimation of technical inefficiency in the stochastic production function model.“ *Journal of Econometrics* 19, 233-238.

Kumbhakar, S.C. (1996). “Efficiency measurement with multiple outputs and multiple inputs.“ *The Journal of Productivity Analysis* 7, 225-255.

Linna, M. and U. Häkkinen. (1996). “A comparative application of econometric frontier and DEA methods for assessing cost efficiency of Finnish hospitals.“ accepted for publication.

Luoma, K., M-L. Järviö. (1992). *Health center productivity in Finland. Productivity change from 1980 to 1990 and productivity differences in 1990.* Helsinki: Discussion Paper 42, Government Institute for Economic Research.

Luoma, K., M-L. Järviö, I. Suoniemi, R.T. Hjerppe. (1996). “Financial incentives and productive efficiency in Finnish health centers.“ *Health Economics* 5, 435-445.

Magnussen, J. (1996). “Efficiency measurement and the operationalization of hospital production.“ *Health Services Research* 31, 21-37.

Nunamaker, T. (1985). “Using data envelopment analysis to measure the efficiency of non-profit organisations: a critical evaluation.“ *Managerial and decision economics* 6, 50-58.

Pekurinen, M., E. Pitkänen, H. Sintonen, V. Alander and D. Coyle. (1991). “Hospital productivity in Finland. Further analysis.“ *Liiketaloudellinen aikakausikirja* 39, 50-67.

Ray, S. and H. Kim. (1995). “Cost efficiency in the US steel industry: A nonparametric analysis using Data Envelopment Analysis.“ *European Journal of Operational Research* 80, 654-671.

Salonen, M., U. Häkkinen, I. Keskimäki and M. Linna. (1995). *DRG-ryhmien kustannuspainot suomalaisella aineistolla.* Helsinki: Stakes Aiheita 35.

SCI Journal Citation Reports. (1994).

Seiford, L.M. and R.M. Thrall. (1990). “Recent developments in DEA. The mathematical programming approach to frontier analysis.“ *Journal of Econometrics* 46, 7-38.

Shephard, R. (1970). The theory of cost and production functions, (Princeton University Press).

Smith, P. (1997) "Model misspecification in data envelopment analysis," *Annals of Operational Research* 73, 233-52.

Suomen Kuntaliitto. (1994). *Suomen sairaalataloustilasto,* Helsinki.

Virtanen, M., M. Brommels and I. Keskimäki. (1995). *FinDRG potilasryhmitys*. Helsinki: Luokitusopas 2, Ohjeita ja luokituksia, Stakes.

Wagstaff, A., G. Lopez. (1996). “Hospital costs in catalonia: a stochastic frontier analysis.“ *Applied Economic Letters* 3, 471-474.

Zuckerman, S., J. Hadley and L, Iezzoni. (1994). “Measuring hospital efficiency with frontier cost functions.“ *Journal of Health Economics* 13, 255-280.

Table 1. Variable definitions.

	Variable name	Definition
Output variables (y_i)		

Outpatient treatment:	EMVIS	Total number of emergency visits
	VISITS	Total sum of scheduled and follow-up visits
Inpatient treatment:	ADMISSIONS	DRG-weighted number of total admissions
	BED-DAYS	Total number of bed-days exceeding the cutoff point defined in the outlier analysis
Teaching variables:	RESIDENTS	Number of residents receiving 1 year of training at the hospital
	NURSE-EDU	Total number of on-the-job training weeks of nurses
	STUEDU	Total number of clinical training weeks of medical students
Research variable:	RESEARCH	Total number of impact-weighted scientific publications
Cost variable (TC)	TOTALCOST	Net operating costs of a hospital
Input variables (x_i)		
Doctors' working hours	DOCFTE	Average total working hours of doctors
Other employees' working hours	OTHFTE	Average total working hours of other employees
Costs of materials and equipment	MATCOSTS	Total costs of materials, equipment and other costs
Price variables (w_i)	DOCWAGES	Average hourly wage rate of doctors
	OTHERWAGES	Average hourly wage rate of other employees
	MATPRICE	Average price of materials, equipment and minor capital assets, assumed to be uniform across all hospitals

Table 2. Parameter estimates for the frontier cost function.

	Box-Cox -model	
	OLS Model	Frontier-model SFMODEL
Variables	Coefficient (*t*-value)	Coefficient (*t*-value)
Constant	6.17	5.98

	(36.68)	(31.98)
EMVIS	0.019	0.016
	(2.02)	(1.67)
VISITS	0.025	0.028
	(1.92)	(2.00)
ADMISSIONS	0.237	0.246
	(13.31)	(13.33)
BED-DAYS	0.086	0.080
	(6.01)	(5.08)
RESIDENTS	0.016	0.017
	(2.53)	(2.93)
RESEARCH	0.014	0.013
	(3.15	(2.70)
NURSE-EDU	0.061	0.062
	(3.51)	(2.98)
STUDEDU	0.010	0.008
	(2.85)	(3.05)
OTHERWAGES/ DOCWAGES	0.60	0.54
	(3.36)	(3.40)
R^2	0.975	-
LogL		54.5
Heteroscedasticity		
Breusch-Pagan c^2 (8)	15.7	
Chow-test		
F(42,44)	1.0	
Box-Cox -analysis		
H_0: $l = 0$		
LR, c^2 (1)	5.11	
Endogeneity test		
Hausman, c^2 (1)	0.46	
Multicollinearity		
(CI - index)	31.3	

Table 3.

Efficiency measure	Mean	Minimum	Maximum
Stochastic frontier model			
Cost efficiency (SFMODEL)	0.86	0.63	0.97
DEA models			
Cost efficiency:			
(DEACE1)	0.85	0.42	1
(DEACE2)	0.90	0.56	1
(DEA3)	0.87	0.53	1
(DEA4)	0.92	0.55	1
Technical efficiency (CRS)	0.91	0.44	1
Technical efficiency (VRS)	0.95	0.59	1
Allocative efficiency (CRS)	0.93	0.73	1
Allocative efficiency (VRS)	0.95	0.74	1
Scale efficiency	0.96	0.75	1

Table 4.

	Mean (CRS model)	Mean (VRS model)
Total sample		
Technically efficient working hours of physicians/	0.867	0.981

Optimal working hours of physicians		
Technically efficient working hours of other personnel/ Optimal working hours of other personnel	1.069	1.116
Technically efficienct expenditure on capital and materials/ Optimal expenditure on capital and materials	1.061	1.056
University hospitals		
Technically efficient working hours of physicians/ Optimal working hours of physicians	0.862	0.971
Technically efficient working hours of other personnel/ Optimal working hours of other personnel	1.005	1.071
Technically efficienct expenditure on capital and materials/ Optimal expenditure on capital and materials	1.130	1.063
Central district hospitals		
Technically efficient working hours of physicians/ Optimal working hours of physicians	0.822	0.984
Technically efficient working hours of other personnel/ Optimal working hours of other personnel	1.089	1.061
Technically efficienct expenditure on capital and materials/ Optimal expenditure on capital and materials	0.955	1.065
Local hospitals		
Technically efficient working hours of physicians/ Optimal working hours of physicians	0.874	0.960
Technically efficient working hours of other personnel/ Optimal working hours of other personnel	1.156	1.221
Technically efficienct expenditure on capital and materials/ Optimal expenditure on capital and materials	0.933	0.983

Table 5. Cost efficiency scores explained by various factors.

	Stochastic frontier estimates	DEA estimates			
	OLS regression	Tobit analysis			
Explanatory variables	SFMODEL	DEACE1 (CRS)	DEACE2 (VRS)	DEA3 (CRS)	DEA4 (VRS)
	Coefficient (*t*-ratio)	Coefficient (*t*-ratio)	Coefficient (*t*-ratio)	Coefficient (*t*-ratio)	Coefficient (*t*-ratio)

Constant	-0.221	1.198	1.791	0.582	0.712
	(-0.80)	(3.21)	(4.09)	(2.25)	(1.92)
OUT-ITI	-0.295	-1.209	-1.123	-1.221	-1.331
	(-4.61)	(-10.49)	(-8.69)	(-13.19)	(-9.75)
DRG-ITI	-0.008	0.033	0.033	0.033	0.039
	(-0.41)	(1.17)	(1.15)	(1.19)	(1.33)
DRGHIGH	-0.041	-0.358	-0.071	-0.364	-0.101
	(-0.29)	(-1.58)	-0.27)	(-2.09)	(-0.46)
READMIS	0.597	-0.629	-0.575	-0.365	-0.258
	(1.30)	(-0.95)	(-0.80)	(-0.79)	(-0.42)
TRANSF	-0.001	0.006	0.002	0.010	0.007
	(-0.23)	(0.80)	(0.27)	(1.72)	(0.92)
OUTPAT	-0.210	0.355	0.123	0.242	-0.049
	(-1.29)	(1.40)	(0.45)	(1.26)	(-0.21)
DOCSHARE	-0.583	-2.113	-2.472	-1.130	-2.265
	(-1.39)	(-3.07)	(-3.34)	(-2.49)	(-3.51)
DISTANCE	0.0006	-0.002	-0.001	0.002	-0.0007
	(0.38)	(-1.01)	(-0.51)	(1.24)	(-0.28)
PINCOME	0.013	-0.008	-0.014	-0.0001	0.0005
	(1.52)	(-1.10)	(-1.53)	(-0.01)	(0.01)
HEALTHIND	0.001	-0.0006	-0.0005	0.002	0.0007
	(1.19)	(-0.33)	(-0.25)	(1.29)	(0.38)
PINVEST	-0.001	-0.0008	-0.002	-0.004	-0.001
	(-0.70)	(-0.23)	(-0.65)	(-1.84)	(-0.54)
SIZE	-0.002	0.0001	-0.0008	0.0001	-0.0008
	(-2.24)	(0.39)	(-5.85)	(0.61)	(-3.82)
RESIDENT	-0.810	-3.073	-3.091	-1.932	-2.541
	(-2.38)	(-6.73)	(-5.85)	(-6.15)	(-5.59)
MARKET	0.002	0.0003	-0.0004	-0.001	-0.0003
	(0.51)	(0.46)	(-0.60)	(-1.06)	(-0.51)
s	$R^2 = 0.34$	0.129	0.127	0.091	0.103
		(12.05)	(9.86)	(11.43)	(9.10)
Log-likelihood	-	37.109	16.324	56.763	22.137

Table 6. Technical, allocative and scale efficiency explained by various factors.

	DEA estimates				
	Tobit analysis				
Explanatory variables	Technical efficiency (CRS)	Technical efficiency (VRS)	Allocative efficiency (CRS)	Allocative efficiency (VRS)	Scale efficiency
	Coefficient (*t*-ratio)	Coefficient (*t*-ratio)	Coefficient (*t*-ratio)	Coefficient (*t*-ratio)	Coefficient (*t*-ratio)

Constant	0.896	1.062	0.212	0.881	0.293
	(1.91)	(1.90)	(0.94)	(3.20)	(1.14)
OUT-ITI	-1.290	-1.206	-0.362	-0.390	-0.376
	(-7.99)	(-6.30)	(-5.57)	(-5.16)	(-4.24)
DRG-ITI	0.012	-0.005	0.019	0.034	0.003
	(0.36)	(-0.11)	(1.09)	(1.55)	(0.13)
DRGHIGH	-1.004	-0.651	0.197	0.251	-0.369
	(-3.14)	(-1.99)	(1.43)	(1.86)	(-2.14)
READMIS	-0.944	-0.167	0.382	0.434	-0.044
	(-1.01)	(-0.17)	(0.98)	(0.99)	(-0.09)
TRANSF	0.014	0.010	-0.002	-0.0008	0.005
	(1.50)	(0.94)	(-0.48)	(-0.14)	(1.06)
OUTPAT	0.179	0.082	-0.018	-0.144	-0.199
	(0.55)	(0.24)	(-0.12)	(-0.86)	(-1.00)
DOCSHARE	-0.632	-0.983	-1.778	-2.213	-0.245
	(-0.72)	(-1.00)	(-4.21)	(-4.65)	(-0.59)
DISTANCE	-0.005	-0.004	0.002	0.001	-0.003
	(-1.83)	(-1.22)	(1.16)	(0.65)	(-1.61)
PINCOME	-0.011	-0.008	0.004	-0.007	-0.007
	(-1.22)	(-0.74)	(0.76)	(-1.20)	(-1.35)
HEALTHIND	-0.0001	-0.001	0.001	-0.0003	-0.0007
	(-0.04)	(-0.52)	(0.87)	(-0.23)	(-0.59)
PINVEST	0.003	0.001	-0.003	-0.002	0.003
	(0.91)	(0.41)	(-1.34)	(-0.96)	(1.24)
SIZE	0.0002	-0.0008	-0.00005	-0.0003	0.0005
	(0.96)	(-2.57)	(-0.54)	(-2.22)	(4.65)
RESIDENT	-4.363	-4.014	-0.572	-0.892	-1.475
	(-6.29)	(-4.74)	(-1.97)	(-2.75)	(-3.94)
MARKET	0.001	0.0005	-0.0004	-0.0004	0.001
	(1.61)	(0.40)	(-0.97)	(-0.90)	(2.78)
s	0.144	0.142	0.082	0.083	0.077
	(9.78)	(7.50)	(11.83)	(9.56)	(8.49)
Log-likelihood	10.583	0.096	68.579	35.070	26.953