EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

EFFICIENCY EVALUATIONS FOR HOSPITALS

Torben Holvad

Thesis submitted for assessment with a view to obtaining
the Degree of Doctor of the European University Institute

Florence
August 1994

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

EFFICIENCY EVALUATIONS FOR HOSPITALS

Torben Holvad

The Thesis Committee consists of:

Prof. Hans Keiding, University of Copenhagen
" Alan Kirman, E.U.I., Supervisor
" Stephen Martin, E.U.I.
" Henri Tulkens, Université Catholique de Louvain
" David Ulph, University of Bristol

To Aurora

Chapter 1

Introduction.

This thesis has two main fields of interest: efficiency measurement models and hospitals. Efficiency measurement models within the area of production theory a[illegible]es such methods are directed to indicate whether and to what extent producers' performance differ from what should be expected. The importance of such information is stressed by Fried, Lovell & Schmidt (1993, p. vii):

"Assuming that high levels of economic efficiency and productivity, and high rates of productivity growth, are desirable goals, then it is important to define and measure efficiency and productivity in ways that respect economic theory and provide useful information to managers and policy makers".

Although a study of the properties of the efficiency measurement models is interesting in its own right, I believe that it is in the application of these methods to real world data that the advantages and disadvantages are revealed. Moreover, such an application can provide empirical results concerning the efficiency structure for the chosen sample of production data. This brings me to the second field of interest, hospitals. I have chosen to apply the efficiency measurement models on data which originate from hospitals. Applying methods with a clear basis in economic theory, such as the efficiency measurement models to hospitals is interesting due to the very special position hospitals and health care provision have in relation to economic analysis. Most of the usual assumptions in microeconomic theory are not valid for hospitals. Thus medical care commodities have different characteristics: "... from the usual commodity of economic textbooks ... which although not individually unique to this market ... taken together, they do establish a special place for medical care in economic analysis", Arrow (1963, p. 948). Among other special characteristics the medical care commodity has a high degree of product uncertainty, since the outcome of a treatment depends on number of factors outside the control of the producer. In addition and related to this the problems of defining and measuring hospital outputs makes efficiency measurement for hospitals a non-trivial and interesting area for detailed research.

The thesis thus has two aims. Firstly, it is concerned with a theoretical examination of efficiency measurement models originating from the production frontier tradition. This approach involves the construction of a common benchmark for a set of empirical observations, i.e. a production frontier, derived from the best-practices among these observations. The production frontier measures the maximum attainable output for any given input levels and all possible, efficient observations are, therefore, those which are situated on the frontier. The distance from the observation to the frontier thus serves as a measure for the degree of inefficiency for an observation. In particular, the non-parametric efficiency measurement methods, Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH), are analysed. These methods seem relevant for empirical analyses in situations where information about the production processes, at best, are weak.

Secondly, the thesis includes empirical applications of the non- parametric efficiency measurement methods to 3 hospital data sets in order to illustrate the functioning of the methods and to examine the applicability of the efficiency results obtained [illegible] applications provide possibilities of examining hypotheses concerning the efficiency structure of the hospital sector. Two of the hospital data sets are from a sample of 80 Danish hospitals with data on inputs and outputs for each one. The first data set does not allow for case-mix standardization for outputs, while the second data set does. Among other issues these data sets will be used to examine the effect of case-mix differences on the efficiency results. One data set consists of data on inputs and outputs for a sample of private British hospitals. This data set will be used to investigate the degree of efficiency within a group of private firms. These applications show that the range of sector types which can be analysed by efficiency measurement models are not restricted to public sector firms for example.

The thesis consists of 7 chapters organised in the following way. Chapter 2 includes a short literature survey. An indepth theoretical examination of the properties of the non-parametric approach to efficiency measurement represented by the models Data Envelopment Analysis and Free Disposal Hull is included in chapter 3[1]. This chapter also looks at the relationship between these models and estimation of production functions. The possible applications of the results from efficiency analyses will be discussed in this chapter as well.

Chapter 4 contains a description of the Danish hospital sector as well as an analysis of the problems and advantages of this structure primarily oriented towards the ability of enabling an efficient use of resources.

In chapter 5 the problems of defining production data for hospitals are discussed. It focusses on the problems related to the definition and measurement of the outputs of hospitals since the difficulties are mainly concerned with establishing relevant measures for the outputs of hospitals whereas the inputs are less problematic.

[1]Chapter 3 is partly based on the EUI Working Paper: "Measuring technical input efficiency for similar production units: A survey of the non-parametric approach"; ECO No. 93/20 (jointly written with J. L. Hougaard).

The empirical applications of the efficiency measurement models appear in chapter 6. Section 6.2[2] includes an efficiency analysis of 80 Danish hospitals where case-mix differences between hospitals are ignored. This is followed up in sections 6.3-6.6 where a sample of Danish hospitals with case- mix dependent output measures is used to analyse the extent of efficiency. In section 6.3 the effects on efficiency with and without case-mix adjustment is analysed. Section 6.4 uses the efficiency measurement models to analyse the extent to which Danish hospitals are producing at an optimal scale. In section 6.5 the efficiency results from 6.3 are compared with efficiency measures from a parametric frontier model in order to examine the similarity of the two sets of results. Section 6.6 includes an analysis of the capacity utilization in the hospitals based on the efficiency measurement models. Section 6.7 contains an efficiency analysis of a sample of private British hospitals.

Chapter 7 is a summary of the main findings in the thesis as well as some suggestions for future research.

[2]Section 6.2 is based on the EUI Working Paper: "Measuring technical input efficiency for similar production units: 80 Danish hospitals"; ECO No. 93/36 (jointly written with J. L. Hougaard).

Chapter 2

A survey of the frontier approach and its relation with other performance evaluation methods.

In this chapter the non- parametric measurement of productive efficiency, the chosen approach of the this thesis, will be surveyed and related to other methods available for performance evaluation of producers. Firstly it will describe which methods can be applied in order to analyse related performance aspects. Then, the focus turns to methods that can be used to analyse the limited aspect of performance related to the transformation of inputs into outputs, including the frontier approach. Then the different models within the frontier approach are surveyed in order to relate the subject of the present thesis with other available methods.

The present thesis is concerned with methods to evaluate the performance of producers' transformation of inputs to outputs. This process can be viewed as a sub-process in the transformation of resources to utilities, which is one of the main elements in any economic system. The transformation process of resources to utility can be characterised in the following way:

Resources are transformed to inputs which, through a production process, are transformed to outputs. These outputs induce effects which can give utility to the users of the outputs and to non- users if external effects are related to the consumption of the outputs. The separation of outputs and effects of outputs might seem strange from a standard economic theoretical point of view, where it is argued that it is the consumption of outputs which provide utility. However it should be admitted that the presented analysis also can be valid since it should be the effects from outputs which have importance for the utility gained. Indeed this approach have been used in health economics, see Mcguire et al. (1988). The examination of this process can take different levels as shown in figure 1. The broadest and most comprehensive analysis is the cost-benefit analysis, where the resource use is compared

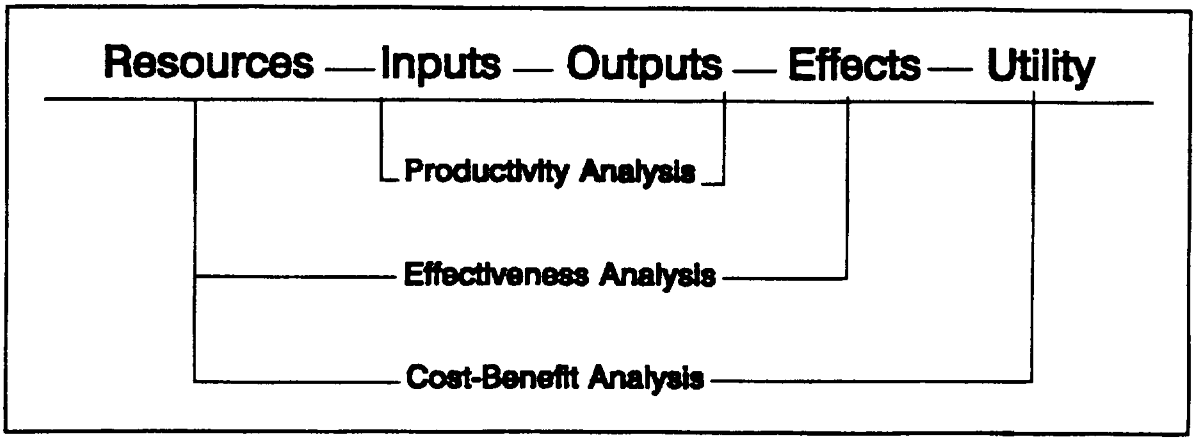

Figure 1. The transformation of resources to utility.

Source: Christensen et al. (1991).

to the utilities obtained from a project, see e.g. Gramlich (1984) for an examination of cost-benefit analysis. In practical cost-benefit analysis it is however the costs of project which are compared to an estimated money value of the benefits. However, as it is the broadest analysis method, it has also many problems. As an alternative the less demanding effectiveness analysis (cost- effectiveness analysis) is available. This method examines the relation between resources (very often approximated by inputs) and effects of outputs in order to analyse whether the effects are provided efficiently; that is whether the same level of effects could be obtained by a lower level of resources (e.g. by changing to more "effect"-efficient outputs. Thus, the outputs in effectiveness analysis are not taken as given. However even data on the effects of outputs can, in many situations, be difficult to obtain since it has to be justified that the measured effects come from the output and are not due to some external factor. This brings us to the third type of analysis, productivity analysis, which analyses the relationship between inputs and outputs; that is the efficiency in the production process of transforming inputs to outputs is examined. In productivity analysis the outputs from a micro producer are taken as given, so that the efficiency of the outputs with respect to effects are not considered. In principle, it is this aspect of the transformation of resources to utility that is in focus in this thesis. In general the efficiency measurement methods have been used to examine the efficiency of micro level production processes. However, it should be noted that, to the extent that data can be provided for the effects of outputs, such information can easily be included in the efficiency measurement methods, and therefore, these methods should be renamed as effectiveness measurement methods.

Having restricted the subject in question to the measurement of the performance of micro units' in their transformation of inputs to outputs, several methods can be applied to examine this aspect. The following list is not exhaustive, but includes some of those most frequently used in empirical analyses: (1) Ratio analysis, (2) Average production function estimation, (3) Frontier models.

Ratio analysis has been used intensively in the evaluation of the transformation of inputs to outputs by micro units. The method is based on the so-called productivity ratio:

$$Productivity = \frac{weighted\ outputs}{weighted\ inputs}$$

In order to use this method it is necessary to obtain data on the micro units' outputs and inputs and in addition weights for the outputs and inputs. See Hjalmarsson (1990) for a survey of productivity-based analysis of performance for micro units. The most likely reasons for the popularity of the ratio analysis seem to be that it is simple to construct and apparantly easy to interpret. For some applications see e.g. Gathon (1989), Bradley & Baron (1993) and Andersen (1989). However it is problematic to base an evaluation solely on such ratios since it is difficult to apply in the case of multiple outputs or multiple inputs, where it is necessary to either define weights to compute composite outputs and inputs or distribute inputs on outputs. This gives the method an ad-hoc characteristic. See Bowlin et al. (1985) and Sherman (1984) for critical comparisons of ratio analysis with frontier methods like Data Envelopment Analysis.

An alternative to ratio analysis is the estimation of an average production function based on a sample of observations with data on inputs and a single (composite) output. This requires an a priori definition of the functional form of the relationship between inputs and outputs. Then the estimation can be carried out using standard econometric techniques such as Least Squares. The micro units can be compared in terms of efficiency by looking at the residuals obtained where an observation with a residual equal to zero can be characterised as being of average efficiency. This method has the advantage of allowing the possibility of using statistical tests in order to examine the relationship, such as marginal substitution possibilities and scale properties between inputs and outputs.

However, this method has a number of important drawbacks. First of all it is inconsistent with the theoretical implication of a production function showing the maximum attainable output level for a given level of inputs, since observations can be placed both above and below the estimated function due to the estimation of a central tendency between inputs and outputs. See Aigner & Chu (1968) for a critique of the estimation of a central tendency between inputs and output. Moreover, the method is restrictive because of the a priori choice of functional form which, most likely, will result in misspecification due to the lack of knowledge about the true relation between inputs and output. This point is emphasised in Bowlin et al. (1985). In addition, the relationship between inputs and output is likely to be different for inefficient observations and efficient observations e.g. represented by different substitution possibilities of inputs, see Sengupta (1989) and Lovell (1993)[1]. In terms of comparing productive efficiency for a sample of observations this can be done by relating the observation's output level to the estimated output level, but no information is obtained concerning the observation's distance to maximum output level for the given input level. It should be noted that although the method has been described above with a single output, it can be used in case of multiple outputs by estimating an average cost function, as is done in Feldstein (1967) in his study on production relations for British hospitals. This might be more complicated since it can require the estimation of a system of equations in order to obtain consistent and efficient estimates, see Johnston (1984).

The lack of consistency between average production function estimation and production theory suggest using methods where the relationship between inputs and output shows the maximum attainable output for given inputs in order to reach such consistency. The frontier approach, which will be described briefly below and analysed intensively in the following chapters, includes such methods.

The frontier approach to the measurement of efficiency corresponds to production theory in the sense that observations are compared to a standard which is identical to the theoretical notion of a production function; see Lovell & Schmidt (1988). This approach involves the construction of a production frontier obtained from observed data on inputs and outputs, where these observations are measured on the micro level. The frontier is thus a best-practice frontier with a relative comparison of performance in contrast to a relationship between inputs and outputs as a blue-print technology with an absolute comparison of performance. The comparison of observations to this frontier indicate which observations are efficient and inefficient. Those observations which are situated on the frontier are candidates for being efficient. However, as it will be shown later on, being on the frontier is only a necessary condition but not a sufficient one for efficiency. Inefficient observations are those observations which are not placed on the frontier, i.e. are inside the production possibility set. Moreover, the distance from an observation to the frontier can serve as a procedure for comparing observations with respect to efficiency. Efficient observations have a distance equal to zero, while inefficient observations are some positive distance from the frontier; and the larger this distance is the more inefficient is the observation.

The frontier approach involves two related issues. First, since the production frontier and the production possibility set are generally unknown they have both to be derived from the data on inputs and outputs. A second issue is how to measure the degree efficiency of a given observation relative to this constructed production frontier. In fact this problem consists of two: (i) Whether an input-output combination is efficient and (ii) how to measure the degree of efficiency, i.e. construct an efficient reference against which the given input-output

[1]This is confirmed in Westerberg (1989), where an average production function is compared with a parametric frontier production function for Swedish Dental care. The parameter estimates for the inputs vary considerably between the two functions.

combination can be evaluated, see Pestieau & Tulkens (1993). The frontier approach to efficiency measurement covers competing proposals to both issues. Notice, however, that the efficiency measurement according to the frontier approach does not, as such, provide explanations for inefficiency that could serve as a background for theories of inefficient micro unit behaviour; see Bs (1988).

For general surveys of the frontier approach see e.g. Färe et al. (1985), Schmidt (1985-86), Aguilar (1988), Sengupta (1989) and Lovell (1993).

In general, productive efficiency for a micro unit has two components: technical efficiency and economic efficiency (or allocative efficiency). Technical efficiency refers to whether, for given inputs it is possible to increase outputs or for given outputs to decrease inputs, i.e. whether a point is situated on the frontier or not. Economic efficiency refers to whether the inputs and outputs are in optimal proportions in terms of the prevailing input and output price ratios. In this case a behavioural goal for the micro unit is defined, such as cost minimisation or revenue maximisation, and observed costs or revenues are compared with optimal costs (minimum costs) and optimal revenues (maximum revenues) respectively. Technical efficiency is examined when data is quantitative based, while analyses of economic efficiency requires data on both quantities and prices.

Farrell (1957) showed in his seminal paper on efficiency measurement how the overall efficiency for an observation can be decomposed into technical and allocative efficiency. Consider a situation where two inputs, x_1 and x_2, produce a single output y and assume that the production technology has constant returns to scale so that the Farrell decomposition can be shown in a single isoquant diagram as shown in figure 2.

The isoquant SS' represents the technically efficient combinations of the two inputs to the given output level. Points above the isoquant are technically inefficient since it is possible to reduce the inputs and for the output level y still to be feasible. However not all points on the isoquant are also allocatively efficient in the sense that the inputs are employed in the optimal proportions corresponding to the input price ratio p_1/p_2. The point P is inefficient, Q is technically efficient but allocatively inefficient if the input price ratio is as represented by the line AA'. In this case only Q' is technically and allocatively efficient; the point R has the same costs as Q'. Notice however that there exists a price ratio which will make Q allocatively efficient.

The Farrell decomposition of the inefficiency of P can now be derived as follows. Technical efficiency can be measured as $\frac{OQ}{OP}$ and overall efficiency can be measured as $\frac{OR}{OP}$ according to the observed costs for P relative to the minimum costs of producing the output level y. Therefore allocative efficiency can be computed as:

$$\frac{OQ}{OP}/\frac{OR}{OP} = \frac{OQ}{OR}$$

All these efficiency measures take values in the interval]0;1], where a value equal to

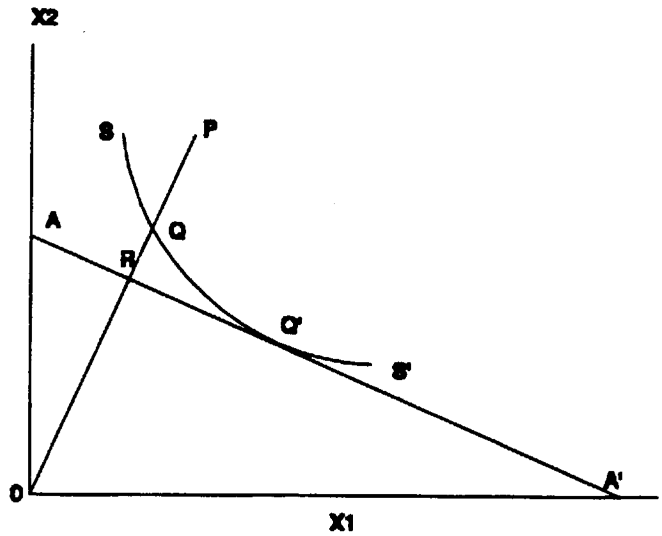

Figure 2. The technical, allocative and overall efficiency measure.

1 implies that the observation is technically or allocatively efficient. In addition to the three micro unit related efficiency measures Farrell (1957) introduced an efficiency measure with respect to the whole sample of observations i.e. a measure of the structural efficiency for an industry which is computed from the single technical efficiency measure for each observation by weighting with the output levels and then summed over the observations. This measure can be used to examine the dispersion of efficiency within an industry, see Førsund & Hjalmarsson (1979).

Two approaches concerning the calculation of efficiency measures have been developed: Radial measures (Debreu (1951) and Farrell (1957)), which are the ones almost always used in empirical applications following the procedure described above, and non- radial measures (see e.g. Färe & Lovell (1978), Färe, Lovell & Zieschang (1983), Zieschang (1984), Russell (1985), Russell (1988), Russell (1990) and Dmitruk & Koshevoy (1991)). The difference between these methods can be explained more deeply by putting aside the problem of how to construct the efficiency frontier. Thus to assume that the frontier is known. Consider a situation where two inputs (x_1 and x_2) are used in the production of one output y. In figure 3 the input requirement set[2] (the set of input vectors that can produce y) is assumed to have the indicated form, where the boundary is identical to the well known isoquant. A point like P is certainly inefficient since it is possible to reduce both inputs and still obtain y. Following the Debreu-Farrell tradition, a radial efficiency measure for the point P will be calculated along a ray from origin through P, where the efficiency measure is equal to $\frac{OR}{OP}(0 < \frac{OR}{OP} < 1)$. This measure indicates that reducing the amount of inputs used at P by

[2]The input requirement set is used solely for graphical reasons

$(1-\frac{OR}{OP})$ gives a feasible input vector placed on the boundary of the input requirement set. The efficient reference is, in this case, R where the inputs used at point P are reduced in an equi-proportionate way to reach this point. The point R would give an efficiency degree equal to 1 if this point were evaluated. However points exist on the isoquant that seem better than R because they use less of the input x_1. The problem arises when the isoquant has horisontal or vertical segments and the observation is located at such a segment. Thus the application of a radial method to obtain the efficiency measure can produce "spurious" information such that observations are termed efficient without being efficient. In the non-radial tradition such a situation cannot exist, since the efficiency measure for an observation is equal to 1 if and only if it is impossible to reduce any input and still obtain a feasible point, i.e. that belongs to the input requirement set. Thus the efficient reference in the non-radial approach for the point P will be the point M. This procedure for computing technical efficiency corresponds to Koopmans' (1951) definition of technical efficiency, where a producer is technically efficient if an increase in any output requires a reduction in at least one other output or an increase in at least one input, and if a reduction in any input requires an increase in at least one other input or a reduction in at least one output. This definition of technical efficiency is thus more restrictive than the Debreu- Farrell radial tradition where a technically efficient producer cannot reduce proportionately the inputs or increase proportionately the outputs. Unfortunately there are also problems involved with these non-radial measures; see Kopp (1981). In particular the interpretation of the efficiency measures is less clear with non-radial measures.

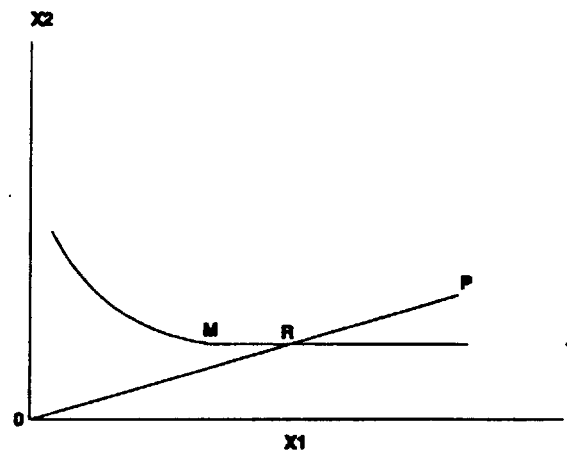

Figure 3. Radial and non-radial efficiency measures.

In relation to the problem of the construction of a production frontier two main alternative methodologies are available - parametric methods and non-parametric methods - with different advantages and problems. With reference to the above discussion of radial and non-radial efficiency measures it should be noted that all the frontier methods described below apply a radial method to the calculation of efficiency measures, once the frontier has been constructed.

The parametric approach assumes, a priori, a specific functional form of the relationship between inputs and output, e.g. Cobb- Douglas, CES or translog, given a set of unknown parameters that are to be estimated. The parametric approaches can be further classified as either deterministic or stochastic. A deterministic frontier function is obtained when the parameters are estimated such that the observations will be placed on or below this function (see e.g. Richmond (1974) and Schmidt (1976)). The residuals can, thus, be taken to be due solely to technical efficiency.. But this is seldom the case. Probably some of the variation in the error term can be explained by other factors outside the control of the evaluated units, such as measurement and stochastic errors. In that case some observations should also be found above the frontier. To some extent this has been solved by so-called stochastic frontier models (originally proposed by Aigner, Lovell & Schmidt (1977) and Meeusen et al. (1977)) where the error term is divided into two components. The first component is assumed to be one-sided reflecting inefficiency, whereas the other error term is a symmetric (two-sided) component which reflects random noise, measurement errors and factors outside the evaluated units control. The efficiency measure for each observation is then determined relative to this estimated function and the symmetric error component. However it is difficult to separate the obtained error term into the two sub-components and the proposals that have attempted to solve this problem are not particularly applicable[3].

An important problem of the parametric approach is that it is difficult to use in the case of multiple outputs. This problem is especially relevant for efficiency evaluation in the public sector where prices generally do not exist prices that otherwise could have been used to obtain a weighted output measure. In order to deal with both multiple inputs and multiple outputs in a parametric approach it is necessary to estimate a cost function, which in addition to data on inputs and outputs requires data on input prices. The estimation of a cost function can either be formulated as a single equation estimation or as a system with multiple equations with a cost function and input share equations. The multiple equation approach for estimation of cost frontiers was initially proposed by Lovell & Schmidt (1979). The estimation of a single equation implies that the obtained efficiency measures are containing technical inefficiency and allocative inefficiency, which can be decomposed as proposed by Kopp & Diewert (1982). The estimation of a single equation based cost function is inefficient compared to the estimation of multiple equation cost functions. The problem with that procedure is however to decompose the efficiency into technical and allocative parts. See Bauer (1990) for an analysis of the possible ways to make such a decomposition.

Furthermore, and more importantly it seems to be too restrictive that the approach

[3]See e.g. Jondrow et al. (1982) and Sengupta (1989).

demands an a priori specification of the frontier since the frontier is unknown generally. There is no reason why the choice of function should be correct.

For surveys of the different methods in the parametric approach see Førsund et al. (1980), Bauer (1990) and Greene (1993). As examples of recent empirical studies in the parametric tradition one could mention: Perelman & Pestieau (1988) who analyse the efficiency of European postal services and railways using a deterministic method and the production function is assumed to be of the translog kind. In Bjurek et al. (1990) efficiency in Swedish local social insurance offices is studied using a Cobb-Douglas production frontier and a deterministic approach. Green et al. (1991) analyse efficiency in the UK and Australia manufacturing industry using a stochastic frontier method where the production frontier is specified as a translog function. In Ferrier & Lovell (1990) a stochastic cost function is used (plus input share equations) in order to study the efficiency of US banks.

The non-parametric approach for obtaining the production frontier does not, a priori, specify a functional form of the relationship but imposes some properties concerning the production possibility set, e.g. convexity and free disposability of inputs and outputs[4]. These properties are assumed to be satisfied by the data. For each observation it is, then, determined whether the observation could be considered as a member of the frontier given the specified assumptions concerning the production possibility set.

The different methods of the non-parametric approach diverge with respect to the specified assumptions about the production possibility set. Farrell (1957) is often viewed as the starting point for empirical analyses of efficiency using a non- parametric approach. The assumptions he imposed on the production possibility set were convexity, free disposability of inputs and outputs and constant returns to scale. This approach is adopted in the so-called Data Envelopment Analysis (DEA) whose development was started by Charnes, Cooper & Rhodes (1978), (1981). It was shown that the Farrell method could be formulated as an LP-problem but the assumptions concerning the production possibility set were essentially the same. In Banker, Charnes & Cooper (1984) the DEA model was extended with less restrictive scale assumptions, i.e. non-increasing returns to scale and variable returns to scale. The measurement of technical efficiency with respect to production technologies with different scale assumptions makes it possible for DEA to determine for each observation the extent of technical inefficiency and scale inefficiency. Optimal scale is defined as operating at constant returns to scale (maximum average productivity) whereas scale inefficiency implies that an observation is operated at decreasing or increasing returns to scale. The use of DEA to examine scale properties has been studied in Banker (1984), Banker, Charnes & Cooper (1984) and Banker & Thrall (1992), see also Chang & Guh (1991) for a critique view of this application of DEA.

Another line of extension of the standard DEA model has been through the relaxation of

[4]See Varian (1984) for formal definitions of these concepts.

the assumption on free (strong) disposability of inputs and outputs due to the possibility that inputs might not be disposable and outputs unwanted. The extension of DEA to allow for inputs and outputs to be only weakly disposable implies that the technical efficiency measure can be decomposed into pure technical efficiency, scale efficiency and congestion; see Färe & Svensson (1980), Färe & Grosskopf (1983), Färe, Grosskopf & Lovell (1983), Färe & Grosskopf (1985) and Färe, Grosskopf & Lovell (1987).

Although the preceding survey of DEA has dealt with technical efficiency it is, indeed, possible to compute allocative efficiency measures within a DEA model, see e.g. Banker & Maindiratta (1988) and Lovell (1993) for such models. The former approach provides upper and lower bounds on the technical, allocative and overall efficiency measures, whereas only a single measure is obtained in the latter approach. In Morey, Fine & Loree (1990) and Ferrier & Lovell (1990) empirical measures of allocative efficiency are computed along with technical efficiency. As stated earlier the calculation of allocative efficiency measures requires the availability of input price data.

For surveys of recent developments in DEA, see Sengupta (1989), Charnes & Cooper (1990), Seiford & Thrall (1990) and Ali & Seiford (1993). Empirical applications of DEA are numerous, as illustrated in the Seiford (1990) bibliography on DEA which includes more than 400 works. As examples of some of the most recent studies the following can be mentioned: Banker, Conrad & Strauss (1986) analyse North Carolina hospitals, Lovell, Walters & Wood (1990) study US schools, Bjurek, Kjulin & Gustafsson (1992) study efficiency at public day care centres in Sweden, Berg, Førsund, Hjalmarsson and Suominen (1993) study efficiency of Nordic banks and Kittelsen & Førsund study efficiency of Norwegian courts.

Deprins, Simar & Tulkens (1984) propose a non-parametric method (the so-called FDH method) which only imposes an assumption about the free disposability of inputs and outputs, thereby removing the convexity assumption of the DEA models. Thus the frontier is found as the boundary to the free disposal hull.

For surveys of the FDH analysis see Thiry & Tulkens (1989) and Tulkens (1993). Empirical applications of this method are few but the following examples can be given: Deprins, Simar & Tulkens (1984) analyse the Belgian postal service, Thiry & Tulkens (1988)[1] analyse Belgian bus companies, Tulkens (1990) study Belgian courts, Vanden Eeckaut, Tulkens & Jamar (1993) study Belgian municipalities and Fried, Lovell & Vanden Eeckaut (1993) study US credit unions.

A mixed position between assuming convexity of the production possibility set (DEA) and not asssuming convexity at all (FDH) is Petersen (1990) who relaxes the convexity of the production possibility set, but retains the assumptions of convexity of the input set and of

[1]This application is very interesting since it combines a parametric method with the FDH model: the FDH model is used to determine which units are efficient. These efficient observations are then used in a second step to estimate a translog production function with a two-sided error term.

the output set.

The main advantage of the non-parametric methods compared with the parametric approach is that they require few a priori assumptions. In particular, the FDH model is free from restrictive assumptions with the only assumption being on free disposability of inputs and outputs. Furthermore, DEA and FDH are directly applicable in the case of multiple inputs and outputs. However the standard models are deterministic in the sense that all deviation is interpreted as technical inefficiency. Possible noise around the frontier is ignored; see Tulkens (1993).

In the choice between a parametric approach and a non-parametric approach the following remarks are relevant: the parametric approach allows for not defining all deviation from the frontier as due to inefficiency. This is not the case in the standard models in the non-parametric approach. On the other hand the non- parametric approach does not require an a priori choice of functional form of the relation between the inputs and outputs. This is not the case in the parametric approach, where a specific function has to be imposed a priori. Thus the parametric approach implicitly assumes a high information level with respect to the production technology for the sample of observations to be analysed. On the other hand the non-parametric approach assumes that no information about the production technology is available, except that the free disposability of inputs and outputs and the convexity of the production possibility set (if DEA is used) are satisfied for the sector to be analysed.

In most applied studies the actual information level with respect to the production technology lies between these two extremes. The analyst cannot propose a particular functional relationship but will, in general, know something about the production relationship, i.e. more information than implied in the non-parametric efficiency measurement models. A possible solution could be to include such information in the non-parametric modelling of the production technology. It should be noticed that such a step removes the generality of the non-parametric approach.

The above comments seem to indicate a call for mixed approaches or methods which combine the attractive properties from the parametric and the non-parametric approaches. Attempts have been made to introduce a stochastic element within the DEA models in order to remove the problem of true inefficiency being confounded by noise. One such approach is to analyse the deviations between the frontier and the observations using a non-parametric sample estimator of the distribution of deviations; see Sengupta (1989) and Sengupta (1990). Alternatively, the efficiency distribution could be obtained by the use of the non-parametric "bootstrap" technique; see Härdle (1990). For applications of the bootstrap technique see Simar (1991) who studies the efficiency distribution for European Railway companies and Färe & Whittaker (forthcoming) who study the efficiency distribution of dairy producers in the US. Another approach adopted to introduce a stochastic element into DEA models is through chance-constrained programming which transforms the deterministic inequalities in the DEA LP problem into probability statements. This approach has been followed in Land, Lovell & Thore (1988) and Petersen & Olesen (1989).

Alternatively the starting point could be with the parametric approach and not taking the functional form as known. This could be obtained by choosing non-parametric regression techniques, where no fixed form is assumed only that the relation between inputs and outputs belongs to some class of functions, e.g. concavity, see e.g. Härdle (1990) for an examination of non-parametric regression techniques. However, the mixed approaches are still in their initial phase and no conclusion about the merits of these can yet be reached.

Chapter 3

Non-parametric efficiency measurement models

3.1 Introduction

Performance evaluation of production units based on traditional production theory has received growing interest during the past decade. The so called Data Envelopment Analysis method (the DEA method as proposed by Charnes, Cooper & Rhodes (1978)) and Free Disposable Hull method (the FDH method as proposed by Deprins, Simar & Tulkens (1984)) relate to evaluation of similar production units (or Decision Making Units as they are often called) where outputs and inputs are measurable and data on prices are not necessarily available. These characteristics are often found in production units which provide services such as libraries, hospitals etc. It is on the productive efficiency aspect of performance of such organisations that these methods focus. In the following it is mainly the technical side of productive efficiency that is considered, in other words the organisations' ability to transform multiple inputs via a production process into outputs.

These methods are non-parametric since this approach do not impose any a priori assumptions on the functional relationship between inputs and outputs in the production process. In particular, the free disposal hull procedure represents the closest approach to an evaluation procedure completely based on the information obtained from the production dataset.

This chapter is organised as follows. Section 3.2 contains some preliminary definitions from production theory. Given a set of observations on a group of similar activities the aim is to filter these activities such that good and bad performances are identified. One can attempt to apply such a filtering through the non-parametric approaches, FDH and DEA, as examined in section 3.3. In section 3.4 different efficiency measures are analysed and explained and their relation to production function estimation is shown. The following section 3.5 examines the possible applications of the results from a DEA or FDH analysis. Section 3.6 presents a non-radial efficiency measure. The next sections includes analyses of how to allow for categorical variables (sec. 3.7) and quality variables in DEA (sec. 3.8). In section 3.9 the purely technical information obtained from the efficiency scores are extended by the inclusion of institutional factors. Some part of the technical inefficiency may be explained by such factors in a regression approach. Section 3.10 extends the previous static analysis to cover dynamic aspects. Finally 3.11 discusses the possible advantages and problems of DEA and FDH analysis.

3.2 Preliminaries

Let $\mathbf{R}_+^m$ denote the non-negative Euclidean m-orthant. For every $x, x' \in \mathbf{R}_+^m$ I write $x \geq x'$ and $x > x'$ respectively, if for every $i = 1, ..., m, x_i \geq x_i'$ and $x_i > x_i'$ respectively. I shall consider the production of s outputs denoted by $y \in \mathbf{R}_+^s$ from a set of m inputs, where inputs are denoted by $x \in \mathbf{R}_+^m$.

Define the production possibility set as $Y = \{(x, y) | x \in \mathbf{R}_+^m, y \in \mathbf{R}_+^s, (x, y) \text{ is feasible}\}$. The production possibility set is said to satisfy free disposability of inputs if $(x, y) \in Y$ and if $x' \geq x$ then $(x', y) \in Y$. Likewise it is said to satisfy free disposability of outputs if $(x, y) \in Y$ and if $y' \leq y$ then $(x, y') \in Y$. Free disposability of inputs means that if a given input-output combination belongs to the production possibility set (i.e. a feasible combination of inputs and outputs) and there is another input-output combination characterised by the same output level but with a higher level of some inputs, then this combination will also belong to the production possibility set (i.e. this combination of inputs and outputs is also feasible). Likewise, free disposability of outputs means that if an input-output combination belongs to the production possibility set and another input-output combination is characterised by the same input level but with some outputs at a lower level then this combination will also belong to the set. The assumption of free disposability of inputs and outputs is quite standard in economic models and should be regarded as a rather weak assumption.

Let $L(y) = \{x | (x, y) \in Y\}$, the input requirement set, where this set includes those input vectors which can produce the given output vector y and let $Q(x) = \{y | (x, y) \in Y\}$, the output possibility set, where this set includes those output vectors which can be produced given the input vector x. $x \in L(y*)$ is an efficient input vector for $y*$ if there is no $\mu \in [0, 1[$ such that $\mu x \in L(y*)$. Likewise $y \in Q(x*)$ is an efficient output vector for $x*$ if there is no $\delta \in]0, 1[$ such that $y/\delta \in Q(x*)$. The frontier S of Y is then defined as $S(Y) = \{(x, y) | x \text{ is efficient and } y \text{ is efficient}\}$. Efficiency is in this way defined as no possibility for proportionate reductions in inputs and proportionate augmentations of outputs.

Define both the dominated and the dominant set of a given production vector (x', y') as $DO(x', y', Y) = \{(x, y) \in Y | x \geq x' \text{ and } y \leq y'\}$ and $D(x', y', Y) = \{(x, y) \in Y | x \leq x' \text{ and } y \geq y'\}$ respectively. The dominated (dominant) set includes those input-output combinations which use the same or more (less) of the inputs to achieve the same or less (more) output levels compared with the production vector (x', y').

Let $Y_0 = \{(x_k, y_k) | k = 1, ..., n\}$ be a set of n production vectors e.g. observed production activities from n production units (DMUs). Moreover, let the production possibility set be defined by the free disposal hull technology of Y_0 - that is:

$$Y_{FDH} = \{(x, y) | (x, y) = (x_k, y_k) + \sum_j \mu_j [e_j^m, 0^s] - \sum_i \tau_i [0^m, e_i^s],$$

$$(x_k, y_k) \in Y_0 \cup \{0^m, 0^s\}, \mu_j \geq 0, \tau_i \geq 0, j = 1, ..., m, i = 1, ..., s\},$$

where e_i^s is the i'th column of the s-dimensional identity matrix and e_j^m is the j'th column of the m-dimensional identity matrix. Obviously this set satisfies the free disposability requirements defined above. Notice that the free disposal hull technology can be expressed as $Y_{FDH} = \cup_{k=1}^n DO(x_k, y_k) \cup \{0^m, 0^s\}$.

Consider the following simple example concerning the production activities of six DMUs each producing one output (y) by one inputs (x):

DMU	(x, y)
A	(3,2)
B	(4,4)
C	(6,5)
D	(7,5)
E	(5,2)
F	(3,1)

Figure 1 illustrates the free disposal hull technology based on the six observations.

If the assumption that the production possibility set satisfies convexity is added it becomes necessary to distinguish between several types of technologies, each identical to the free disposal convex hull (FDCH). [1] If, for example, the technology is characterised by constant returns to scale (which was what Farrell (1957) originally assumed) the following possibility set is obtained:

$$Y_{CON} = \{(x, y) | (x, y) = \sum_k \Psi_k (x_k, y_k) + \sum_j \mu_j [e_j^m, 0^s] - \sum_i \tau_i [0^m, e_i^s],$$

[1] A convex production possibility set implies that if two input-output combinations belong to the production possibility set then the linear combination of the two will also be in the set; i.e. will provide a feasible input-output combination. Assuming convexity is quite common, but must be viewed as restrictive. At least in the short run, it can be difficult to adjust the input-output levels of some observations in order to obtain a linear combination. However in the long-run, imposing convexity seems less demanding.

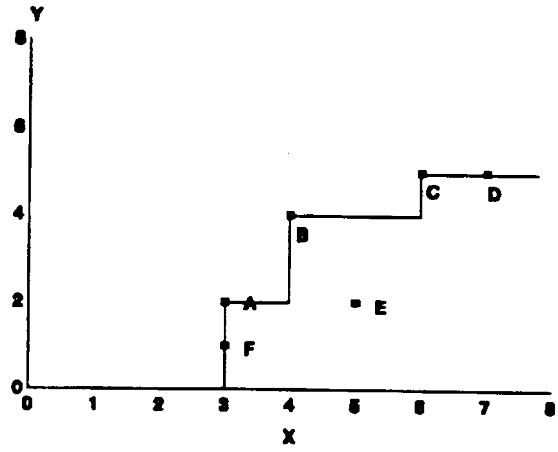

Figure 1. The free disposal hull technology of A, B, C, D, E, F.

$$\Psi_k \geq 0 \forall k, (x_k, y_k) \in Y_0 \cup \{0^m, 0^s\}, \mu_j \geq 0, \tau_i \geq 0, j = 1, ..., m, i = 1, ..., s\}.$$

Y_{CON} is different from Y_{FDH} because linear combinations of observations are allowed to be elements of the possibility set as well as the observations themselves. If further the assumption that $\sum_k \Psi_k \leq 1$ is added, the technology will be characterised by decreasing returns to scale. Denote by $Y_{DEC} = \{Y_{CON}$ and $\sum_k \Psi_k \leq 1\}$ the decreasing returns to scale FDCH-technology. If the zero-vector $(\{0^m, 0^s\})$ is excluded and $\sum_k \Psi_k = 1$ is added, the technology is characterised by variable returns to scale. Denote by $Y_{VAR} = \{Y_{CON} \setminus \{0^m, 0^s\}$ and $\sum_k \Psi_k = 1\}$ the variable returns to scale FDCH-technology (note that due to convexity, increasing returns to scale is excluded as a general property and only possible if it is followed by decreasing returns to scale). Notice that Y_{VAR} is contained within Y_{DEC} which, again, is contained within Y_{CON}, see Grosskopf (1986). Thus in general:

$$Y_{FDH} \subseteq Y_{VAR} \subseteq Y_{DEC} \subseteq Y_{CON}$$

Figure 2 illustrates the different technologies in the simple example of one-input-one-output case.

3.3 Filtering the data set

Given the data set Y_0 the data can be divided into two subgroups according to their reflected performance level. Ideally this partitioning would result in one group containing units with good performances, i.e. the efficient DMU's and another group containing those with bad performances, i.e. the inefficient DMU's.

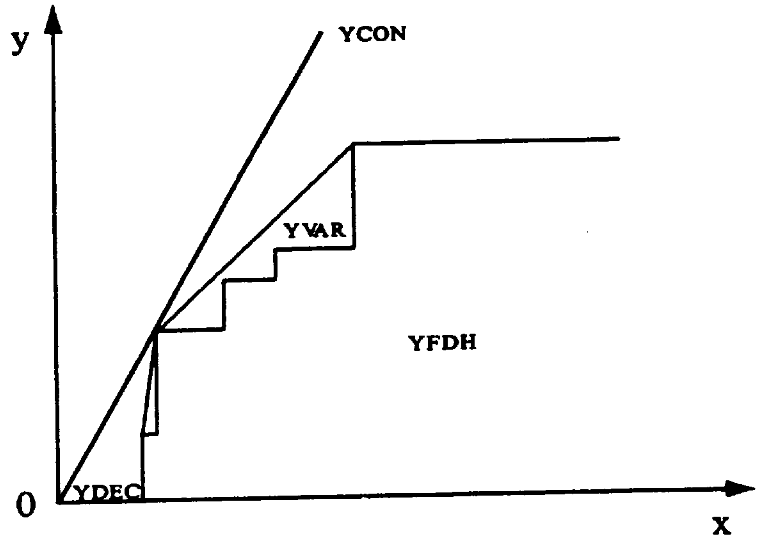

Figure 2. DEA and FDH production technologies.

Among others, such a partition is possible through the use of a free disposal hull technology filter, though one can discuss whether the procedure is fair to all DMU's.

3.3.1 The FDH-procedure

The free disposal hull technology provides a natural partitioning of a data set since the technology is based on a principle of dominance. Dominance appears to be a well suited concept for the filtering of the DMU's because if one DMU dominates another then it uses less of at least one of the inputs and achieves at least the same level of outputs and possibly more of some of them.

Formally, look at each element in Y_0 and construct the set $D(x_k, y_k, Y_0)$ where $(x_k, y_k) \in Y_0$.

Definition: The k'th DMU is not dominated if and only if $D(x_k, y_k, Y_0)$ is a singleton (only contains the k'th element), otherwise the DMU is dominated.

Due to the definition of dominance there exists a group of DMU's which are not dominated yet not dominating any other DMU. This is obviously the case for highly specialised units where production is limited in either inputs or outputs but also for units which are dominated in some dimensions and dominating in others.

Empirical work by Tulkens (1990) seems to indicate that a large part, 50-90 %, of the data set will be declared undominated and of this part about 50-70 % are non-dominating units. In general the number of undominated yet non-dominating units increases when the dimension of the product vector (x, y) increases and when the number of units (n) decreases. Thus there does not exist a clear cut relationship between dominance and efficiency. Though it would be convenient to declare efficient all the undominated units it raises serious questions reguarding the role of the undominated but non-dominating units. The question is whether it is fair to assume that such units are efficient or whether they have obtained their position either due to the way dominance was defined or due to pure specialisation. This is of particular importance when the strategic responses of the DMUs to FDH-evaluation (or control) are considered.

In general the FDH-procedure can be characterised as highly unrestrictive since the only assumption included is free disposability of inputs and outputs. This makes it difficult to characterise the efficient units as illustrated above. However, the dominated units are easier to characterise as inefficient since they are declared so under very weak restrictions.

3.3.2 The DEA-procedure

Convexity of the production possibility set is the underlying assumption of the DEA-procedure. Therefore the choice between FDH and DEA methods relates fundamentally to whether one accepts the assumption of convexity or not. In DEA the data filtering is not directly built upon the concept of dominance, rather the procedure uses some kind of collective production function defined in section 3.2 as the frontier of the chosen FDCH-technology. Hence, the observations placed on this frontier are declared efficient and the elements in the interior of the FDCH-technology are declared inefficient, that is:

Definition: The k'th DMU characterised by $(x_k, y_k) \in Y_{FDCH}$ is efficient if and only if $(x_k, y_k) \in S(Y_{FDCH})$ and is otherwise inefficient.

Two notable properties occur in relation to the FDH-procedure. Firstly, an undominated observation in Y_{FDH} will not necessarily be an element of the FDCH-frontier and hence efficient. On the other hand an element of the FDCH-frontier will always be undominated in Y_{FDH}, as illustrated in figure 3. This follows directly from the fact that $Y_{FDH} \subseteq Y_{FDCH}$ as mentioned in section 3.2. Secondly, the dominated observations are not only dominated by observations in Y_0 but also by linear combinations of some of these observations. This fact may have consequenses which concerns the implementation of the filtering result. It is easier to explain to the manager of a dominated production unit that he is dominated by another, actually existing, unit than by a linear combination of e.g. the division in Copenhagen and the division in Stockholm and, perhaps more importantly, it is easier to suggest efficiency improvement strategies since it is straightforward to copy the strategy of the dominating (and actually observed) unit.

Also important is the fact that the partitioning of the data set depends to a grat degree on the assumed type of underlying FDCH-technology. In figure 2 it is easy to see that the same observation can be declared either efficient or inefficient depending upon the type

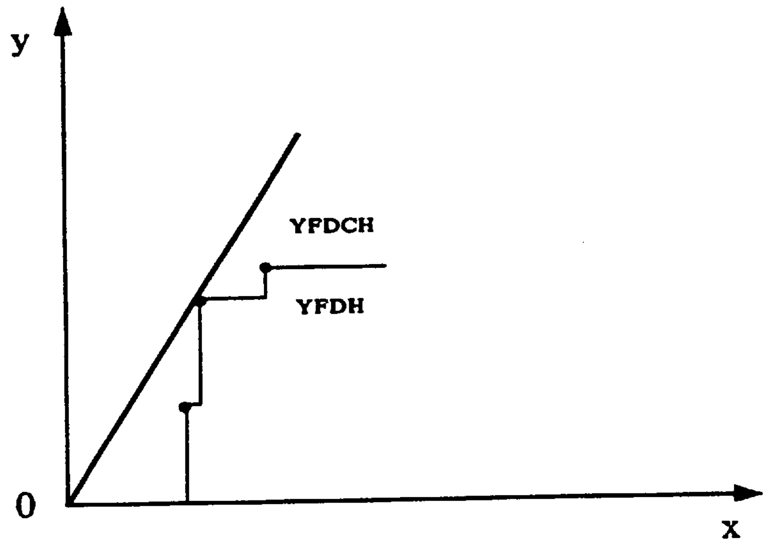

Figure 3. Undominated observations are not necessarily efficient.

of FDCH-technology (Y_{CON}, Y_{DEC} or Y_{VAR}) chosen. Since $Y_{VAR} \subseteq Y_{DEC} \subseteq Y_{CON}$, efficient observations in Y_{CON} will also be efficient under both Y_{DEC} and Y_{VAR}-technologies. Likewise efficient observations in Y_{DEC} will also be efficient under the Y_{VAR}-technology.

It is worth noticing that an assumption of convexity may seem inappropriate in the short run since activities constructed as linear combinations of existing activities are functioning as efficient references. Such hypothetical observations will rarely be attainable in the short run because of fixed technological constraints supporting of free disposability as the sole assumption on the technology. The efficient references in the free disposal hull technology are existing (and dominating) activities and hence their performance level ought to be attainable even in the short run. However, in the long run convex combinations may, of course, represent attainable activities.

3.4 Measures of efficiency

The main purpose of the data filtering was to divide the DMUs into two subgroups; one containing the efficient and the other containing the inefficient units. The next natural question is; how inefficient are the inefficient units? Obviously great injustice can be done if the group of inefficient units is considered as one. In other words information about the degree of efficiency of each unit is needed. Such information can be obtained for example, through efficiency indices which can be either radial as treated in this section, or non-radial, as treated in 3.6.

3.4.1 Radial efficiency indices

In economic theory the radial efficiency measures have a long history. However, the concept of technical efficiency and hence the measurement of technical efficiency is considered to be introduced by Farrell (1957), who again was inspired by Debreu's coefficient of resource utilisation (Debreu (1951)). These can be either input or output efficiency indices. Farrell's efficiency indices are radial indices in the sense that the input (output) efficiency measures the efficiency of an observation along a ray from the origin in the input (output) space to the frontier of the production technology .

Definitions: Considering output as fixed Farrells index of input efficiency can be defined as:

$$E_F^i(x, L) = \min\{\theta | \theta x \in L\},$$

where $L = \{x | (x, y) \in Y\}$.

Considering input as fixed Farrells index of output efficiency can be defined as:

$$E_F^o(x, L) = \min\{\delta | y/\delta \in Q\},$$

where $L = \{x | (x, y) \in Y\}$. Notice that $0 < E_F^i \leq 1$ for $x \in L$ and $0 < E_F^o \leq 1$ for $y \in Q$. Thus E_F^i (E_F^o) measures the maximal proportionate reduction (augmentation) in inputs given the feasibility constraint $x \in L$ ($y \in Q$). The radial input efficiency index is illustrated by figure 4.

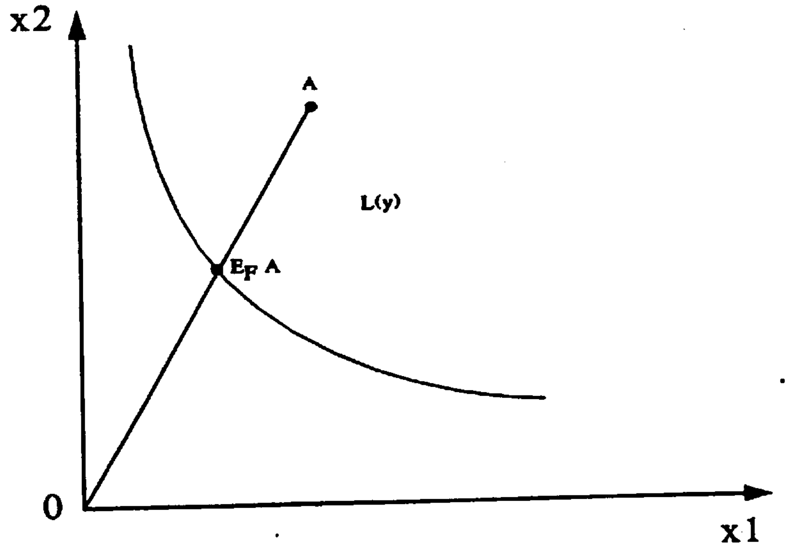

Figure 4. Farrell's index of input efficiency.

3.4.2 DEA-V efficiency measures.

First I will consider how radial efficiency measures for single observations can be constructed with respect to the DEA-C, DEA-D and DEA-V frontiers. If the different efficiency measures is constructed with variable returns to scale then the following can be established with respect to observation E from the above-mentioned example; see figure 5. These definitions was proposed in Førsund & Hjalmarsson (1979), see also Førsund & Hjalmarsson (1987):

- The pure technical input efficiency measure: $E_1 = GH/GE = 3/5$
- The pure technical output efficiency measure: $E_2 = JE/JK = 4/9$
- The gross scale efficiency measure: $E_3 = GI/GE = JE/JL = 2/5$
- The input scale efficiency measure: $E_4 = E_3/E_1 = (GI/GE)/(GH/GE) = 2/3$
- The output scale efficiency measure: $E_5 = E_3/E_2 = (JE/JL)/(JE/JK) = 9/10$

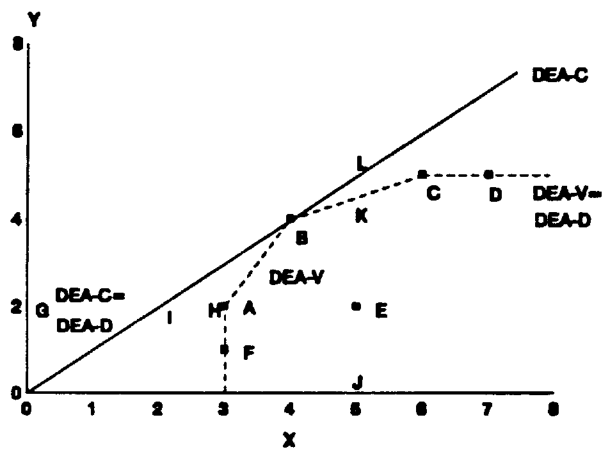

Figure 5. DEA efficiency analysis of A, B, C, D, E, F.

All values are constructed such that they take values in the interval]0; 1]. E_1 measures the ratio of required input to the observed input level for observation E given its output level (required to be situated on the DEA-V frontier). If $E_1 = 1$ it would have implied that E was DEA-V efficient, but E_1 is equal to 3/5 so E is inefficient in terms of inputs, this observation has produced its level of output using an excessive amounts of the input. Thus this measure shows how much of the input E should use in order to be DEA-V efficient. $(1 - E_1)$ indicates how much, in percentage term, E could reduce its input level.

E_2 measures the ratio of observed output to potential output (potential output with respect to the DEA-V frontier). It shows the short fall between the actual level of output and the potential output, given the input level. If E_2 is equal to 1 then the actual output level is identical to potential output and thus no inefficiency is present. Otherwise E_2 is less than 1 and it is possible to increase the output level without increasing the input level. $(1 - E_2)$ expresses the percentage by which the output should be increased in order to be on the DEA-V frontier.

E_3 measures the ratio of the optimal input level to the observed input level, output being taken as given. As a general property, the same E_3 is obtained if it is expressed by the ratio of observed output to potential output given the input level. E_3 can contain two sources of inefficiency : pure technical inefficiency equivalent to E_1 (or E_2) and scale inefficiency due to a position less than the optimum. Therefore $E_1 \geq E_3$ (and $E_2 \geq E_3$).

E_4 and E_5 can be seen as measures of pure scale inefficiency due to the elimination of the pure technical inefficiency component in E_3, computed either in terms of input (using E_1 to eliminate pure technical inefficiency) or in terms of output (using E_2). A value of E_4 equal to 1 implies that $E_1 = E_3$, i.e. that all inefficiency (if any) stems from excessive input usage but not from a non-optimal scale that gives smaller than maximum productivity. The same interpretation holds for E_5. See e.g. Banker (1984) for a similar analysis of scale efficiency.

The analysis of the efficiency measures E_1 to E_5 has been developed with respect to a DEA-V frontier. Thus in order to calculate E_3 (to relate the input usage of observation E to the point I) it is necessary to extrapolate from B to the point I. I is a hypothetical observation with the same productivity level as B but with an output level corresponding to E. Alternatively, if the efficiency measures E_1 to E_5 alternatively are constructed in a set-up where the production technology is assumed to satisfy constant returns to scale then the frontier corresponds with the line of maximum productivity through B. Notice that in a DEA-C model $E_1 = E_2 = E_3 = E_4 = E_5$ since all possible inefficiencies are interpreted as being technical and there is no distinction made between output or input based measures.

The returns to scale for a given observation can be determined from the E_1 and E_3 measures based on the DEA-V model; see Ferrier & Lovell (1990) and Hjalmarsson (1990). If $E_1 = E_3$ then constant returns to scale are present and if $E_1 \neq E_3$ (that is $E_1 > E_3$) either decreasing or increasing returns to scale are present. In order to judge whether decreasing or increasing returns to scale prevail it is necessary in addition to calculate E_1 with respect to the DEA-D frontier[2]. If $E_{D1} = E_{V1}$ decreasing returns to scale are present and if $E_{D1} < E_{V1}$ increasing

[2]The DEA-D frontier is equivalent to the DEA-C frontier until observation B and then follows the DEA-V frontier.

returns to scale are present[3]. In the example the returns to scale for the 6 observations can be outlined as follows:

$$A : E_{V1} = 1, E_{V3} = E_{D1} = 2/3 \Rightarrow IRS$$
$$B : E_{V1} = E_{V3} = E_{D1} = 1 \Rightarrow CRS$$
$$C : E_{V1} = E_{D1} = 1, E_{V3} = 5/6 \Rightarrow DRS$$
$$D : E_{V1} = E_{D1} = 5/6, E_{V3} = 5/7 \Rightarrow DRS$$
$$E : E_{V1} = 3/5, E_{V3} = E_{D1} = 2/5 \Rightarrow IRS$$
$$F : E_{V1} = 1, E_{V3} = E_{D1} = 1/3 \Rightarrow IRS$$

Alternatively, the scale properties can be determined by comparing E_1 and E_2 both calculated with respect to the variable returns to scale frontier. If $E_1 > E_2$ it implies increasing returns to scale and if $E_1 < E_2$ the observation shows decreasing returns to scale. This property can easily be understood. Consider observation D from the example where $E_1 = 5/6$ but $E_2 = 1$. Thus by reducing the input usage E_1 will increase and E_2 will remain unchanged until $E_1 = 1$ equal to observation C. This implies that a move from C to D, where the input increases from 5 to 6, does not provide an increase in output corresponding to the textbook definition of decreasing returns to scale. On the other hand consider observation F where $E_1 = 1$ but $E_2 = 1/2$. Keeping the input level and increasing the output level results in moving along the DEA-V frontier until observation A is reached and $E_1 = E_2 = 1$. This shift along the frontier has provided an output increase for a constant level of input, i.e. increasing returns to scale is present for observation F.

The above analysis of efficiency measures for each observation are provided by solving a LP-problem for each observation. The observed input vector for observation $k0$ can be listed as

$$X_{k0} = (x_{k01} ..., x_{k0m})$$

and the output vector

$$Y_{k0} = (y_{k01}, ..., y_{k0s})$$

The input efficiency measure E_1 can then be calculated by the following LP problem:

$$(1) \ \min E_{1,k0}$$

$$s.t.$$

$$(1a) \ \sum_k \delta_k x_{kj} \leq E_{1,k0} x_{k0j}, j = 1, ...m$$

[3] E_{D1} is the E_1 efficiency measure calculated relatively to the DEA-D frontier, whereas E_{V1} is the E_1 efficiency measure calculated relatively to the DEA-V frontier.

$$(1b)\ \sum_k \delta_k y_{ki} \geq y_{k0i}, i = 1, ..., s$$

$$(1c)\ \sum_k \delta_k = 1, k = 1, ...n$$

$$(1d)\ \delta_k \geq 0$$

where the $\delta_k' s$ determine the reference technology, that is, the unit on the frontier with which $k0$ is compared. The restrictions (1a)-(1d) secure that the production technology for the n observations satisfies free disposability, convexity and variable returns to scale. (1a) expresses that the inputs of $k0$ adjusted with the efficiency measure E_1 must be equal to or larger than the reference technology. Note that all inputs are adjusted with the same factor E_1, i.e. the observed inputs are changed by the same proportion. (1b) implies that the observed outputs of observation $k0$ must be equal to or smaller than the reference technology. These inequalities can be interpreted intuitively in the following way: For each observation (x_{k0j}, y_{k0i}) it is analysed whether there exists a linear combination of observations which dominates this observation, i.e. a combination which uses the same amount or fewer inputs in order to produce the same or greater level of outputs. (1c) and (1d) provides a convex production technology with variable returns to scale. If (1c) is substituted with $\sum_k \delta_k \leq 1$ then the production technology is assumed to satisfy non- increasing returns to scale. If (1c) is removed, the production technology is assumed to satisfy constant returns to scale. Note that in order to calculate E_3 the LP-problem (1), with (1c) removed, will be sufficient to provide this, that is:

$$(2)\ \min E_{3,k0}$$

$$s.t.$$

$$(2a)\ \sum_k \delta_k x_{kj} \leq E_{3,k0} x_{k0j}, j = 1, ...m$$

$$(2b)\ \sum_k \delta_k y_{ki} \geq y_{k0i}, i = 1, ..., s$$

$$(2c)\ \delta_k \geq 0$$

The output increasing E_2 measure can be obtained by changing the restrictions (1a) and (1b) to:

$$(3a)\ \sum_k \delta_k x_{kj} \leq x_{k0j}$$

$$(3b)\ \sum_k \delta_k y_{ki} \geq \frac{1}{E_2} y_{k0i}$$

and then maximising $\frac{1}{E_2}$. The scale efficiency measures E_4 and E_5 can be derived from E_1 and E_3 resp. E_2 and E_3.

The optimal solution to (1) consists of $E^*_{1,k0}$ and $(\delta^*_1,...,\delta^*_n)$ where $(\delta^*_1,...,\delta^*_n)$ determines the efficient input-output combination with which the observation $k0$ is compared to. Thus $(\sum_k \delta^*_k X_k, \sum_k \delta^*_k Y_k)$ is the reference unit. $k0$ is efficient if $E^*_{1,k0} = 1, \delta^*_{k0} = 1, \delta^*_k = 0k \neq k0$. Otherwise $k0$ is inefficient with $E^*_{1,k0} < 1$ and $\delta^*_{k0} = 0$. E_1 to E_5 are independent of the units of measurement.

However an important problem concerning the formulation of the restrictions in (1)-(3) is that they allow for observations to be deemed falsely efficient. Such observations receive an efficiency score equal to 1 but other observations exist which use smaller levels of some inputs and achieve the same output level. Consider the following example with 3 observations (A, B and C) where each is using 2 inputs (x_1, x_2) to produce 1 output (y):

DMU	(x_1, x_2, y)
A	(2,1,1)
B	(1,1,1)
C	(1,2,1)

These observations imply an l-shaped DEA-isoquant as illustrated in figure 6.

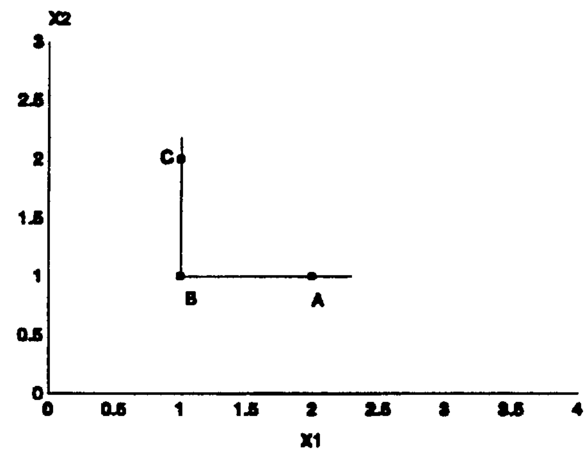

Figure 6. False efficiency evaluation situation.

All 3 observations will attain efficiency scores equal to 1, but for both A and C other observations exist that perform better (e.g. B), i.e. other observations reach the same output level and use less of either one of the inputs. The problem arises because the models (1)-(3) declare an observation efficient if it is not possible to reduce all inputs proportionately (or augment all outputs proportionately) and at the same time still be a feasible input-output combination (i.e. still belonging to the production possibility set). Thus if the inputs of observation A (from the example) are proportionately reduced this input-output combination is not possible since the output level cannot be produced from this level of inputs. This problem can though be "solved" by a slight reformulation of (2) (the same holds for (1) and (3))[4]:

$$(4)\ \min E_{3,k0} - \alpha(\sum_j s_j^+ + \sum_i s_i^-)$$

$$s.t.$$

$$(4a)\ 0 = E_{3,k0} x_{k0j} - s_j^+ - \sum_k \delta_k x_{kj}, j = 1, ...m$$

$$(4b)\ y_{k0i} = -s_i^- + \sum_k \delta_k y_{ki}, i = 1, ..., s$$

$$(4c)\ \delta_k \geq 0, \alpha, s_j^+, s_i^- \geq 0$$

s_j^+ and s_i^- are slack-variables for the j'th input and the i'th output respectively, α is an infinitesimally small constant. This model was originally proposed by Charnes, Lewin, Morey & Rousseau (1981). For some further developments see Banker et al. (1984). In this model an observation is efficient only if $E_3 = 1$ and $s_j^+ = s_i^- = 0$ for all i, j. If E_3 is less than 1 it implies that this observation is not situated on the frontier. If $E_3 = 1$ but some s_j^+ or s_i^- are different from 0 then this observation is placed on the frontier but other observations on the frontier reach the same output level using fewer inputs in some dimensions, i.e. where the slack variables are different from 0. Obviously it is also possible to obtain a solution where $E_3 < 1$ and some slack-variables are positive. If an observation has slack- variables different from 0 it implies that it is possible to reduce those inputs with the value of the slack without changing the value of E_3. But slacks will affect the overall efficiency measure $E_{3,k0} - \alpha(\sum_j s_j^+ + \sum_i s_i^-)$. As noted by Lovell (1993) the magnitude of the overall efficiency measure is dependent on the units of measurement of inputs and outputs. It should be noted that the overall efficiency measure is affected by the choice of α. Originally Farrell (1957) solved the slack problem within the Y_{CON}-technology by adding the hypothetical observations $(\infty, 0, ...), (0, \infty, 0, ...), ..., (0, ..., \infty)$ to Y_0. However, this 'solution' is only possible when the convexity of Y is assumed.

The LP problem in (4) has a dual form represented by the non- linear problem (5):

[4]Another solution could be to use non-radial efficiency measures, see e.g. Färe & Lovell (1978) and Russell (1985). See section 3.6 for more details.

$$(5)\ \max E_{3,k0} = \frac{\sum_i u_i y_{k0i}}{\sum_j v_j x_{k0j}}$$

$$s.t.$$

$$(5a)\ \frac{\sum_i u_i y_{ki}}{\sum_j v_j x_{kj}} \leq 1$$

$$(5b)\ u_i, v_j \geq 0$$

where u_i and v_j are the weights given to the i'th output resp. the j'th input. Alternatively, a non-linear minimisation problem can be defined as in (5'), which results in the same solution as in (5); see Seiford & Thrall (1990):

$$(5')\ \min E_{3,k0} = \frac{\sum_j v_j x_{k0j}}{\sum_i u_i y_{k0i}}$$

$$s.t.$$

$$(5a')\ \frac{\sum_j v_j x_{kj}}{\sum_i u_i y_{ki}} \geq 1$$

$$(5b')\ u_i, v_j \geq 0$$

(5) (and (5')) is solved by choosing the weights such that the efficiency of $k0$ is maximised. Thus the efficiency evaluation of $k0$ is made in the best possible light, see Lewin & Morey (1981). In general large output weights will be assigned to those outputs which the unit produces at relatively high levels and large input weights to those inputs which the unit use in relatively small amounts. Thus, if a unit is deemed inefficient after the best possible weights have been used then this inefficiency is difficult to neglect. On the other hand a unit can appear to be efficient by an "appropriate" choice of weights, e.g. putting large weights to inferior outputs. This has led to the development of DEA-models with weight restrictions in order to restrict the possibility for observations being efficient due to inappropriate weights, see e.g. Charnes et al. (1990) and Ali & Seiford (1993). Since $k0$ is one of the n observations $E_3, k0 = \frac{\sum_i u_i y_{k0i}}{\sum_j v_j x_{k0j}} \leq 1$. The observations for which $\frac{\sum_i u_i^* y_{ki}}{\sum_j v_j^* x_{kj}} = 1$ with the optimal weights for $k0$ form the reference set for $k0$. For each of the n problems values are provided for the input and the output weights (v_j and u_i). The non-linear problem can be formulated as an ordinary LP-problem, that is:

$$(6)\ \max E_{3,k0} = \sum_i u_i y_{k0i}$$

$$s.t.$$

$$(6a)\ \sum_i u_i y_{ki} \leq \sum_j v_j x_{kj}$$

$$(6b)\ \sum_j v_j x_{k0j} = 1$$

$$(6c)\ u_i, v_j \geq 0$$

Notice that (6) is the dual of (4)

(4)-(6) represent the original DEA formulation proposed by Charnes et al. (1978). An interesting property of the efficiency concept in this model is the clear correspondence with productivity defined as

$$Productivity = \frac{weighted\ sum\ of\ outputs}{weighted\ sum\ of\ inputs}$$

The only difference is that in DEA the input and output weights are determined as a part of the solution of an optimization problem whereas the productivity definition above requires fixed weights a priori. Thus there is a clear relationship between the efficiency measure in the DEA-models and the more traditional performance measures. The choice between (4) and (6) (i.e. the choice between the primal and dual form) cannot be made clear. However the number of restrictions in (4) are smaller than in (6) since the sum of the inputs and output categories are smaller than the number of observations, thus (4) will be computationally be faster to solve than (6), see Boussofiane et al. (1991).

3.4.3 Free Disposable Hull model.

Measures of efficiency in the FDH-model (see e.g. Deprins et al. (1984), Thiry & Tulkens (1989) and Pestieau & Tulkens (1993)) are computed as follows.

Consider the n sets $D(x_k, y_k, Y_0), k = 1, ..., n$. These sets contain information about the observations which dominate the k'th observation (the evaluated unit). As mentioned in section 3.3.1, the k'th observation is undominated (efficient) if and only if $D(x_k, y_k, Y_0)$ is a singleton. Otherwise it is dominated (inefficient) by the other observations included in $D(x_k, y_k, Y_0)$.

Let the observations in $D(x_k, y_k, Y_0)$ be indexed by $h = 1, ..., H$. Consider a given observation $k0$ in Y_0. This observation is the observation under evaluation. The aim is to measure the input (and output) efficiency of $k0$ based on the radial Farrell index defined above, i.e. E_1 (and E_2). One way to do this is to build a procedure directly upon the dominance set $D(x_k, y_k, Y_0)$. This was done in the one-input -multiple-output case by Deprins, Simar & Tulkens (1984) and in the multiple-input-one-output case by Thiry & Tulkens (1988). The following stepwise procedure extends this approach to the multiple-input-multiple-output case and result in the input efficiency measure E_1:

Step 1. Determine $D(x_{k0}, y_{k0}, Y_0)$.

Step 2. For each of the H elements in $D(x_{k0}, y_{k0}, Y_0)$ calculate the input ratios $\theta_{hj} = x_{hj}/x_{k0j}, h = 1, ..., H, j = 1, ..., m$. θ_{hj} is the h'th unit's amount of input j in relation to the amount of input j used by the evaluated unit $k0$, hence $0 < \theta_{hj} \leq 1$. θ_{hj} indicates how much the evaluated unit $k0$ can reduce the amount of input j in order to be as good as unit h with respect to input j.

Step 3. For all h calculate $\theta_h = \max\{\theta_{hj}\}$. Notice that this implies that $k0$ is compared with each of the H units with respect to the most favorable input level. If this were not the case then $\theta_h x_{k0j}$ would not be included in Y_{FDH}.

Step 4. Calculate $E_{1,k0} = \min \theta_h$ in order to move $(E_{1,k0}x_{k0}, y_{k0})$ to the frontier of Y_{FDH}.

A similar procedure can be constructed for an output efficiency measure $E_{2,k0}$.

In the following this procedure will be illustrated using the previous example with 6 observations. In the case of the example only the observations E, D and F are dominated, the rest are not in the position where other observations use a lower level of input and produce the same amount or more output. Observation E is dominated by A and B, observation D is dominated by C and observation F is dominated by A. Measurement of efficiency in the single input-single output case is simple. If the efficiency is measured in terms of input two candidates for observation E appear: 3/5 and 4/5. Thus the input efficiency measure becomes equal to 3/5. In terms of output, the candidates are 2/2 and 2/4, implying an output efficiency measure equal to 2/4 for observation E. Note that only observation E will have efficiency measures of less than 1 for both types of measure. Observation F will receive an input efficiency measure equal to 1, while observation D will receive an output efficiency measure equal to 1.

The efficiency evaluation in terms of inputs of a given observation $k0$ in the FDH-model can be formulated as a mixed integer problem, e.g. shown in Tulkens (1990):

$$(7)\ \min E_{1,k0}$$

$$s.t.$$

$$(7a) \sum_k \delta_k x_{kj} \leq E_{1,k0} x_{k0j}, j = 1, ...m$$

$$(7b)\ \sum_k \delta_k y_{ki} \geq y_{k0i}, i = 1, ..., s$$

$$(7c)\ \sum_k \delta_k = 1, k = 1, ...n$$

$$(7d)\ \delta_k \geq 0, \delta_k \in \{0, 1\}$$

Note that this model also has the problem with slacks in inputs and outputs even for observations with $E_1 = 1$. However, a slack-augmented model can also be formulated with a FDH model. An output efficiency measure E_2 can be obtained using a slightly different version of the mixed integer programming expression in (7). (7) and (1) are similar, where the only difference appears to be the restrictions on the intensity parameter δ_k. Due to the restrictions on δ_k in the FDH-model only one observation in the left-hand side of the inequalities in (7) can have a $\delta_k \neq 0$, where this value will, by definition be equal to 1. Thus the efficiency evaluation of a given observation $k0$ can be compared with only one observation and not with some linear combination of observations. If the optimal solution to (7) turns out to be such that $\delta_k = 0$ for all $k = 1, ..., n$ with $k \neq k0$ and $\delta_{k0} = 1$ then $k0$ is undominated and will have $E_{1,k0} = 1$. The efficiency measures in the FDH-model will be greater than or equal to the efficiency measures in the DEA- model because it is more difficult to find input-output combinations that dominate a given observation since this model does not allow for linear combinations of observations to form the dominant observation. The following holds in particular: $E_{1,DEA-C} \leq E_{1,DEA-D} \leq E_{1,DEA-V} \leq E_{1,FDH}$. The same holds for E_2.

In contrast to the DEA-model it is not possible to derive immediately the measures of scale efficiency in the FDH set-up, that is E_3, E_4 and E_5. The E_3 measure is constructed with specific reference to the DEA-model and E_4 (and E_5) is a derived measure where E_3 is related to the pure technical efficiency measure E_1 (or E_2) from the DEA-V model. Of course it is possible to compute E_4 (or E_5) by the ratio $\frac{E_3}{E_1}$ (or $\frac{E_3}{E_2}$), with E_3 from the DEA-model and E_1 from the FDH-model, but such a measure will not have any clear interpretation. In a DEA set-up this ratio has a clear meaning in the sense that it measures inefficiency caused by scale (either by being too large or too small). If E_4' is constructed as the ratio $\frac{E_{3,DEA-C}}{E_{1,FDH}}$ it is not possible to say that E_4' expresses only scale inefficiency since it includes both scale inefficiency and measured inefficiency caused by introducing convexity (by moving from the FDH model to the DEA-V model). Much more information can be provided by constructing E_4' and E_5' as follows:

$$E_4' = \frac{E_{3,DEA}}{E_{1,FDH}} = \frac{E_{3,DEA}}{E_{1,DEAV}} \frac{E_{1,DEAV}}{E_{1,FDH}}$$

$$E_5' = \frac{E_{3,DEA}}{E_{2,FDH}} = \frac{E_{3,DEA}}{E_{2,DEAV}} \frac{E_{2,DEAV}}{E_{2,FDH}}$$

The first ratio in the last expression for E_4' measures the scale efficiency, whereas the second ratio measures the change in efficiency by imposing convexity. The same interpretation can be given to the expression for E_5'. Thus, these ratios indicate whether differences in DEA-C and FDH efficiency scores are caused by scale inefficiency or due to imposing convexity in the DEA-model.

3.4.4 Measuring the efficiency of the undominated units

Though it is possible to measure the degree of efficiency of each observation in the data set the problem still remains that a large number of observations, in fact, will be efficient (or undominated) as mentioned in section 3.3. Hence, it would be preferable if the efficient (or undominated) units could be ranked by some measure which can be interpreted along the same lines as the original efficiency measure.

Andersen & Petersen (1989) propose a procedure to solve this problem within a DEA framework. The basic idea is to construct a measure for each efficient observation which determines by how much inputs can be increased proportionately, provided that the observation stays efficient relative to the data set. Consider the following reformulation of the DEA model where $Y = Y_{CON}$:

(8) $\min \theta$

s.t. $\sum_k \delta_k x_{kj} \leq \theta x_{k0j}, \sum_k \delta_k y_{ki} \geq y_{k0i}, \delta_k \geq 0,$
$j = \{1, ..., m\}, i = \{1, ..., s\}, k = \{1, ..., n\} \setminus \{k0\}.$

The only difference between the original DEA model and this reformulation is the exclusion of the $k0$'th observation from the envelopment of the frontier as illustrated by figure 7 (using the example from figure 6). Notice that inefficient observations remain inefficient and obtain

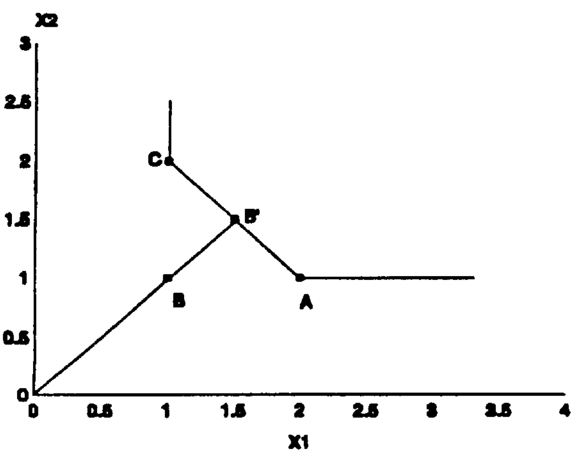

Figure 7. Measuring B against the hypothetical frontier.

the same degree of efficiency since their exclusion does not change the frontier. However, observations with $\theta = 1$ in the original DEA-model will receive a degree of efficiency larger than or equal to 1 in the respecified model. Those observations which maintain their efficiency degree equal to 1 are those which in the original model have slacks in one or more input dimensions.

However, Andersen & Petersen notice that if the assumed technology is changed to Y_{DEC} or

Y_{VAR} situations may occur where (8) has no solution. Consider figure 8 where $Y = Y_{DEC}$. The observation C represents a situation where (8) has no solution since the removal of C from the frontier implies that the input level of C has no frontier reference.

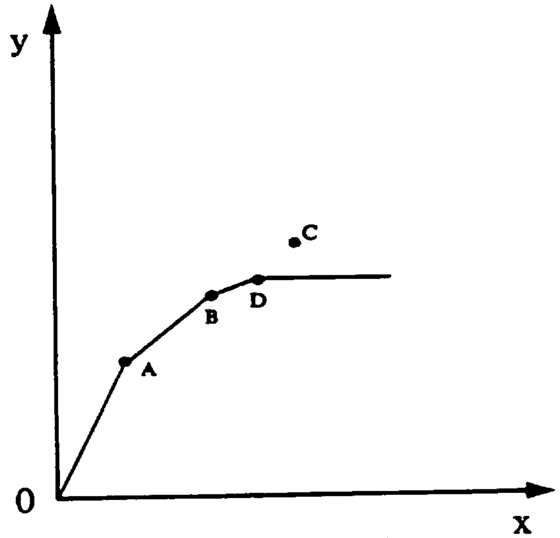

Figure 8. No solution to (8).

Y_{FDH} is characterised by additional restrictions on the δ's compared to both Y_{DEC} and Y_{VAR}. Therefore there is an even larger probability of finding observations within Y_{FDH} which have no solution to the programming problem. For the free disposal hull technology Y_{FDH} the following n reformulated mixed integer programming-problems are constructed:

(9) $\min \theta$

s.t. $\sum_k \delta_k x_{kj} \leq \theta x_{k0j}, \sum_k \delta_k y_{ki} \geq y_{k0i}, \sum_k \delta_k = 1, \delta_k \in \{0,1\}$,
$j = \{1,...,m\}, i = \{1,...,s\}, k = \{1,...,n\} \setminus \{k0\}$.

It is easy to see that (9) has no solution when the evaluated observation is characterised by a maximal output amount in relation to all other units in Y_0 for at least one output category since it would imply that at least one of the i restrictions $\sum_k \delta_k y_{ki} \geq y_{k0i}$ are violated.

3.4.5 The relationship between efficiency measurement methods and estimation of production relations.

Implicit in the non-parametric efficiency evaluations of micro units as provided by DEA and FDH is a production relation between some inputs and some outputs. Thus the methods of efficiency measurement are, in fact, closely connected to the estimation of production relations. Therefore, the efficiency evaluation models can be seen as providing a basis for estimation of production relations at a micro level while allowing for inefficiency. This contradicts the traditional procedure for estimating micro production relations, which seeks to provide an "average" relation with less and more efficient points placed both below and above the estimated relation, such that it is assumed that the units, in general, produce efficiently. Moreover this standard method to obtain a production relation follows a parametric approach where the functional form for the relationship between inputs and outputs has to be specified a priori. This specification can be correct but will most likely be wrong. In contrast the production relations obtained in DEA have a non-parametric basis per se. This criticism also concerns the parametric efficiency measurement methods where the production relation is estimated before the computation of efficiency measures, as described briefly in chapter 2.

In this section the relation between the non-parametric efficiency measurement model, DEA, and the estimation of production relation is analysed. It is assumed that the n observations use m inputs in the production of a single output y where a typical observation, k, is (x_k, y_k) with $x_k = (x_{k1}, ..., x_{km})$ and $k = 1, ..., n$

Consider the following production function $f(x)$ originally proposed by Afriat (1972):

$$(10)\ f(x) = \max[\sum_k y_k \delta_k \mid \sum_k x_k \delta_k \leq x, \sum_k \delta_k = 1, \delta_k \geq 0]$$

This function is defined in the case of a single output being produced by m inputs and can be shown to be non-decreasing and concave provided non-negative input and output values.

$f(x)$ envelops the data not more distantly than any other non- decreasing and concave function. This production function is essentially similar to the boundary of the production possibility set from a Data Envelopment Analysis efficiency measurement model with variable returns to scale since we previously showed that the reference technology (the production possibility set):

$$(11)\ Y_{DEA-V} = \{(x, y) \mid y \leq \sum_k y_k \delta_k, x \geq \sum_k x_k \delta_k, \sum_k \delta_k = 1, \delta_k \geq 0\}$$

where the restrictions on inputs and outputs correspond to free disposability of inputs and outputs and convexity of the production possibility set. Thus the frontier to this production possibility set can be derived by using the concept of a production function as the maximum output for given inputs and it appears as:

$$(11a)\ f_T = \max[y_k \delta_k \mid x \geq \sum_k x_k \delta_k, \sum_k \delta_k = 1, \delta_k \geq 0]$$

that is (10) = (11a). Afriat was concerned with whether the k observations were consistent with a given production function, such as the one expressed in (10). Consistency implied that the data are situated on the production function, i.e. efficiency is precluded. If the production function is given by (10) the consistency of the data reduces to seeing whether the observations are placed on the DEA-V frontier or not. The condition for the n observations to be consistent with (10) is

$$(12) \quad \sum_k x_k\delta_k \leq x'_k \Rightarrow \sum_k y_k\delta_k \leq y'_k$$

where $\sum_k \delta_k = 1, \delta_k \geq 0$. (12) is satisfied if $y_k = f(x_k)$. The intuitive meaning of (12) can be explained as follows. Consider the case where $\sum_k x_k\delta_k \leq x'_k$ and $\sum_k y_k\delta_k \geq y'_k$. This is clearly in contrast to efficient production since an input combination $\sum_k x_k\delta_k$ at a lower level than x'_k produces a higher level of output $\sum_k y_k\delta_k$ compared with the output level y'_k. Such a situation indicates that $y_k < f(x_k)$.

As noted by Afriat it is likely that with samples of empirical observations the test considered will fail due to inefficiencies in production.

Another procedure to show the correspondence between DEA and the production function estimation is to consider the dual form of the DEA-problem, i.e. equation (5'). With a single output the DEA minimisation problem can be formulated as the following LP-problem:

$$(13) \quad \min \theta'_{k0} = \sum_j \beta_j x_{k0j}$$

$$s.t.$$

$$\sum_j \beta_j x_{kj} \geq y_k$$

$$\beta_j \geq 0, k = 1,...,n, i = 1,...s, j = 1,...,m$$

The optimal $\beta_j^* = \beta_j^*(k0)$ determines the efficiency of the k0'th unit:

Define $y_{k0}^* = \sum_j \beta_j^*(k0)x_{k0j}$: If $y_{k0} = \sum_j \beta_j^*(k0)x_{k0j}$ and $s_{k0} = y_{k0}^* - y_{k0} = 0$ then $k0$ is efficient, otherwise it is inefficient

The formulation in (13) has a very clear correspondence with the production function concept. Thus $\sum_j \beta_j^*(k0)x_{k0j}$ can be interpreted as indicating the maximum level of output in the specific region of the production possibility set to which $k0$ belongs. But this is exactly what a production function measures. For each of the n observations a parameter vector $\beta_j^*(k)$ will be obtained which can be interpreted as parameters defining an appropriate production function. Running (13) for all observations can therefore be used to reveal the surface of the production possibility set, that is the best-practice production frontier. Notice that this production function is not a global function between inputs and outputs in parameter constancy, thus the parameters β_j vary for the different observations. The relation derived between inputs and output is the union of n local production functions since the parameters β are determined for each observation. However, there will not be n different solutions of $\beta_j^*(k0)$ since, for some of the solutions, the only difference is a constant. In this way the

DEA-model's production relation is very similar to the so-called random parameter model from econometrics (see Johnston (1984)). The production relation that is implicit in the standard DEA-model can, therefore, be characterised as piecewise linear, where each piece corresponds to a different β vectors[5].

The relation between DEA efficiency evaluation and production function estimation can also be illustrated by looking at the average parametric[6] production function

$$(14)\ y_k = \sum_j \beta_j x_{kj} + e_k$$

where e_k is a random error term that can take positive as well as negative values. This relation can e.g. be estimated by OLS with the resulting optimal parameter vector $\beta^* = (\beta_1^*, ..., \beta_m^*)$. Notice that the β^* are chosen such that the sum of squared residuals is minimised. Notice that the estimation of the optimal parameters in (14) is based on all observations, whereas the optimal parameters obtained by (13) is based only on a subset of observations (the efficient observations), implying relatively more outlier influence in the DEA-model. In order to obtain a production relation which corresponds with the theoretical meaning of a production function (that is the maximum output given the value of the input), the following restriction has to be imposed on the error term $e_k : e_k \leq 0$. In this case the yk cannot exceed the estimated production level $y_k^* = \sum_j \beta_j^* x_{kj}$. This function could be estimated by OLS minimising $\sum_k e_k^2$, but this implies that extreme observations will have a great influence on the resulting estimates (see Timmer (1971)). In addition, the restriction on the ek implies that the OLS estimator will result in inconsistent estimates of β.

Instead Timmer (1971) proposed to estimate the parameters by minimising the sum of the absolute value of the errors, $\sum_k | e_k |$, which, since the e_k are non-positive, leads to the following LP- problem:

$$(15)\ \min \theta_0 = \sum_j \beta_j \bar{x}_j$$

$$s.t.$$

$$\sum_j \beta_j x_{kj} \geq y_k$$

$$\beta_j \geq 0$$

[5]See however Banker & Maindiratta (1986) for a DEA-model which implies a loglinear production relation. See also Färe, Grosskopf & Njinkeu (1988) for a specification of the production technology as a piecewise technology covering both linear and non-linear relations which includes the standard DEA piecewise linear production relation and the piecewise loglinear relation as special cases.

[6]It is parametric in the sense that a global function is assumed a priori, which in this particular case is specified as a linear function.

where $\bar{x}_j = \frac{1}{n}\sum_k x_{kj}$, the average level of input j over the whole sample. The estimator of β based on minimising the sum of the absolute value of the errors is known as the LAV estimator (Least Absolute Value) and can be shown to give less importance to extreme observations compared with the OLS estimator based on the squared residuals (see Rousseeuw & Leroy (1987)). In contrast to the LP-problem in (13) which represents n LP's, the model in (15) is a single LP and therefore only one parameter vector $\bar{\beta}$. It should be noted that (15) also represents an optimal production relation rather than an average tendency between inputs and outputs due the restrictions on the LP-problem. This LP problem can also be obtained by summing the n objective functions from (13) and dividing each component by n, the number of observations. Thereby the relationship between the average production function and the DEA production relation is shown. An interpretation of the LP problem (10) with minimisation of the weighted sum of the average inputs is that it represents in some sense the representative observation within the empirical sample. Thus, if inputs for each observation are close to the calculated average level then the difference between the $\bar{\beta}^*$ $\beta^*(k0)$ will be low implying that the efficiency rankings will be close.

Several tests can be applied in order to examine the consequences of estimating an average production function rather than an optimal production function as formulated in (13) or (15); see e.g. Sengupta (1987) and Sengupta (1989). An example of such a procedure is the following: consider the efficiency ranking obtained from running the n LP's. Select a subsample among the observations representing the most efficient. Estimate a production function of the following form

$$(16)\ y_k = \sum_j \beta_j x_{kj} + \alpha D_k + c$$

where $D_k = 1$ if the k'th observations is efficient and $D_k = 0$ otherwise. Then enlarge the data set with less efficient observations in an order such that the most efficient observations are included first and run a regression for each enlarged sample of the form as in (16). By including more observations with smaller efficiency levels the estimated production relation will move from an optimal production function towards an average production function. A general property will be that the R^2 will decline. The extent of this decline in explanatory power can indicate the importance of using an optimal production function rather than an average production function. The parameter estimate α for the dummy variable will be significant if the observations can be grouped into disjoint sets of both efficient and inefficient observations. Notice that this procedure can be characterised as establishing an interface between non-parametric DEA models and parametric production function estimation.

3.5 Applications of non-parametric efficiency measurement methods.

The analysis of the possible applications of non-parametric efficiency measurement methods involves two issues: (1) the information obtained from a DEA-analysis, (2) how this information can be used. In the following both facets will be examined[7]

The main output from a DEA analysis is the efficiency score for each of the observations, as described in sec. 3.4. This efficiency score indicates the possible reduction in inputs or the possible increase in outputs. The efficiency scores provide a criteria by which a given set of observations can be ranked, from high to low.

In addition to these efficiency scores the DEA-model provides other sources of important information. If (4) is considered, the optimal solution consists of

$$(E^*_{3,k0}, \delta^*_1, ..., \delta^*_n, s^{*+}_1, ..., s^{*+}_m, s^{*-}_1, ..., s^{*-}_s)$$

, where the observations with $\delta^*_k \neq 0$ are those efficient observations which form the dominating group of observations relative to $k0$. These observations are said to form a peer group with respect to $k0$. Thus, by multiplying each of these observations' inputs and outputs with the corresponding δ^*_k an input-output combination situated on the efficient frontier, i.e. this combination uses less inputs to produce greater output levels than the observation $k0$. Moreover, for each observation with $\delta^*_k \neq 0$ its contribution to the constructed reference input-output combination can be computed, where the largest contributor is the most important observation in the efficiency evaluation of $k0$. It should be noted that the observations which form the reference combination for $k0$ are similar to $k0$ in terms of input-output orientation. Thus the solution of the DEA-model for a given observation $k0$ provides information concerning how the efficient observations operate compared with $k0$. This is of particular importance with respect to the observations which are the main contributors to the reference combination. The DEA-solution can, therefore, give suggestions how inefficient observations should change production patterns in order to become more efficient.

Obviously changing the inputs and outputs for $k0$ in the following way will render it efficient:

$$(17) \qquad \Delta x_{k0j} = (1 - E^*_{3,k0})x_{k0j} + s^{*+}_j$$
$$\Delta y_{k0i} = s^{*-}_i$$

which can be derived directly from the solution to (4); see Charnes & Cooper (1985b) and Ali & Seiford (1993). These changes in inputs and outputs will give exactly the input-output combination which serves as the reference point in the efficiency evaluation. However, it should be noted that (17) is not the only input-output adjustment available for a given

[7]The arguments below will concentrate on the DEA-model.

observation $k0$. In an infinite range of possible adjustments exist, since the frontier contains an infinite number of points.

If, instead, the dual form to (4) is considered, i.e is (6), this solution also provides other types of information than the efficiency scores alone. The optimal weights u_i^* and v_j^* can be applied in the same way as δ_k in order to identify the group of efficient observations which form a peer group for $k0$. These weights will make the peer units efficient where

$$\sum_i u_i^* y_{hi} = \sum_j v_j^* x_{hj}, h = 1, ..., H$$

h is an index number for the peer units. The group of peer units can, therefore, be found by examining the n restrictions to (6) with the optimal weights for $k0$ included. Those of the n restrictions which are satisfied the equality indicate the group of peer units for $k0$; see Boussofiane et al. (1991) and Sengupta (1989). The identified peer units in (6) are identical to those in (4), provided that the solutions are unique. However, as noted previously the weights u_i and v_j can, in the present form of (6), be chosen without restrictions. Thus the optimal weights can be such that they are inappropriate, e.g. by giving a high weight to an output category which is not considered to be the most important product. One way to approach this problem is to take into account that (6) does not only provide an efficiency score for $k0$, but can also provide efficiency scores for the other observations through the n restrictions. Thus, information about how the other observations will be rated according to the optimal weights for $k0$ is available.

If (6) is solved for each of the n observations and calculate how it will be rated by the other $n - 1$ observations weights, an impression of the robustness of the obtained efficiency results is gained. If an observation is efficient with its own optimal weights, but turns out to be inefficient with the weights of any other observation then this would be a clear indication that the efficiency of this observation is more a result of the choice of weights than that of an efficient way to organise the production. Another procedure that can be used to detect such inappropriately chosen weights could be to solve (6) using the average inputs and outputs of the observations. This will provide a set of weights $(\bar{u}_i, \bar{v}_j)$ to be compared with the set of n optimal (u_i^*, v_j^*). If, for some observation $k0$, (u_i^*, v_j^*) are distant from $(\bar{u}_i, \bar{v}_j)$ then the weights for this observation to are inappropriately chosen.

A possible solution to the problem of inappropriate weight- structure for some observations has been to introduce restrictions on the possible values the weights can take, e.g. that the weight of one output is restricted to take values below that of another output; see Dyson & Thanassoulis (1988), Charnes et al. (1990) and Ali & Seiford (1993). Indeed such an approach secures that the weights will take values within the specified ranges, but imposing weights restrictions can only be based on ad hoc considerations and not some general principles and therefore requires sector-specific knowledge. Thus the introduction of weights restrictions remove the desirable property of the DEA-model of being assumption free.

Another source of information directly obtainable from the solution to (6) is the products $u_i^* y_{ki0}$ and $v_j^* x_{kj0}$, the so-called virtual outputs and inputs, see Charnes & Cooper (1985b). Each of the products $u_i^* y_{ki0}$ measures the i'th virtual outputs contribution to the efficiency score, since the sum over all virtual outputs is equal to the overall efficiency score for $k0$. The sum of the virtual inputs for $k0$ is restricted to equal 1. Thus the value of the single virtual input shows the proportion for this virtual input with respect to the sum of all virtual inputs. If the size of the virtual inputs and outputs are examined for the efficient observations information about which inputs and outputs these observations give importance to in the efficiency evaluation is obtained. It is possible in this way to point out to the inefficient observations how their production should be organised in order to become more efficient. The efficient observations can, indeed, give differing importance to the inputs and outputs, but this could be used to suggest to the efficient observations how they can change their production and become even more productive. However, this last remark is made with the assumption that the weights are appropriately chosen.

An interesting interpretation of the weights is to consider them as shadow prices. If market prices for inputs and outputs are available these can be compared with the shadow prices in order to examine the consistency of the two sets. In any case the weights contain information about how the observations should value their inputs and outputs in order to achieve the highest possible efficiency rating. Obviously this valuation does not need to correspond exactly to the valuation of the inputs and outputs that the unit in fact has because of the differing goals of the units or due to market conditions.

In addition, the weights can be used to calculate marginal rates of transformation of outputs, marginal rates of technical substitution of inputs and marginal productivities, see e.g. Sengupta (1989), Charnes & Cooper (1985) and Banker et al. (1986). For any pair of outputs y_{k0i} and y_{k0g} (with $g = 1, ..., s,\ i = 1, ..., s$ and $g \neq i$) it is possible to measure the marginal rate of transformation of output y_i for output y_g by the ratio $\frac{u_i^*}{u_h^*}$. This ratio shows by how much the output y_g could be increased if the output y_i is reduced by 1 unit. Similar for any pair of inputs x_{k0j} and x_{k0l} (with $j = 1, ..., m, l = 1, ..., m$ and $j \neq l$) the marginal rate of technical substitution of input x_j for input x_l can be measured by the ratio $\frac{v_j^*}{v_l^*}$. This ratio indicates by how much the input x_{k0l} has to be increased in order to keep the same efficiency level if x_{k0j} is decreased by 1 unit. The marginal productivity of input x_{k0j} with respect to output y_{k0i} can be measured by the ratio $\frac{v_j^*}{u_i^*}$.

3.5.1 Managerial efficiency versus programme efficiency.

A problem often encountered in empirical analyses of organisations is a situation where two or more programmes[1] have to be compared in order to decide which one is the "best" and

[1]Programmes could be: public vs. private hospitals, schools with specific steps towards disadvantaged pupils vs. schools without such steps, libraries with a bus service vs. libraries without this.

the organisations have employed different programs. Comparisons of programs under these circumstances can lead to erroneous inferences since the program inefficiency is mixed up with managerial inefficiency. Thus a given program can be deemed inefficient not because the program is bad but because the organisations which supply the program are managed in an inefficient way. Charnes, Cooper & Rhodes (1981) introduced a procedure that can be used to decompose the inefficiency of each observation into program and managerial inefficiency within a DEA model and thus provides a basis for examining which programs are most efficient. Charnes et al. (1981) applied this procedure in the comparison of two school programs in the US. For a similar procedure see Grosskopf & Valdmanis (1987), who test the efficiency differences between public and private hospitals in the US. Another application is Magnussen (1992), who compares the efficiency of small hospitals with that of big hospitals in Norway.

Consider that the n observations can select among L different programs (indexed as $l = 1, ..., L$) and let n_l be the number of observations which use the program l such that $n_1 + n_2 + ... + n_L = n$. For each of L subsets the DEA model can be runned (e.g. (1)) and obtain efficiency measures for the observations in each subset. A given observation is, in this case, evaluated with respect to the other observations which use the same program and therefore L separate frontiers are constructed. These efficiency measures cover the managerial part of efficiency since the effect from using different programs has been isolated by only comparing observations within the same program. The next step adjusts the inputs of each observation such that each observation becomes efficient with respect to its own frontier, i.e. any managerial inefficiency is removed. These adjusted observations are now pooled in one set instead of being L subsets. The DEA-model is rerunned for each of the n observations. Because the managerial inefficiency was removed from the observations the only remaining inefficiency for a given observation is due to program inefficiency. This efficiency measure, calculated in this way, reflects a measure of program efficiency. An alternative method of obtaining measures of program efficiency that do not rely on adjusting the inputs to the separate frontiers is available. This method also evaluates each observation with respect to its separate program frontier. Then the non-adjusted observations are pooled and an efficiency measure is calculated for each observation with respect to the pooled frontier. The ratio between the pooled efficiency measure and the separate efficiency measure provides a measure of program efficiency for each observation.

This procedure is illustrated in figure 9 in the two inputs (x_1, x_2) one output y case case, where it is assumed that all units produce the same level of the output (however this assumption is only necessary because the procedure is illustrated in two dimensions).

The line *ss* represents the frontier for one of the L programs and the line *pp* represents the pooled frontier for all observations. The observation A belongs to the program with the frontier *ss*. Thus A is inefficient with respect to the best practice of the program, where the measure of managerial efficiency is equal to $\frac{OA'}{OA}$. Furthermore if A is compared with the pooled best-practice frontier A will also be judged inefficient, since this frontier represents

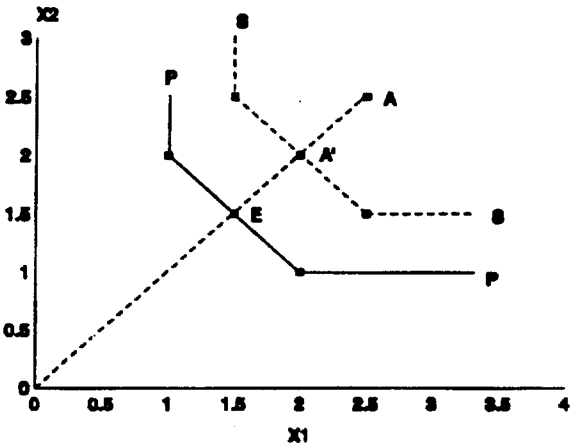

Figure 9. Program efficiency and Managerial efficiency.

the best performances over the whole sample. The measure of "pooled" efficiency becomes identical to $\frac{OE}{OA}$, where $\frac{OE}{OA} \leq \frac{OA'}{OA}$ (equality holds if the best-practice frontier of the individual program corresponds with the pooled frontier). The measure of program efficiency for A can be found as $\frac{OE}{OA'} = \frac{(OE/OA)}{(OA'/OA)}$ and thus equal to the ratio of the separate and pooled efficiency measure. The interpretation with respect to the obtained measure of program efficiency for observation A is that A uses an inferior program[9], since a part of the inefficiency is caused by the choice of program ($\frac{OE}{OA'} < 1$) and other programs are obviously at hand that perform better. In figure 9 the analysis is very simple since all observations in the program, represented by ss, are inefficient due to the program, implying that this program cannot be justified given the present information level. In general the results will be more dispersed such that for some observations the program will be efficient (that is some segments of the separate program frontier will correspond to the pooled frontier) and, for other observations, the program will be inefficient. Thus the information DEA provides can not normally state that program A is more efficient than program B, but that in some regions program A is more efficient than program B and in other regions the opposite holds. The DEA results provide more detailed and disaggregated information than usual concerning the performance of one program relative to another (corresponding to the micro orientated origin of DEA).

A slightly different version of this procedure can be constructed in order to compare hospitals in two countries (e.g. Danish hospitals vs. Swedish hospitals). The change of the procedure described above procedure involves only the definition of a program. Instead of a program being organisational specific, it is now country specific. For each hospital efficiency

[9]At least in terms of technical efficiency.

is measured according to: (1) a country specific frontier, (2) a pooled frontier (where observations from both Denmark and Sweden can form the frontier). The ratio of these two efficiency measures provides a measure of "country" efficiency for each observation, similar to the previous notion of program efficiency. It should however be noted that the country efficiency measures have to be interpreted with care. In the previous case where the program, for instance, was a special treatment for some customers and this program was inefficient relative to some other program for this observation, then a clear suggestion would be to let the observation adopt the superior program. Obviously, such a suggestion cannot be given in the case of observations with country inefficiency. What the information can be used for is to indicate whether one country's hospitals are more efficient than anothers and for which segments of the best-practice frontier this is the case. If this is the case then attention should be directed towards analyses of how the other country's hospitals are organised. To my knowledge no such application of the Charnes et al. (1981) procedure have been attempted. There have been intercountry efficiency comparisons (see e.g. Perelman & Pestieau (1988), Gathon (1989) and Deprins & Simar (1989)), but they do not decompose the inefficiency into managerial and country inefficiency.

3.5.2 DEA results and the decision of the allocation of resources.

The previous arguments have been concerned primarily with the information supplied by DEA models. Now, I will consider briefly how this information might be used.

The most obvious use of DEA results is to view them as tools for increasing the information level of the public authorities with respect to the performance of a group of similar public organisations, where the public authority has been set up in order to control and decide on the allocation of resources to the group of organisations. As described above, a number of interesting aspects of the performance of organisations can be derived from the DEA results. Moreover, this kind of information is, indeed, of importance for public authorities.

However, the DEA model suggests that other and more direct uses of the results can be applied in relation to the organisations under evaluation. The DEA results provide information about the extent to which the included organisations can reduce their input use (or increase the output produced), indicating that the DEA model can be used by the public authority for allocating resources between the organisations and to control the organisations such that the goals of the public authority are reached. These two related applications will be analysed below.

Before turning to these issues, I will briefly consider the possible consequences of using DEA for resource allocation decisions (either in terms of the general resource allocation or in terms of some award system implemented for purposes of controlling the organisations). If the organisations know that the resources provided will depend, to some extent, on the DEA evaluation, then it is expected that the organisations take this fact into account when they plan production. The crucial issue is whether this situation is incentive compatible, that is, whether it is consistent with the goals of the public authority. The organisations to be evaluated by DEA will focus on outputs and inputs which result in the best possible evaluation, and therefore expand inputs and outputs in directions which secure this. Problems occur if the specification of inputs and outputs included in the DEA-model are incomplete according to the actual production structure and, in particular, to the goals of the public authority; see Premchand (1993). As an example, consider the case of the evaluation of a group of hospitals where output is specified as the number of operations. In order to obtain a high DEA score the best procedure the organisations can follow will be to achieve the highest level of operations, resulting in unnecessary operations or operations which are of low quality. Of course this example assumes an extreme specification of the outputs of hospitals, but more realistic specifications could also imply problems if there are inconsistencies with respect to the goals of the public authority. Therefore, the DEA results have to be used with care regarding resource allocation in order to prevent perverse production behaviour.

3.5.2.1. DEA and resource allocation decisions.

If the possible use of DEA-results with respect to resource allocation is examined then this provides quite clear indications of how much the inputs could be reduced and still achieve the same production level. Thus if an observation has received an input efficiency score equal to 0.85, then it implies that this observation's inputs could apparently be reduced by 15 per cent. It is possible to carry out this calculation can we carry out for all observations and thus obtain a measure of how much the resources could be reduced in the sector overall. The DEA model provides a method by which the reductions in resources of a group of similar organisations can be decided. This is a desirable property in these times where the resources to the public sector are scarce and need to be prioritised.

Unfortunately this property of the DEA-model is open to various kinds of criticisms; see Boussofiane (1991). First of all the efficiency results obtained should be possible to reproduce with other input and output specifications. If not, the above mentioned link between DEA and resource allocation is invalid, simply because the allocation will depend on the specification chosen. It is, indeed, very likely that the efficiency results obtained in one set-up are not found in another.

In addition the computed inefficiencies could be the result of the scale at which these observations operate because the DEA evaluation is carried out with the assumption of constant returns to scale. If it is not possible to change these observations such that they operate at the optimal scale then these observations will be deemed inefficient. Otherwise the only possibility for removing the inefficiency will be to eliminate these observations. This possibility can, in practice, be quite limited due to various reasons, e.g. the need of keeping these observations in spite of their inefficient production organisation or political opinions. This last argument brings me to another criticism of the use DEA in resource allocation decisions. DEA assumes that it is possible to reduce the inputs by, (1 - the efficiency score), but this might not be possible due to either technical limitations on the transferral of resources from a specific unit to another or political restrictions.

Moreover, the DEA results are static in the sense that they measure the efficiency of a given organisation relative to the other organisations. Thus it is known how this organisation is evaluated at present with its given input and output levels. What is not known is what happens when the input levels actually is changed of this organisation. The organisation might turn out to be even more inefficient if changes in the current inputs are attempted. In any case since the evaluation is relative, any changes in the input (or output) levels imply that the DEA evaluation has to be carried out again with the changed data for the complete sample of observations. Because of these problems in relation to the use of DEA results for the allocation of resources a more practical way of applying the efficiency results would be to interpret them as indications of the possible presence of inefficiencies and then carry out more detailed analyses of these organisations in order to verify the extent of inefficiency.

3.5.2.2. DEA as a control instrument.

Another aspect regarding resource allocation is the extent to which DEA evaluations can be used to set up a payment scheme in order to award those organisations which have showed a "good" performance, when public authorities attempt to control its organisations; see e.g. Guesnerie & Laffont (1984) and Jackson (1983). These payment schemes should not be viewed as part of the original budget, but as some additional resources given to those organisations with a better performance.

In the analysis below it will be assumed that there is a positive relationship between the DEA efficiency score and the payment award achieved, in other words the higher efficiency score the higher the payment received[10]. This way of applying the DEA method should be interpreted as a form of controlling or influencing the organisations to perform in a way corresponding to the goals of the public authority. Further it will be assumed that the organisations have two main goals: to obtain the highest possible award and to put as little effort in the work as possible. The relation between the work intensity and the output level is positive such that for a given input level the higher the work intensity the higher the output level. The goal for the public authority is to minimise the payments to the organisations and to encourage them to be highly productive (i.e. employ a high work intensity level). Finally, it is assumed that all organisations know that they will be evaluated by a DEA analysis. Given this knowledge the firm choose production plans accordingly. Thus I examine the problem of awarding the organisations ex ante, where the organisations can still change their production plans and can react in a strategic way to the fact that they will be evaluated and awarded according to the DEA results. This way of specifying the resource allocation problem is in line with the so-called principal agent models[11] where the agents are assumed

[10] The following is based mainly on Christensen et al. (1991) and Bogetoft (1990).

[11] See Grossman & Hart (1982) and Martin (1993).

to take care of some task for the principal. In this case the principal is the public authority and the agents are the organisations.

The role a DEA evaluation can play in this problem depends crucially on the way the award payment schedule is specified, see Bogetoft (1991). I will consider 3 different ways of specifying the payment schedule:

1. The payment is proportionally to the DEA-efficiency score, without an a priori fixed total amount.
2. The payment is related to the DEA-efficiency score, with an a priori fixed total amount.
3. The payment is related to the DEA efficiency score, without an a priori fixed total amount[12].

3.5.2.2.1. The payment is proportional to the DEA-efficiency score, without an a priori fixed total amount.

This payment schedule relates in a very direct manner the DEA efficiency score to the payment received by the organisation. The maximum payment received is when the organisation is efficient (i.e. with an efficiency score equal to 1). Thus even when there is no fixed total amount a priori, there is an upper limit to the total payments which is reached when all organisations receive a DEA efficiency score equal to 1. With this payment schedule the question is, how should the organisations construct production plans in an optimal way in order to satisfy the two goals of the highest possible payment and the lowest possible work intensity. Consider a given organisation, A. If it operates at a high work intensity level such that it becomes efficient, it will receive the maximum payment, but at the cost of sacrifying the other goal of low work intensity. Relative to this organisation there exist two groups of organisations; those which are dominated by organisation A due to a combination of similar production pattern and a lower work intensity and there is a group which are undominated by A[13], some with lower work intensity and some with higher work intensity compared with A, all receiving efficiency scores equal to 1. The group of organisations with a higher work intensity than A face the same problem, namely a non-optimal (organisational) production plan as a result of the high work intensity and are, therefore, not interesting to A if it wants to reach a production plan that corresponds to a higher level of achievement of both goals. The same holds for the group of organisations dominated by A: they have, indeed, a lower work intensity, but achieved at the cost of receiving a smaller award than A. However the organisations with a lower work intensity which are not dominated by A but receive the

[12] This payment schedule has not been proposed previously.

[13] Some of these organisations will be dominated by other organisations than A, but I will ignore this fact since it does not influence the proceeding arguments.

same payment are of interest to A. If A imitated one of these organisation's production plans it will reach a more optimal situation since it receives the same payment for a lower work intensity. Another possible strategy for A could be to specialise in a way such that it is no longer dominated and can choose a lower work intensity level without the risk of sacrifying the maximum payment.

In general, the strategies of imitation and specialisation where all organisations choose the lowest possible work intensity level are equilibrium strategies for all organisations, that is given the other organisations' choice of production plan each observation cannot improve its situation. If all organisations imitate (or specialise) they receive the same maximum payment and if one organisation increases its work intensity in order to obtain a higher payment it will still receive the same (maximum) payment[14]. Thus the organisations have no incentive to change from their low work intensity position. This payment schedule in combination with the DEA-evaluation does not help the public authority create incentives for high work intensity levels.

3.5.2.2.2. The payment is related to the DEA-efficiency score, with an a priori fixed total amount.

According to this payment schedule the public authority fixes a priori the total amount of payments to the organisations for rewarding good performance. The payments are still positively related to the size of the efficiency scores, although not proportionately. If all organisations receive the same efficiency score they will receive identical shares of the total amount of payments. In particular, if they all adhere to the strategies of imitation or specialisation (at the lowest work intensity level) they will receive efficiency scores equal to 1 and hence each organisation receives $1/n$ of the total amount. However, these strategies are not equilibrium strategies for the organisations, since it is optimal for the organisations to deviate.

Consider the situation where all imitate the low work intensity level and receive identical payment shares. Then if one organisation increases its work intensity level it will receive efficiency score equal to 1, but the other organisations will receive lower efficiency scores. Therefore, this organisation will obtain a share larger than $1/n$ of the fixed total amount of payments. The exact deviation from the imitation strategy will depend on how the organisations weight the utility of achieving higher payments relative to the disutility of the higher level of work intensity. The same holds for the strategy of specialisation. This specification of the payment schedule combined with the DEA evaluation secures that the organisations will attempt to reach higher levels of work intensity. The change has occured because of the trade off between the two goals of the organisations - high payment and low work intensity.

[14] Notice however that by increasing the work intensity the organisation will destroy the other organisations' optimal position. Therefore, the analysis assumes that this behaviour will not take place.

3.5.2.2.3. The payment is related to the DEA-efficiency score, without an a priori fixed total amount.

This payment schedule is similar to the first in the sense that the total amount of payments is not fixed a priori. However by contrast, this schedule does not give the payments to the organisations based only on the DEA efficiency score. In addition, the payment also depends on the number of times a given organisation dominates other organisations, that is the number of times an organisation appears as a peer unit. Thus the more times a unit dominates other organisations the higher the payment will be. How will the results change compared with the payment schedule where the payment is determined by the DEA efficiency score alone?

Consider the case where all organisations specialise at the lowest possible work intensity level such that they receive efficiency scores equal to 1. This strategy is no longer an equilibrium strategy because if one of the organisations deviate by imitating another organisation's production plan but at a higher work intensity level then this organisation will receive a higher payment since it obviously dominates the one with the lower work intensity. The exact increase in work intensity will depend on how the organisation weights the utility from increased payments against the disutility of higher work intensity level. Alternatively, consider the strategy where all organisations imitate the lowest possible work intensity level and that they still all receive DEA efficiency scores equal to 1. This strategy is no longer an equilibrium strategy. If one organisation changes its work intensity level it will imply that it dominates the rest of the organisations (the imitation strategy is defined as the case where all organisations undertake the same production plan). This organisation will receive the maximum level of payment since by dominating all other organisations it cannot dominate any other organisation in order to increase the payment. Thus this payment schedule can also provide a procedure that enables the public authority to induce the organisations to employ a high level of work intensity. Therefore, a slight modification of the relationship between the DEA evaluation and the payment schedule is enough to induce organisations to choose strategies with a high work intensity level.

In this section I have shown that the extent of the applicability of using DEA evaluations in the control of organisations by the public authority hinges on how the payment schedule is specified. If the payment schedule encourages competition between organisations then the DEA evaluations can be used.

3.6 The non-radial input efficiency index

As noted in section 3.4 radial efficiency indices are problematic due to the possibility of characterise observations as efficient which, in fact, should be inefficient. This problem is caused by radial efficiency indices being based on the Debreu-Farrell efficiency concept, which considers on input-output combination to be efficient if equi proportionate reductions (augmentations) in inputs (outputs) are not feasible. In contrast Koopmans's (1951) efficiency concept defines an efficient input-output combination to be one where a reduction (augmentation) in any input (output) is not possible. Therefore the slack-augmented DEA (FDH) was proposed. However, another solution is represented by non-radial efficiency indices which are based directly on Koopmans's efficiency concept. I will in the following analyse on such efficiency measure in terms of inputs. However the analysis could also have been based on outputs.

Definition: Consider output as fixed. In the case of strictly positive inputs define the non-radial Färe-Lovell (1978) input efficiency index as:

$E^i_{FL}(x, L) = \min\{\sum_j \theta_j/m | (\theta_1 x_1, ..., \theta_m x_m) \in L, \theta_j \in]0,1] \; \forall \; j, x \in R^m_{++}\}$,

where $L = \{x | (x, y) \in Y\}$.

If $\forall \; j, \; \theta_j = \theta$ then $E^i_{FL} = E^i_F$, making E^i_{FL} a generalization of E^i_F. E^i_{FL} is consistent with Koopmans concept of efficiency since $E^i_{FL} = 1$ if and only if $\theta_j = 1 \; \forall \; j$.

Consider $Y = Y_{CON}$. Using the Färe-Lovell input efficiency index the following n LP programming problems are obtained:

(18) $\quad \min \sum_j \theta_j/m$

s.t. $\sum_k \delta_k x_{kj} \leq \theta_j x_{k0j}, \sum_k \delta_k y_{ki} \geq y_{k0i}, 0 < \theta_j \leq 1, x_j > 0, \delta_k = 1, \geq 0$

$j = 1, ..., m, i = 1, ..., s, k = 1, ..., n$.

Notice that the degree of efficiency obtained by (18), $\sum_j \theta_j/m$, does not have the same intuitive interpretation as Farrell's input efficiency index θ since $((\sum_j \theta_j/m)x_k, y_k)$ is not necessarily an element of Y_{CON}. However, the information contained by the partial θ_j's is in some sense more valuable (that is, specific) than the aggregated information contained in Farrell's radial efficiency measure.

Compared to the slack model in section 3.4, the model based on the Färe-Lovell index seems less ad hoc. Further, it has the advantage that the slack is a relative measure and hence is independent of the units of measurement as opposed to that approach. In the above definition it does require however that inputs must be strictly positive which at first sight seems fairly harmless but in practice might involve some problems. Observations which include input amounts equal to zero will mainly originate from specialised units. Hence, either one has to design the data categories carefully or one has to aggregate the input categories which seems less satisfactory.

3.7 Allowance for categorical and exogenous inputs and outputs.

The efficiency measurement models presented in the preceding sections all have as implicit assumptions that the included inputs and outputs are continuous. Moreover if an input-oriented model was chosen the included inputs were assumed to be controllable and if an output-oriented model was chosen the included outputs were assumed to be controllable. Finally if the general DEA-model was chosen both inputs and outputs were assumed to be controllable. If non-controllable variables are present in the model as inputs or outputs it does not make sense for a solution to show that an inefficient unit can reach a efficient production level by reducing such inputs (or increasing such outputs). A similar problem holds in the case of the inclusion of categorical variables as inputs and outputs, since the maximum proportionate reduction (or augmentation) is likely to be ill-defined leading to meaningless results. The two types of variables imply that combining them gives 4 set of variables: (1) controllable categorical inputs and outputs, (2) non-controllable categorical inputs and outputs, (3) controllable non-categorical inputs and outputs, (4) non-controllable non-categorical inputs and outputs in principle. The preceding sections have only considered (3): controllable non-categorical inputs and outputs.

It can be claimed that both types of variables exist in the case of hospitals. An example of an exogenously fixed input could be the total number of beds determined at a level outside the hospital. A categorical input for a hospital could be the presence of a CAT-scanner.

In Banker et al. (1986a) the DEA model is altered to consider non-controllable non-categorical variables, whereas in Banker et al. (1986b) the DEA model is adapted to consider both non-controllable categorical variables and controllable categorical variables. Kamakura (1988) and Rousseau & Semple (1993) provide some refinements of the formulation in the original model of categorical controllable variables by Banker et al. (1986b). This section shows initially how to consider non-controllable non-categorical variables, then introduces non-controllable categorical variables and finally controllable categorical variables are included.

3.7.1 A DEA model with non-controllable non-categorical variables.

Consider the following input oriented DEA-V model with slack in the objective function:

$$(19)\ \min(\theta - \alpha(\sum_j s_j^+ + \sum_i s_i^-)$$

$$s.t.$$

$$\sum_k \delta_k x_{kj} + s_j^+ = \theta x_{k0j}, j = 1, ..., m$$

$$\sum_k \delta_k y_{ki} - s_i^- = y_{k0i}, i = 1, ..., s$$

$$\sum_k \delta_k = 1, k = 1, ..., n$$

$$s_j^+, s_i^-, \delta_k \geq 0$$

The unit $k0$ can achieve efficiency by reducing the j'th input by $(1-\theta)x_{k0j}+s_j^+$ and increasing the i'th output by s_i^-. However the above formulation assumes that the units control their inputs and outputs, since it indicates that the evaluated unit can reduce all the inputs and increase all the outputs in order to be efficient. If some of the m inputs are fixed then the above formulation is not appropriate in evaluating the efficiency of the units. The model must be such that only the maximum reduction in the non-fixed (discretionary) is considered given the level of the fixed inputs. Let m be the total number of inputs with the first m' controllable, that is the $(m - m')$ inputs are non-controllable. The modified DEA model, allowing for non-controllable inputs, can be specified as:

$$(20) \quad \min(\theta' - \alpha(\sum_j^{\prime} s_{j'}^+ + \sum_i s_i^-)$$

$$s.t.$$

$$\sum_k \delta_k x_{kj'} + s_{j'}^+ = \theta' x_{k0j'}, j' = 1, ..., m'$$

$$\sum_k \delta_k x_{kj} + s_j^+ = x_{k0j}, j = m' + 1, ..., m$$

$$\sum_k \delta_k y_{ki} - s_i^- = y_{k0i}, i = 1, ..., s$$

$$\sum_k \delta_k = 1$$

$$s_{j'}^+, s_j^+, s_i^-, \delta_k \geq 0$$

The main difference between (19) and (20) is that only the m' controllable inputs are reduced proportionately by the factor $(1 - \theta)$ and only the slack from the controllable inputs are included in the objective function. The constraint with respect to the non-controllable inputs is expressing simply that the solution must be such that the reference point has a level of fixed inputs equal to or smaller than the fixed input level for the evaluated unit. In general the efficiency score θ' is equal to or smaller than θ, since the solution to (19) is also feasible in (20), therefore with fewer restrictions with respect to the minimisation of θ', a lower value can be reached. It should be noted that a similar formulation can be achieved when some outputs are non-controllable. Notice further that allowing for non- controlable inputs (outputs) can be based on a slack-free DEA model.

3.7.2 3.7.2. Non-controllable categorical variables.

A categorical variable can be defined as a variable which can only be given a finite value. As examples of categorical variables regarding hospitals the following can be mentioned: the population basis for a hospital divided into a number of size related groups, a variable measuring patient satisfaction by groups such as very satisfied, satisfied, non satisfied, a 0-1 variable indicating the presence of a CAT-scanner at the hospital. If a non-controllable variable is categorical it implies that the preceding DEA-model (20) with allowance for non-controllable variables can create problems with respect to the interpretation of the solution. Therefore the reference point to which the $k0$ unit is evaluated might be ill-defined for the categorical variable since the reference point is defined as the convex combination of different units, which does not necessarily belong to the categorical scale. As an example consider the 0-1 variable equal to 1 if the hospital has a CAT-scanner and 0 otherwise and assume that the $k0$ hospital is dominated by a composite unit obtained by two hospitals with equal weight in the composite unit, where one of the hospitals has a CAT-scanner and the other does not have a CAT-scanner. Then the composite unit has half of a CAT-scanner, which is clearly impossible to interpret.

The convexity assumption combined with the presence of categorical non-controllable variables also creates problems in the way the units are evaluated. Consider again the categorical (non-controllable) input indicating the presence of a CAT-scanner. A hospital without a CAT-scanner can be evaluated in (20) with respect to a composite unit obtained by two hospitals one with CAT-scanner and the other one without. In this construction, it is implicitly assumed that the factor "CAT- scanner" has a constant marginal productivity, that is the output level is independent whether the hospital has a CAT-scanner or not, which might not be the case. The model should be changed such that the units are evaluated with respect to equal or worse environmental conditions. In the above example the model should imply that a hospital without a CAT-scanner is only compared to hospitals without a CT-scanner, whereas a hospital with a CAT-scanner can be compared to both hospitals which have a CAT-scanner and hospitals which do not. The inclusion of categorical non-controllable variables reduces the possible reference set used in the evaluation of each unit.

The problems of taking convex combinations of categorical non-controllable variables mentioned above can be solved by changing (20) in the following way, where for simplicity the analysis is restricted to one categorical variable[15]: assume that one of the $(m - m')$ non-controllable inputs is categorical e.g. the m'th input and suppose that the variable can take two levels (corresponding to the previous example). The evaluation of the units concerning this m'th non-controllable and categorical input is intended to be such that the reference point consists of units with the same or lower levels of this variable (corresponding to the statement that the evaluation in connection with categorical variables should be made by comparing the unit to other units in the same or worse conditions). In this example, a hos-

[15]Another possibility could be to perform the efficiency evaluation separately for each category, as noted by Ali & Seiford (1993).

pital without CAT-scanner should only be compared with other hospitals a scanner, whereas a hospital with a CAT-scanner can be compared with hospitals with or without. That is, for each unit that is contributing to the composite unit the following inequality has to hold with respect to the m'th input: $x_{k'm} \leq x_{k0m}$, where k' indexes the units contributing to the composite unit, i.e. $k' = 1, ..., h < n$ and $\delta_k > 0$. These constraints can be incorporated into the DEA-model (20) by introducing Q new variables d^q_{km}, the so-called descriptor variables, where $Q + 1$ is equal to the number of levels of the categorical variable can take. These descriptor variables are binary variables and therefore are similar to dummy variables. Therefore with the categorical input CAT- scanner/no CAT-scanner it is needed to introduce one descriptor variable $d^1_{km} = 0$ if CAT-scanner is not present and $d^1_{km} = 1$ if CT- scanner is present. The DEA model (20) can now be reformulated with the new variable included:

$$(21)\ \min(\theta' - \alpha(\sum_{j'} s^+_{j'} + \sum_i s^-_i)$$

$$s.t.$$

$$\sum_k \delta_k x_{kj'} + s^+_{j'} = \theta' x_{k0j'}, j' = 1, ..., m'$$

$$\sum_k \delta_k x_{kj} + s^+_j = x_{k0j}, j = m' + 1, ..., m - 1$$

$$\sum_k \delta_k y_{ki} - s^-_i = y_{k0i}, i = 1, ..., s$$

$$\sum_k \delta_k d^1_{km} \leq d^1_{k0m}$$

$$\sum_k \delta_k = 1$$

$$s^+_{j'}, s^+_j, s^-_i, \delta_k \geq 0$$

Thus if the hospital does not have a CAT-scanner $d^1_{k0m} = 0$ and in order for the last constraint to hold the units which form the composite unit must also have $d^1_{km} = 0$. On the other hand if the hospital has a CAT-scanner then $d^1_{km} = 1$ and the constraint can be satisfied with $d^1_{km} = 1$ as well as $d^1_{km} = 0$.

Comparing (21) with (20) it should be noted that the average efficiency score in (21) will be equal to or larger than the average efficiency score in (20), since each unit will have the same or a larger efficiency score in (21) relative to (20). This tendency towards higher efficiency scores is caused by the restrictions on the possible units that can be included to form the reference point, that is it becomes more difficult to find a composite unit that dominates the unit to be evaluated. implying higher efficiency scores. However the efficiency score will remain the same for those units unaffected by the restrictions on the reference set. In the example the efficiency scoore can increase for those hospitals without CAT-scanner, but will not change with certainty for those hospitals with one since the latter group can be compared to both types of hospitals.

Notice that it does not impose difficulties to allow for categorical variables with more than two levels, since it is just a matter of adding including descriptor variables and formulating the constraints as in (21). Moreover the above is done with respect to non-controllable categorical inputs but can easily be extended to allow for non-controllable categorical outputs.

3.7.3 Controllable categorical variables.

When turning from non-controllable categorical variables to controllable categorical variables the difference is that the latter type should be included in the objective function. The model will be constructed such that the objective function considers the maximum proportionately reduction in the controllable non-categorical inputs, the slacks in outputs and controllable non-categorical inputs and the largest decrease in the controllable categorical inputs. For simplicity I analyse only the case with a single controllable categorical input taking a range of levels, in which as a starting point, the number of levels will be limited to two corresponding to the previous section of non-controllable categorical variables. The m'th input is still categorical but now controllable. Consider a unit under evaluation at the high level of the m'th input corresponding to $d^1_{k0m} = 1$. The decrease in the m input can take a value equal to 0 or 1 with maximum equal to 1. (21) will be modified by changing the inequality $\sum_k \delta_k d^1_{km} \leq d^1_{k0m}$ to an equality: $\sum_k \delta_k d^1_{km} + t^1_m = d^1_{k0m}$, where t^1_m represents the possible change in the categorical input. Further t^1_m is included in the objective function in the same way as the normal slack variables, giving the following DEA model:

$$(22)\ \min(\theta' - \alpha(\sum_j s^+_{j'} + \sum_i s^-_i + t^1_m)$$

$$s.t.$$

$$\sum_k \delta_k x_{kj'} + s^+_{j'} = \theta' x_{k0j'}, j' = 1, ..., m'$$

$$\sum_k \delta_k x_{kj} + s^+_j = x_{k0j}, j = m' + 1, ..., m - 1$$

$$\sum_k \delta_k y_{ki} - s^-_i = y_{k0i}, i = 1, ..., s$$

$$\sum_k \delta_k d^1_{km} + t^1_m = d^1_{k0m}$$

$$\sum_k \delta_k = 1$$

$$s^+_{j'}, s^+_j, s^-_i, \delta_k \geq 0$$

However a problem with this model is that the objective function includes both continuous and binary slack. The continuous slacks, s_j^+ and s_i^-, depend on the units in which inputs and outputs are measured, while the binary slack is invariant to the unit of measurement[16]. This implies that the solution can be affected by the units of measurement and, in particular, that the degree to which the solution includes continuous slacks or binary slack can be changed. As noted by Rousseau & Semple (1993) this problem can be avoided by solving (22) in two stages: the first stage involves only continuous slacks and the second stage only the binary slack. The extension of (22) to consider more than two levels is straightforward. Let $Q+1$ be the number of levels the categorical controllable input can take, thus requiring Q descriptor variables d_{km}^q where $q = 1, ..., Q$. If the categorical controllable input is at the lowest level then $d_{km}^q = 0$ for all q, if it reaches the second lowest level then $d_{km}^1 = 1$ and $d_{km}^q = 0$ for $q \neq 1$ and finally if the input reaches the highest level then $d_{km}^q = 1$ for all q. Therefore for each unit the vector of descriptor variables consists of a string of ones followed by a string of zeros. (22) can then be reformulated simply as:

$$(23) \quad \min(\theta' - \alpha(\sum_{j'} s_{j'}^+ + \sum_i s_i^- + t_m^q)$$

$$s.t.$$

$$\sum_k \delta_k x_{kj'} + s_{j'}^+ = \theta' x_{k0j'}, j' = 1, ..., m'$$

$$\sum_k \delta_k x_{kj} + s_j^+ = x_{k0j}, j = m' + 1, ..., m - 1$$

$$\sum_k \delta_k y_{ki} - s_i^- = y_{k0i}, i = 1, ..., s$$

$$\sum_k \delta_k d_{km}^1 + t_m^q = d_{k0m}^q$$

$$\sum_k \delta_k = 1$$

$$s_{j'}^+, s_j^+, s_i^-, \delta_k \geq 0$$

Notice that (22) and (23) can be formulated in terms of categorical outputs. This extension is important since in most empirical applications categorical and/or non- controllable inputs and outputs exist.

Another possibility to take into account categorical variables would be to report the binary slacks along with the solution without including them in the objective function. This could be extended to consider continuous slacks such that the objective function only contains

[16]The dependence of the unit of measurement is a general problem with the slack-extended DEA model, as noted in section 3.4.

the maximum proportionate reduction in inputs. All slacks, binary or continuous, should be reported together with the maximum proportionate reduction in inputs. This procedure avoids the solution depending on the units of measurement as noted by Lovell (1993) and uses the extension of DEA to cover categorical variables.

3.8 Quality inclusion.

The inclusion of quality in efficiency evaluations involves at least two sets of problems. The first is related to the measurement of quality variables due the intangible dimension of quality. This problem will not be considered here but will be discussed in chapter 5 together with other data problems. The second set of problems is concerned with the changes of the standard DEA model required in order to take into account quality. Some of the problems have already been analysed in section 3.7 concerning categorical variables since quality-related variables are seldom measured on a continuous scale but only on a categorical scale. In fact, the inclusion of quality-related variables should simply take place in principle as additional elements to the production vector (ignoring the measurement problems) after determining whether the variables should be considered as inputs or outputs. This is the case in Fãre, Grosskopf, Roos and Ödegaard (1992) who extend the standard Malmquist productivity model to take quality into account[17]. They are in principle doing this by extending the production vector with variables measuring quality aspects.

However, the inclusion of quality in a DEA evaluation poses problems related to the formulation of the model. An important problem is the assumptions with respect to the nature of outputs. All the models in the preceding sections have assumed that outputs are strongly disposable, that is, if $(x,y) \in Y$ and $y' \leq y(x,y') \in Y$. This assumption implies that more of any output is preferred to less and that all outputs can be dumped. However, in the case of outputs available at different quality levels it might not be appropriate to assume that low quality outputs are freely disposable. This problem is essentially the problem of undesirable outputs in DEA models. In Fãre et al. (1989) the DEA model was changed to allow for a situation where some outputs are undesirable. The point is that disposing of undesirable outputs should not be costless. By changing the assumption of strong disposability for all outputs to be valid only for the desirable outputs and assuming weak disposability for the undesirable ones the costless reduction in undesirable outputs is restricted. Weak disposability of outputs implies that if $(x,y) \in Y \Rightarrow (x,\tau y)$ for $0 \leq \tau \leq 1$, it allows for the proportionate reduction of both desirable and undesirable outputs. Assuming weak disposability of some outputs allows for the presence of negative output prices. Other non-proportionate reductions in outputs (e.g. reducing only the undesirable outputs) are not allowed for, and may be feasible only if one or more inputs are increased. The meaning of this can best be understood with reference to the empirical application in Fãre et al

[17]The standard Malmquist productivity model is concerned with measurement of productivity due to efficiency changes and technology, i.e. a dynamic approach is followed. For more details see section 3.10.

(1989). They apply DEA in order to measure technical efficiency within a sample of US mills producing paper. Inputs are pulp, capital, labour and energy, the outputs are paper together with 4 pollutants (biochemical oxygen demand, total suspended solids, sulphur oxides and particulates). In this case it is indeed difficult to keep the assumption of strong disposable outputs (paper and the 4 pollutants) where the scope for reducing the 4 pollutants while leaving the output level for paper and input levels unchanged seems quite limited since the pollutants are byproducts of the main production process for paper. It seems likely that reductions in the levels of the 4 pollutants only are possible with corresponding reductions in paper output and/or increasing the input levels. Therefore the assumption of weak disposability for the pollutants can be justified.

The same can be claimed to hold for outputs of differing quality, where outputs of low quality level might only be weakly disposable, since such outputs could be undesirable. This is clearly the case for hospital products of low quality. This approach is followed in Olesen & Petersen (1993) who apply DEA in order to evaluate the performance of schools in Denmark. Quality within schools is measured according to the distribution of grades for pupils' final examinations where the lower grade levels are associated with low quality products which can be regarded as undesirable outputs. The model allows each school to define which particular outputs are undesirable and. desirable, where weak disposability is in force for the undesirable outputs. The results clearly suggest the importance of allowing for quality in an efficiency evaluation.

Quality inclusion also requires restrictions on the output weights in the DEA problem as formulated equation (6), such that the weights for the undesirable outputs do not exceed those for the desirable ones, e.g. in relation to school evaluation the weight to a given grade cannot exceed that given to a higher one. This corresponds to the idea in Charnes et al. (1990), where restrictions are imposed on the weights. This is implied by the model in Olesen & Petersen (1993). In Thanassoulis, Boussofiane & Dyson (1991) the idea of restricting the output weights is implemented in a more direct way in a DEA application on the performance of perinatal (birth clinics) care of DHA's in England. They argued that because of the choice of outputs and quality variables some restrictions on the weights are called for. Otherwise the chosen weights could be inappropriate with respect to the "true" weights. The inputs in their study were: whole time equivalent (WTE) obstetricians, WTE paediatricians, general practitioners' fees, WTE midwives, WTE nurses, number of babies at risk. The outputs were: total number of births performed in a DHA, number of deliveries to mothers resident in the DHA, number of special care consultant episodes, number of intensive care consultant episodes. The quality variables were: number of satisfied mothers, very satisfied mothers, number of abortions, number of babies at risk who survive. They consider several different sets of weight restrictions, but only one will be described. The restrictions are as follows: $v_{risk} = u_{survivors}, u_{survivors} \geq u_{deliveriestoresidents}, u_{survivors} \geq u_{abortions}, u_{survivors} \geq u_{deliveryepisodesinDHA}$. These restrictions are added to the standard DEA-model as formulated in equation (6). The first restriction reflects that the quality measure captures the survival rate which is determined by both the input, babies at risk and the output, babies at risk who survive. Without the equalisation of the weights it would allow for units improving efficiency by having a large number of babies at risk surviving or a low number of babies at risk unrelated to the survival rate. The other restrictions are concerned with placing importance on the quality measure number of babies at risk survivors relative to the other output and quality measures. A DHA cannot obtain a high efficiency evaluation if it is the result of putting a relatively low weight on the number of babies at risk surviving. The restrictions imposed seem reasonable, but, in addition, this procedure is open to a certain degree of ad-hoc orientation in the DEA model and thus removes of the generality of the DEA model. On the other hand, this modification allows for implementing specific knowledge about the valuation of inputs and outputs, which can be valid in an efficiency evaluation of a given sector's performance. Often the information level of the evaluator concerning the valuation of the inputs and outputs is larger than assumed in the standard DEA model where the choice of input output weights are unrestricted in the non-negative space.

3.9 Explaining inefficiency

Many attempts to decompose the radial technical efficiency measure into several sub-measures have been made in order to characterise the observed activities. Also measures of size efficiency have been proposed along with measures of most productive scale size (see e.g. Färe, Grosskopf & Lovell (1983) and Maindiratta (1990))

However, all these measures are built upon the same production data set as the 'original' technical efficiency measure and hence no new information is introduced by their calculation. Trying to "explain" the observed technical efficiency score by calculating e.g the most productive scale size does not provide any explanation of the actual level of technical efficiency. Hence, in order to be able to explain the outcome of DEA and FDH analysis, more information is obviously needed; information which relates to the characteristics not only of the purely technical side of the activity but also of the organisational environment.

One such way to provide explanations of the efficiency scores obtained is to interpret the calculated efficiency scores as a dependent variable which is determined by a set of environmental factors.

Let $\theta = (\theta_1, ..., \theta_n)$ denote the vector of efficiency scores for the n observations and Z as a $n \times L$ matrix of L environmental factors. A general regression model can be formulated as:

$$\theta_k = f(z_k; \beta) + e_k, \quad k = 1, ..., n$$

where β are the parameters to be estimated, z_k are the vector of environmental factors for the k'th unit and e_k is a disturbance term for the k'th unit. In order to estimate the vector of parameters β, assumptions about the functional form of $f(z_k;\beta)$ have to be made. This specification could be non-linear and therefore requires non-linear estimation techniques. However since no a priori knowledge about the relationship between θ and z_k is available the tradition of assuming a linear relationship will be followed, i.e. the model:

$$\theta = Z\beta + e,$$

Notice that although this regression model is linear it is still possible to consider non-linear transformations of the environmental factors provided that the transformed variables are linear with respect to θ.

A classical problem connected with regression analysis – the selection of the independent variables – appears in this model through the determination of the set Z of environmental factors. Obviously it is impossible to insure the inclusion of all relevant variables. The environmental factors can be divided roughly into two categories. One consists of uncontrollable variables exogeneous to the DMU's. In the case of hospitals an example of such a variable could be the patient mix reflecting that the patients are not homogeneous with respect to their demand of resources. This cannot be covered by the standard model since it would require disaggregation of the outputs to an extent which in practice is impossible and theoretically undesirable. The other group consists of variables which describes differences in the organisational structure of the DMU's. For hospitals, an example could be whether a hospital has a research department or not. Assuming that the research department affects efficiency through the input vector, since it uses resources, but if no outputs are registered the efficiency scores will be understated.

If the parameters in the linear model are estimated by OLS problems will occur because the vector θ of efficiency scores are restricted to take values between 0 and 1. This implies biased and inconsistent estimates of β. The estimates of β becomes biased (and therefore inconsistent) since $0 < \theta \leq 1 \Rightarrow 0 < Z\beta + e \leq 1 \Leftrightarrow Z\beta + e > 0$ and $Z\beta + e \leq 1 \Leftrightarrow e > -Z\beta$ and $e \leq 1 - Z\beta$. e is a function of Z and therefore correlated with Z. In order for OLS to give unbiased estimates of β, Z and e must be uncorrelated. This can be derived from the formula for the OLS estimate b:

$b = (Z'Z)^{-1}Z'\theta \Rightarrow b = (Z'Z)^{-1}Z'(Z\beta + e) \Rightarrow$
$b = (Z'Z)^{-1}Z'Z\beta + (Z'Z)^{-1}Z'e.$

Taking expectations of the previous expression gives:

$$E[b] = E[(Z'Z)^{-1}Z'Z\beta] + E[(Z'Z)^{-1}Z'e] \Rightarrow E[b] = \beta + E[(Z'Z)^{-1}Z'e].$$

If Z and e were uncorrelated the last term would disappear but with $0 < \theta \leq 1$ this does not happen.

Some transformation of θ is needed to solve the problem concerning the restrictions on θ. If the procedure of ranking efficient observations is applied (see section 3.4), θ is only bounded below by 0. In this case it is sufficient to use a logaritmic transformation of θ to obtain an unrestricted dependent variable, i.e. the model $\ln \theta = Z\beta + e$. For $\theta \rightarrow 0 \Rightarrow \ln \theta \rightarrow -\infty$ and for $\theta \rightarrow +\infty \Rightarrow \ln \theta \rightarrow +\infty$. This approach is chosen by Lovell, Walters & Wood (1990).

However, another transformation of θ may prove more satisfactory. Consider the following reformulation of the general version:

$$\ln((1-\theta)/\theta) = Z\beta + e,$$

For $\theta \rightarrow 0 \Rightarrow (1-\theta)/\theta \rightarrow +\infty \Rightarrow \ln((1-\theta)/\theta) \rightarrow +\infty$ and for $\theta \rightarrow 1 \Rightarrow (1-\theta)/\theta \rightarrow 0 \Rightarrow \ln((1-\theta)/\theta) \rightarrow -\infty$, i.e. $\ln((1-\theta)/\theta) \in]-\infty, +\infty[$. This transformation of θ alters the limited dependent variable to a dependent variable with an unrestricted range and OLS can be applied. Moreover, using this kind of transformation the problematic method of ranking efficient units is avoided, as proposed by Andersen & Petersen (1989). Notice that the sign of the estimates of the β's relate to the transformation and not to θ itself where the effect has the opposite sign.

If the variation in θ can be explained to a high degree by the Z variables this indicates that to a large extent efficiency scores can be explained to a large extent by specific institutional conditions rather than the mere excessive use of inputs.

3.10 Dynamic aspects

Up till now I have only considered single period problems – that is efficiency measurement of observations from one particular period relative to the technology of that period. But what if panel data were available? Not surprisingly, looking at efficiency in a dynamic context in order to examine long run efficiency trends proves to be more than just a smooth extension of the single period analysis. In general there are difficulties involved in the determination of the relevant time horizon of the analysis. The units may become incomparable over time with respect to previously fixed standards such as production categories, sample size etc, and since incomparability leads to meaningless results the long run efficiency results are easily distorted. Moreover, and just as importantly, there are purely methological problems as well.

As an introduction, consider the following example where the two periods are represented by their own technology:

DMU	(y_1, x_1^1, x_2^1)	(y_2, x_1^2, x_2^2)
A	(1,2,6)	(1,2,6)
B	(1,4,3)	(1,6,3)
C	(1,7,1)	(1,7,1)
D	(1,6,4)	(1,6,4)

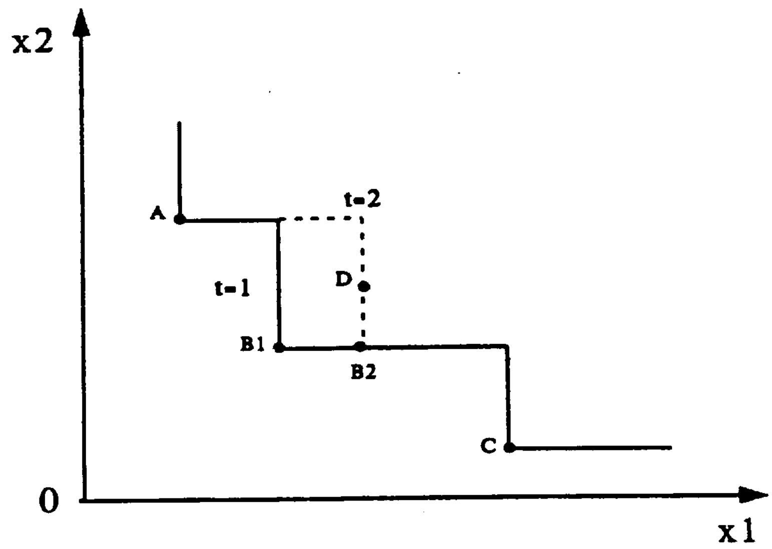

Figure 10. Illustrating the example

The two periods are almost identical except for unit B where the amount of input 1 is increased by two units in the second period. If input efficiency of the four units is measured in period 2 relative to the period 2 technology (the free disposal hull of Y_0^2); $\theta^{A2} = \theta^{B2} = \theta^{C2} = \theta^{D2} = 1$. Focusing on observation D one can notice that moving from period 1 to period 2 it has become input efficient since $\theta^{D1} = 0.75$ and $\theta^{D2} = 1$. But the production vector of D is identical in the two periods. Hence these changes in the degree of efficiency for the same observation cannot be interpreted over time in a direct manner. Additional information about the change in reference technology is needed. The usual approach in such situations is the use of index numbers as in the so-called Malmquist index.

Let $x^t \in \mathbf{R}_+^m$ and $y^t \in \mathbf{R}_+^s$ denote the input and output vectors respectively at time t $(t = 1, ..., T)$ and let Y_t denote the production possibility set at time t based on the set of observations at time $t, Y_0^t = \{(x_{kt}, y_{kt}) | k = 1, ..., n, t = 1, ..., T\}$. Let the history of production data till time t be given by the set:

$$YH^t = \{(x_k^\tau, y_k^\tau) \in Y_0^\tau | \tau \leq t\}.$$

In this way the production data can be viewed sequentially from $t = 1$ to $t = T$. Consider the dominant set $D^t(x_k^t, y_k^t, YH^t) = \{(x, y) \in YH^t; x \leq x_k^t, y \geq y_k^t\}$.

Definition: An observation $k0$ is sequentially undominated if and only if $D^t(x_{k0}^t, y_{k0}^t, YH^t)$ is a singleton. Otherwise the observation is declared sequentially dominated, i.e. previous observations exist which used the same amount or less of inputs and achieved the same or a higher amount of outputs.

Consider the following simple example of two DMU's in two periods:

DMU	(y_1, x_1)	(y_2, x_2)
A	(1,1)	(1,1)
B	(3,3)	(4,2)

YH^t is illustrated in figure 11. Notice that A is declared sequentially undominated in both periods although it has not changed its production vector. So is B, but B has increased its productivity in period 2. Hence the concept of sequential nondominance is rather weak in the sense that it cannot distinguish between observations that reach higher performance levels through time and those which remain unchanged – of course this is an advantage for the specialised units. Obviously it is possible to apply Farrell's input efficiency index to the

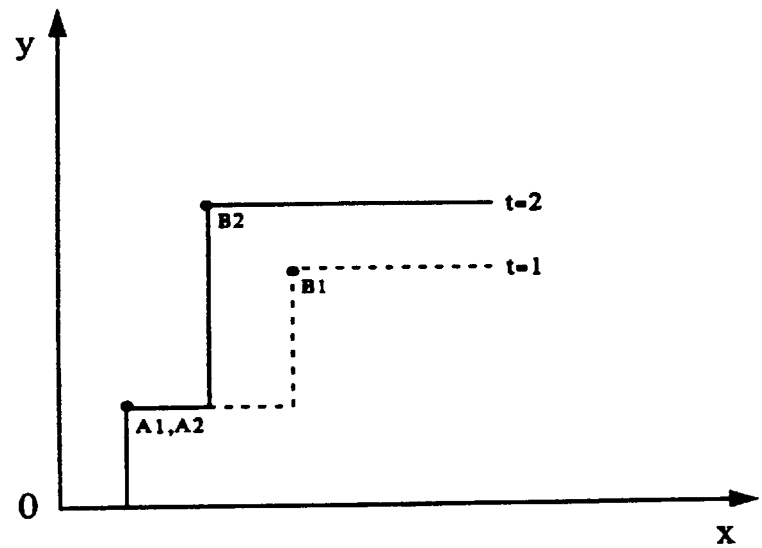

Figure 11. Illustrating YH^t of the example

sequential frontiers. This index will be well defined since $YH^t \subset YH^{t+1}$.

3.10.1 The Malmquist efficiency index

In order to evaluate the change in activity for a given unit from one period to another relative to a given frontier at time d, the Malmquist index approach can be used.

Let E_F^{dt} denote the Farrell input efficiency index for an observation at time t relative to the technology (the frontier) at time d based on Y_0^d.

Definition: Following the approach of Berg et al. (1992) the Malmquist input efficiency index between two periods 1 and 2 is defined as:

$$M_d(1,2) = \frac{E_F^{d2}}{E_F^{d1}} = \frac{E_F^{22}}{E_F^{11}} \frac{E_F^{d2}/E_F^{22}}{E_F^{d1}/E_F^{11}},$$

where $d = 1,2$ represents the reference technology. The first term is the ratio of input efficiency for the two periods i.e. a catching-up effect. The second term represents a frontier shift effect – that is it measures the distance between technology 2 and 1 based on the common reference technology. In figure 12, Q is inefficient in both periods. Hence both E_F^{22} and E_F^{11} are smaller than 1 and the catching-up effect is equal to:

$$\frac{OA/OC}{OB/OD}.$$

The frontier shift effect relative to technology 1 can be written as:

$$\frac{(OE/OC)/(OA/OC)}{(OB/OD)/(OB/OD)} = \frac{OE}{OA}.$$

Thus, $M_1(1,2) = \frac{OE/OC}{OB/OD}$. Hence the Malmquist index captures two different aspects of an efficiency development; efficiency measured relative to the period's own technology, and a shift in the frontier due to a technological change. If $M_d(1,2) > 1$ there has been a positive efficiency development, if $M_d(1,2) = 1$ efficiency has been constant and finally if $M_d(1,2) < 1$ a negative efficiency development has occured.

Consider the above example once more. Focus on the observation D and consider M_1. $E_F^{11} = 0.75$ and $E_F^{22} = 1$. Calculating E_F^{12} and E_F^{21} it is clear that $M_1(1,2) = 1$. Hence the Malmquist index reveals that the performance of D, relative to technology 1, is unchanged despite the fact that the efficiency scores indicate a positive development. The frontier change and the change in efficiency work in opposite directions and cancel out through the Malmquist index.

As noticed by Berg et al. (1992) the Malmquist index satisfies the circular test i.e. $M_d(0,1)$ times $M_d(1,2)$ equals $M_d(0,2)$. However, it is worth noticing that it is not possible to be certain that all Farrell indices involved in the definition of the Malmquist index are well defined. If the sequential frontier is considered in order to take the history of production data into account it is possible to make tables such as table 1. It is easily observed that the efficiency indices of the lower part of the table are not necessarily well defined due to possible observations from the period in question, dominating all previous technologies d. Hence the Malmquist index which involves any such Farrell indices is not generally defined. This may be seen as a major drawback of the Malmquist approach since, although chainable, the index cannot cover all stages of a given production development. Figure 13 illustrates a case where E_F^{12} is not defined.

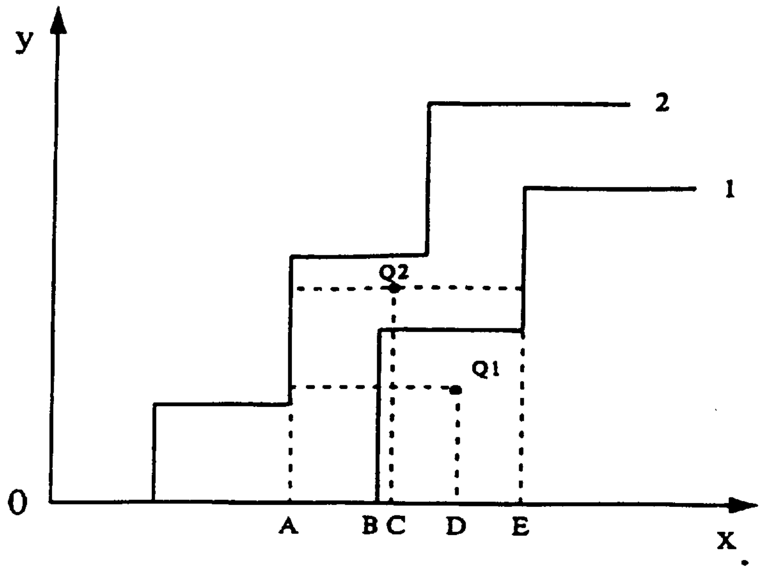

Figure 12. The Malmquist input efficiency index.

unit A	YH^1	YH^2	YH^3	YH^4
period 1	E_F^{11}	E_F^{21}	E_F^{31}	E_F^{41}
period 2	-	E_F^{22}	E_F^{32}	E_F^{42}
period 3	-	-	E_F^{33}	E_F^{43}
period 4	-	-	-	E_F^{44}

Table 1. Sequential performance indices

3.11 Advantages and problems of the DEA model and FDH model.

The main advantage of DEA (and FDH) is that no a priori functional form of the relationship between inputs and outputs has to be assumed, see e.g. Seiford & Thrall (1990). Except for the assumptions of convexity and free disposability of inputs and outputs, the DEA model allows the observations determine the (implicit) production frontier, i.e. the benchmark used for the efficiency comparisons. In the case of a FDH model the only assumption is free disposability of inputs and outputs. The reason for this being an advantage is that, a priori, the information concerning the relationship between inputs and outputs of a particular set of observations (e.g. a group of schools) will normally be quite limited. Thus, assuming a particular form (e.g. Cobb-Douglas) could easily result in cases of misspecification. Of course it is possible that the a priori choice of functional form will be the correct specification,

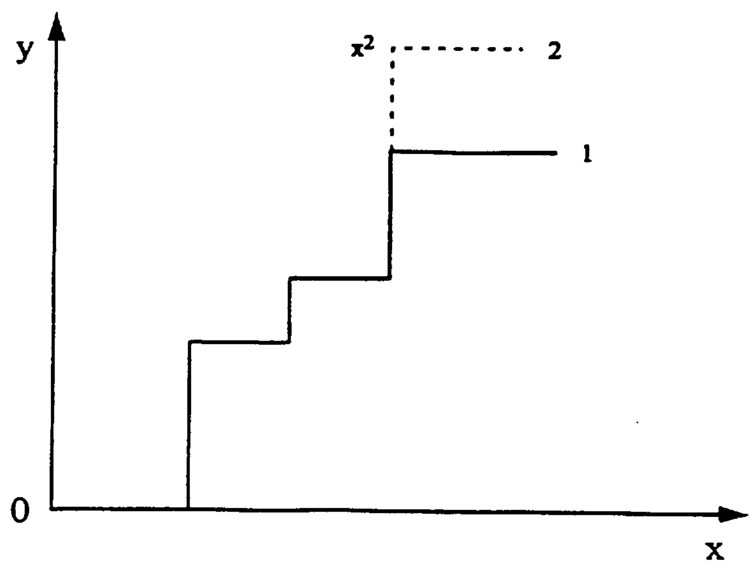

Figure 13. A case where E_F^{12} has no solution.

but in normal circumstances this seems highly unlikely. Therefore a wiser path is to use assumption-free methods such as DEA (or FDH), where the (implicit) frontier is determined by the data.

Another advantage of the DEA and FDH models is the possibility of allowing for efficiency comparisons of organisations that use multiple inputs to produce multiple outputs, see Banker et al. (1986). In the DEA and FDH models it is not necessary to construct single input (and/or) single output representations of the production process. The models are constructed to handle multiple input- multiple output cases which seems to be the typical situation in most organisations. In relation to this it is not necessary to search for weights that can be used to construct aggregated output measures. In particular this is of importance in non- profit making organisations and organisations situated in the public sector where prices, that otherwise could have served as weights, are not available. Outputs and inputs are thus not required to be measured in the same identical units but can be measured in their natural units. In the case of hospitals it is possible to include in same the DEA and FDH models such diversely measured outputs as e.g. the number of patients in diagnosis groups and the number of laboratory tests and inputs such as the number of employees and the expenditure on goods and services.

A further advantage of the DEA model (and FDH model) is that is consistent with production theory since inefficiency is defined relative to a best performance standard corresponding to the definition of the well known production function, which indicates the maximum level of output for a given level of input. This correspondence is in contrast to the traditional estimation of average production functions with observations being placed above and below the estimated function, see Bowlin et al. (1985) for a comparison of DEA with an average production function.

Finally, it should be noted that although the efficiency concept in the DEA model (and FDH model) is weak, it is an advantage since it implies that if inefficiency under these conditions still prevails then this seems a more valid result and that this inefficiency is difficult to ignore.

However, the DEA and FDH model also involve certain problems some of which exist in other performance methods and some are specific to these models. First, the DEA and FDH model obviously requires the specification and measurement of inputs and outputs for the concerned observations; see Epstein & Henderson (1989). Ideally all inputs and outputs relevant for the units should be included in the model. In practice this can be difficult particularly in connection with evaluations of units which produce services, where outputs and inputs typically will be quite intangible (for further details about specification and measurement of inputs and outputs with respect to the hospital sector see chapter 5). Moreover, the units concerned should be characterised by the same inputs and outputs, otherwise the units will become incomparable and incomparable units in the DEA model (and FDH model) implies that these units will be deemed efficient by specialising in the unique output or input. Thus, it is important to examine whether the included units are similar with respect to the input and output categories. This problem is also relevant if these models are used to evaluate a given observation over time, when new products will appear and thus result in non-comparabilities.

Although DEA and FDH is constructed such that multiple inputs and outputs can be included restrictions exist on the possible number that can be included compared with the number of observations in order to preserve the discriminating power of these models. A large number of inputs and outputs result in a large number of possibilities for specialisation for the observations and thus be deemed efficient. Thus the larger the number of inputs and outputs the greater this possibility becomes; see Nunamaker (1985). The number of efficient unit in DEA will reflect more or less the product of the number of inputs and outputs. Unless the number of observations is larger than this product the DEA-evaluation turns out to be meaningless. Therefore, aggregation of inputs and outputs can be necessary in order to reduce the number inputs and outputs relative to the number of observations. However, aggregating data can result in some information being lost. See Olesen & Petersen (1993) for an aggregation procedure for DEA, which does not involve information losses. In connection with this, it is important that the aggregated input and output categories are internally homogenous. Another possibility is the addition of new observations in the original sample, but this can be rather limited possibility due to the previously-mentioned requirement of the similarity of the observations.

Because of either incomparabilities in the data set or inclusion of too many input and output categories the number of efficient units in empirical analyses often turns out to be large.

This limits the discriminating power of these analyses since only a small fraction of the observations can be ranked. If it is not possible to reduce the number of output and input categories and the observations seem comparable but a large number of efficient units have been found it is still possible, in a DEA-C model, to evaluate or rank the efficient units using the procedure proposed by Andersen & Petersen (1989); see section 3.4. However the suggested procedure is not general: Only the DEA model with constant returns to scale will provide solutions for all observations.

Fundamentally the efficiency concept in the DEA and FDH model is problematic since it is based on relative efficiency. A unit which is declared efficient with respect to the other units is only efficient relative to these units. A DEA (or FDH) efficient unit is not necessarily efficient in absolute terms, it is only that the performance of the other units is worse. Thus efficient hospitals can also improve their performance. In order to avoid this problem a blue-print technology has to be constructed, a reference based on "engineering" data. However for most sectors (including hospitals) such a technology is very difficult, if not impossible, to construct. Therefore no solution to this problem seems to be at hand and in the empirical application of DEA (and FDH) it can only be noted that the constructed reference is not an absolute efficiency reference.

A more relevant problem regarding empirical analyses based on DEA is that the construction of the frontier can be seriously affected by the inclusion of outliers in the sample, see Charnes et al. (1985a) for a sensitivity analysis of the DEA-results. The reason for outliers creating a more serious problem than normal is a result of the property of DEA where the constructed frontier is based only on a subset of the total set of observations (recalling the example in figure 1 where the DEA-C frontier was based only on 1 of 6 observations). If some of these observations are special they will distort the results. This has led to the development of stochastic DEA-models, see e.g. Sengupta (1990) and Olesen & Pedersen (1989). Therefore, careful data screening is called for in order to limit the presence of outliers. It should be noted that the outlier problem is less present in the FDH model since the frontier is based on more observations.

For the DEA models the problem of assuming convexity of the production possibility should be mentioned. As such this assumption precludes the production possibility set from obtaining increasing returns to scale. Moreover the same issue of lacking information appears with convexity as it did when assuming a specific functional form. In fact the production possibility set can turn out to satisfy convexity but a priori it is not possible to know this. In addition, the convexity assumption implies that an observation can be compared with linear combinations and be declared efficient against such hypothetical observations. It can, indeed, be difficult to explain to the units concerned that they are inefficient with respect to the linear combination of other observations. A more desirable case will be if an observation is inefficient relative to a real, existing observation.

However, without assuming convexity of the production possibility set as in FDH model, it becomes more difficult to obtain situations where $\sum_k \delta_k x_{kj}$ is strictly less than the observation to be evaluated, i.e. more observations will be deemed efficient in the FDH model compared with the DEA model. Thus the problem of a large fraction of the observations having efficiency scores equal to 1 in the DEA-model is accentuated in the FDH-model. In addition there is scope for observations which are undominated (i.e. no observation in the data set uses less or the same amount of inputs to produce the same or a greater level of the outputs) but neither are they dominating any other observations. The efficiency of such observations is, primarily, the result of the special input-output orientation of these units.

Chapter 4

The Danish health care sector - structure, advantages and problems.

4.1 Introduction.

The preceding chapters have only examined the methodological issues related to the possibility of evaluating the productive efficiency of organisations. In this chapter I will approach the empirical content of the thesis - measurement of efficiency within the hospital sector - by examinating the structure of the Danish hospital sector and the possible advantages and problems related with this specific structure.

The procedure by which I will analyse the Danish hospital sector (and in particular the problems of that sector) is to describe the sector and use this description to understand the problems (and advantages) related to its structure. Some of the problems can be understood as arising from the different agents' reactions to the incentive structure and not something that is external. Therefore by changing the structure it is likely that agents' behaviour will change and so structural changes could provide solutions to the existing problems in the sector. However as into other cases, changing from one structure to another will involve new problems (and advantages), since it entailes is a choice between imperfect structures.

The plan for the chapter can be outlined as follows. In the following section (sec. 4.2) I will present a statistical description of the Danish health care sector. The next section (sec. 4.3) will consider the financial aspects of the Danish health care sector: how the health care sector provides sources for financing the expenditures and how the different agents are paid. Section 4.4 analyses the interaction between the different health care providers. In the following section (sec. 4.5) the structure of the hospital sector is discussed. The final section (sec. 4.6) discusses the problems as well as the advantages of the described structure.

4.2 A statistical description of the Danish health care sector.

In this section I will use figures to describe the structure and importance of the Danish health care sector in general and the hospital sector in particular. I will focus on relative measures in order to avoid drawing conclusions from figures that cannot be related to other countries health care sectors, e.g. due to different units of measurement.

Consider figure 1, where total expenditure on health care, as a share of Gross Domestic Product at factor prices, is depicted together with public health care expenditures as a share of GDP and of total public expenditure for the period 1970-1990. Health care expenditure and GDP are measured in current prices, thus making interpretations of the real growth of the expenditure share dependent on the assumption that health care expenditure experiences price increases parallel to the GDP deflator.

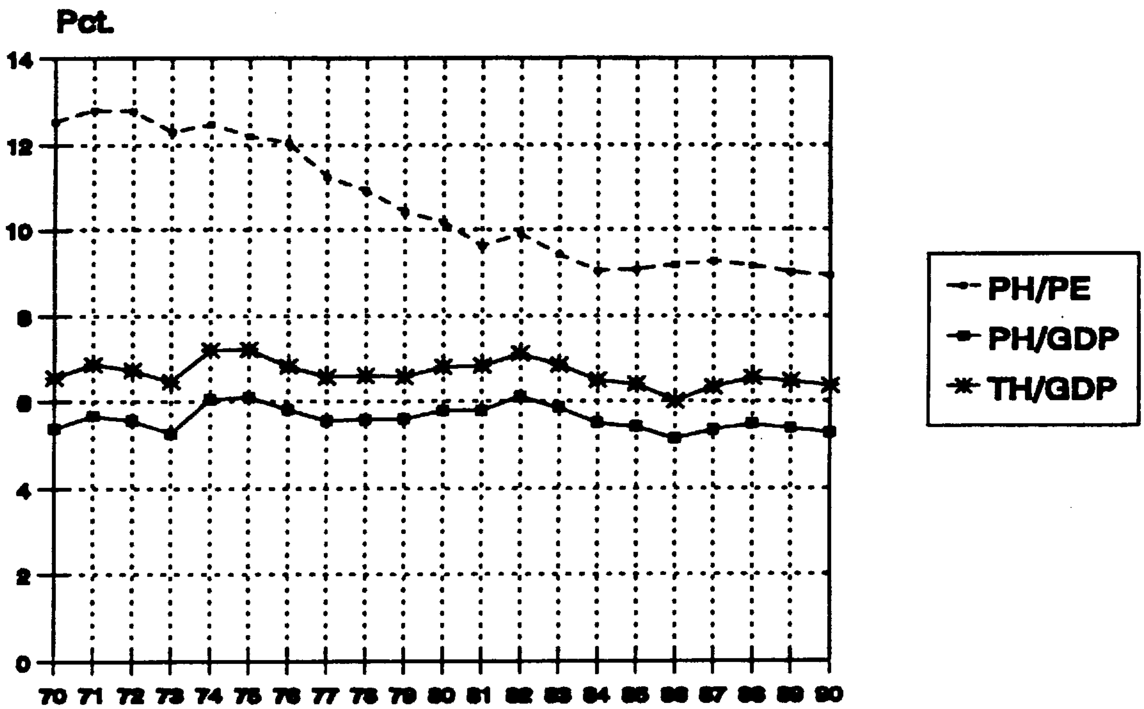

Figure 1. The relative growth of health care expenditure for Denmark in the period 1970-1990.

Note: TH is total health care expenditure, PH is public health care expenditure and PE is public expenditure.

Source: Statistisk Tiårsoversigt, several years.

The main conclusion to be drawn from the figure is that the growth in health care expenditure has been very stable. Total health care expenditure amounted to 6.6. % in 1970 and to 6.4 % in 1990 [1]. Health care expenditure had however, a slight tendency to increase in the first ten-year period, followed by an equally slight tendency to decrease in the second ten-year period.

This picture of stability is confirmed when looking at both private and public share of total health care expenditure, where private health care expenditure in 1970 represented 17.9 % of total expenditure with the corresponding share in 1990 of 17.2 %. This stable growth in the private - public distribution of health care expenditure is illustrated by the parallel growth in the total health care expenditure and public health care expenditure as a purcentage of GDP, that is the increases in public health care expenditure has kept in line with growth in GDP This growth in public health care expenditure represents a decreasing share of the total public expenditure; health care expenditure accounted for 12.5 % of total public expenditure in 1970, while, by 1990, it accounted for only 8.9 %. The constant share of public health care expenditure has been combined with an increasing share of total public expenditure which stood at 42.9 % 1970 and 59 % 1990. From 1974 increased public expenditure were caused partly by the increase in social related expenditure due to (among other factors) the high unemployment level from 1974. The other main reason is related to the first; the increased expenditure were not covered by taxes but by loans resulting in increased interest payments on loans. The largest increases were from 1970-80 when total public expenditure increased by 62 % measured in fixed prices (1980=100). In the period from 1980-1990 the increases were more modest equal to 25.1 %. In 1982 a more restrictive policy of financing was implemented at the governmental level in order to contain public expenditure, because the Danish tax level had reached a magnitude where further increases seemed difficult to obtain and because the balance of payment deficits over two decades had lead to the accumulation of a large external debt. The growth in total public expenditure can provide an explanation of the growth in the share of public health care. The most important decrease in the first ten-year period, when other expenditure increased more than that of public health care. In the second period the restrictive expenditure policy was implemented applying pressure on all categories of expenditure.

Although the above showed the importance of public health care expenditure, this becomes clearer when looking at the public sector expenditure in the counties, where total health care expenditure has a share of total expenditure of 65 %. This proportion is due to the fact that, by law, the counties are responsible for the majority of health care provision in Denmark. Notice however, that around 20 %. of expenditure are covered by general state subsidies (see Larsen (1990)).

[1] This pattern is unusually in comparison to the rest of the OECD-countries, in 1990 the share for Denmark is among the smallest with only the shares for Greece and UK being smaller, see OECD (1993). However, international comparisons of health care expenditure should be interpreted with caution. For some recent international comparisons see e.g. Søgaard (1991), Scieber & Pollier (1990) and OECD (1987).

Next I will consider the different sub-components of total public health care expenditure. In table 1 I have listed the share of expenditure for the different health care categories for 1970 and 1990.

	1970	1990
Public expenditures	82.8	82.5
Hospitals	60.5	61.4
Individual health care sevices	19.6	19.8
Administration	2.1	1.0
Other	0.6	0.3
Private expenditures	17.2	17.5
Medicaments	4.6	5.1
Glasses, hearing instruments etc.	1.9	3.6
Doctors, Dentists	8.0	5.5
Nursing homes, sanatories	1.3	1.5
Health insurances	1.5	1.8
Total health care expenditure	100.0	100.0

	Annual % increase 70-80	Annual % increase 80-90
Public expenditures	3.1	0.7
Hospitals	3.5	0.5
Individual health care sevices	2.5	1.4
Administration	-8.4	5.8
Other	3.1	-5.1
Private expenditures	1.5	2.5
Medicaments	5.9	0.8
Glasses, hearing instruments etc.	6.8	2.5
Doctors, Dentists	-2.8	3.4
Nursing homes, sanatories	0.3	3.7
Health insurances	-3.7	3.8
Total health care expenditure	2.8	1.0

Table 1. The share of expenditure share for the different components of the health care sector.
Note: The private-public shares are slightly different from the numbers reported earlier due to other data.
Source: Hansen (1993).

First of all the table shows the division between public expenditure and private expenditure. Hospitals are, in general, financed from public funds (except for a very few hospitals which are financed mainly through private provided funds). Individual health care services refer to general practioners, specialised practioners, subsidies for medicaments and dental care, nursing services for elderly persons in their own home and health prevention schemes. In addition to hospital expenditure, the counties are also responsible for expenditure related to practioners and subsidies for medicaments and dental care through the county-based health insurance scheme. Expenditure on nursing services for elderly persons in their own homes are incurred at the municipality level as are the majority of the health prevention schemes. Private expenditure on medicaments and dental care, however, are partially covered by public funds. The other categories consist of more specialised health services and different types of support services, which are not included as part of public health care.

Table 1 clearly indicates the importance of hospitals as share of total health care expenditure; hospitals have a share equal to 61.4 % in 1990. This represents 74.4 % of the total public health care expenditure. The individual health care services also account for a large share of health care expenditure, 19.8 % in 1990, of which county expenditure on general practioners, specialised practioners, medicaments and dental care account for 12.6 % (with expenditure on GP's accounting for the largest share, equal to 5.7 %) [2]. Notice that public expenditure on administration has a share of 1 % of total health care expenditure, which, in comparison with other OECD countries is rather small : in the US the share is circa 15 % (see Hansen (1993)) and in Germany circa 15 % (see Petersen (1990)). Payments for medicaments, 5.1 %, physicians' services and dental care, 5.5 %, represent the largest share of private health care expenditure. The second category is mainly made up of dental care expenditure, since physicians' services are usually part of public expenditure. It should be noticed that, although, the shares indicate the importance of hospitals in terms of direct expenditure general practioners (and, to a lesser extent, specialised practioners) have an important influence on health care expenditure. General practioners represent, in most cases, the first contact between a patient and the health care system. Therefore, it is the general practioner who visit the patient initially and decide what steps should be taken, one alternative being whether to hospitalise the patient for more specialised treatment. I will return to the role of the private practitioners in the Danish health care system in section 4.4.

The share of health care expenditure was, more or less, of the same magnitude in 1970 and 1990. However, as table 1, also shows, public expenditure increased in real terms by a greater percentage in the first ten- year period (3.1 %) than did private expenditure (1.5 %), while in the second ten-year period the growth pattern was reversed with public expenditure increases of 0.7 % being smaller than those of private expenditure, 2.5 %. This pattern implies that the shares have remained at the same level between 1970 and 1990. According to Hansen (1993) this stable growth can be confirmed when analysing the employment statistics: in the ten-year period from 1980-90 the number of persons employed in the health care sector was around 147000, a share of the total employment equal to 5.5 %.

In order to examine whether this stability of the overall health care sector can be confirmed when looking at sub-sector level I will analyse the employment and activity growth for the hospital sector for the period 1975-90[3]. Personnel data are calculated as the number of full-time employed persons, where those employed on a part-time basis are included with a weight of 0.5 up to 1985 and, from 1986, according to the number of hours worked. The total number

[2]The other components is specialised practioners 2.0 %, dental care 1.6 % and medicaments 3.3 %. The 12.6 % is the sum of these 4 categories. The source for these data is the Danish county Organisation.

[3]This period was chosen because of the changes in the definition of employment between the years 1970-1974 and 1974 is missing, National Board of Health (1976). Information on the number of out-patient visits is missing from the activity data for general hospitals for the years 1970-75.

of employees in hospitals has increased from 73153 in 1975 to 81905 in 1990, an increase of 12 %. This covers an increase from 1975 to 1983 equal to 18.2 %, after which the number of employed remained constant until 1990 when the number of employees decreased by 4 %. The sub-groups have experienced very different growth patterns, where doctors, nurses and other health care personnel have increased in number while other types of personnel have decreased during the period 1975-1990. The number of doctors increased by 62.3 % from 1975 to 1990, nurses by 43.1 %, other health care educated personnel by 9.9 % and other personnel categories decreased by 17.4 %. This indicates that the personnel structure has changed towards more health care-oriented personnel. In particular, the treating part of the personnel (doctors) has increased, whereas the caring part (nurses and other health care educated personnel) has experienced a more modest increase.

By comparison with the growth in the hospital resources (i.e. personnel) the growth in 3 indicators of hospital production will be described below: the number of discharges, patient days and out- patient visits for both general hospitals and psychiatric hospitals[4].

The number of discharges from general hospitals has increased by 25.2 % for the period 1975-90 (the discharges from psychiatric hospitals decreased by 25.4 %) but the number of patient days from general hospitals decreased in the same period decreased by 21.5 % (the number of patient days from psychiatric hospitals decreased with 68.4 %). The number of out-patient visits from general hospitals have experienced a marked increase in the period 1976-90 equal to 40.9 % (the outpatient visits from psychiatric hospitals increased in the same period by 68.1 %). This growth implies that the average length a patient stays in general hospitals has decreased from 11.2 days in 1975 to 7.1 days in 1990[5] . This decrease in the average length of time a patient is hospitalised should be related to the increase in the number of outpatient visits. Tests and controls previously conducted while patients were hospitalised are conducted increasingly on an out-patient basis. Moreover improved treatment techniques have also implied that for some diagnoses it is now possible to be treated without hospitalisation (or at least with a shorter period of hospitalisation). However, a part of the decline in average length of stay is not real, but a result of the way discharges are registered. A discharge is related to a specific department and therefore if a patient is transferred to another department within the hospital this transfer is counted as a discharge.

[4]However, the activity at psychiatric hospitals is however very different from other hospitals. Notice that while circa 1.000.000 discharges from general hospitals generated circa 3.800.000 patient days, circa 20.000 discharges from psychiatric hospitals generated 1.000.000 patient days. Admission to psychiatric hospitals implies, on average, a long length of stay. Therefore what follows concentrates primarily on general hospitals.

[5]The average length of stay ($MTIME$) is defined as the number of patient days ($PDAY$) divided by the number of discharges ($DISCH$), thus it can be expressed as:

$$MTIME = \frac{PDAY}{DISCH}$$

The expression indicates how many patient days a discharge on average result in.

In addition, there has been a tendency towards an increased number of departments making it more likely, in fact, that patients are transferred. Instead of having a patient for a long period in the same department there has been a development towards shorter periods at several departments. At the aggregate level this will appear as a higher number of discharges and shorter average length of stay and therefore account for part of the increase in discharges and the decline in the average length of stay (Vallgårda (1992)).

The possible change in the average length of stay is clearly dependent on how other parts of the health care sector are functioning, e.g. are there out-patient facilities present or can some patients be treated by practioners (without hospitalisation).

The decrease in the average length of stay should be seen in connection with the change in the hospitals personnel structure. The shorter time patients are hospitalised implies more intensive courses of treatment for the patients. Health care personnel become more important, especially those categories which treat the patients (mainly physicians). Those personnel categories providing care for the patients (nurses and nursing personnel) should become less important because of the shorter time the patients are actually in hospital. When this is not the total explanation, it is probably because of the need for more intensive care during shorter periods of hospitalisation, quality improvements and changes in the job content.

4.3 Financial aspects of the Danish health care sector.

Countries differ according to how the health care sector is constructed. However in one aspect all countries health care systems are similar. A given health care system is always characterised by 3 agents: (1) patients (demanders), (2) producers and (3) the funding system. Unlike most other markets the health care market, on the whole involves payments being made between the funding system to the producers, and not directly between the patients to the producers (or only to a small extent). In this way the health care market is in contrast to other markets where there exists, in general, a direct economic relationship between buyers and sellers; this does not hold for the market for the provision of health care goods. The country differences are not a result of the presence of a third-party financing system but of how this third-party financing system is constructed. The different dimensions of the third-party financing system are: (1) the extent to which the public sector is part of the financing system, (2) how the third-party financing system is provided with funds, that is, the means for obtaining funds and which groups are contributing to this system, (3) how the producers of health care goods (practitioners and hospitals) are paid.

4.3.1 The extent to which the public sector plays a role in financing the health care sector.

At some level, the public sector plays a role in the third-party financing system of the health care sector in all countries. The methods can differ however. Even in countries (e.g. USA) where private insurance companies take an important part of the financing system, the public sector is not absent either because of subsidies for some groups of the population or by providing alternative public funds for those groups not covered by private insurance companies.

The third-party financing system in Denmark is mainly organised within the public sector; that is, public funds provide a major part of the financing resources of the health care sector. In addition, a non-profit making insurance company plays a minor role in financing some specific health care goods not covered by public funds or only partially covered by public funds, e.g. dental care, purchase of remedies, purchase of medicaments etc. Private profit-making insurance companies do not play any significant role in financing the health care sector in Denmark.

4.3.2 The provision of funds to the third-party financing system.

As noticed above the third-party financing system can obtain funds in different ways. Tax-based financing and insurance- based financing are the two most frequently used possibilities. These two possibilities are broadly defined and can take on many forms. Tax-based financing can be in the form of general taxes or taxes with specific reference to the health care sector. In the former case the financing of the health care sector is integrated with the financing of the public sector overall, whereas the latter, by definition, implies that it is possible to separate the financing of the health care sector from the rest of the public sector and that there is a specific political decision to be made about how the level of resources to be allocated to the health care sector. Moreover, a system of tax-based financing can be constructed in many ways, i.e. which groups are taxed, which subjects are taxed and how they are taxed.

Insurance-based financing can also be set up in a number of different forms. Insurance-based financing can be provided on a public as well as private basis (the same is obviously not the case with respect to tax-based financing). In relation to this distinction it raises the issue of a compulsory vs. an optional insurance system, where the former is connected mainly with a system within a public organisation and the latter is, in general, related mainly with a private organisation. Another distinction is whether the premium is actually fair or not, that is whether the premium corresponds with the expected expenditures or not. Insurance systems can differ with respect to the extent of risk coverage and presence of deductibles etc. Similar to tax-based systems, insurance-based systems can be different with respect to which groups contribute to the provision of funds, e.g. employer, employees or patient based contributions.

In Denmark regional public authorities are the main agents of the third-party financing system. These regional public authorities consist of 14 counties and two municipalities Copenhagen and Frederiksberg in the capital. This structure should be seen in connection with the regional organisation of most health care provision in Denmark, which concerns hospitals, practitioners, subsidies for dentists and subsidies for medicaments. The provision of funds are obtained through the counties' possibility of imposing general income taxes on their residents. In addition the state supplies general subsidies to the counties such that these subsidies provide for some of the funds to the regional health care sector. The state subsidies account for around 20 % of the total funds to the regional health care sector. For some health care services, patients have to pay a part of the incurred expenditure, this is in particular the case with respect to dental care and medicaments. For these health care goods the public tax-collected funds only provide partial coverage. However a not-for-profit insurance company, "Danmark", provides, though not complete, additional coverage for these health care goods. These non-public sources of funds for health care are of minor importance, public funding is predominant. Thus hospital services and practitioners' services are in general free of any direct patient charges.

4.3.3 The payments to the producers of health care goods.

The payment of the producers of health care services from the third-party agents (in the Danish case the regional public authorities) can take a number of different forms[6]. Each form has advantages and problems and can be expected to result in certain reactions according to the incentive structure. The most common forms are the following: (1) budgets, (2) fee-for-service payment.

If the payment method of the producers takes the form of budgets it implies that a given amount of funds is supplied to each producer in order to cover all expenditure related to her production in a given period. The budgets can be established in a number of different ways.

In general, the budgets are related to some indications of the activity level. GP's have been paid according to the number of persons connected to the practice where they receive a fee for each person, the so-called per-capita principle. Hospitals can receive the budgets according to the expected number of patient days or admissions. An example of an activity-based budget could be the American Diagnosis Related Groups system, which links the hospitals' diagnosis profile with resources. The link between the budget and the activity level is constructed through unit costs; see chapter 5 and Petersen (1990).

An important problem of budget based financing is the possible reactions if the actual expenditures exceed the budget approved. If the budget is fixed such a system allows for the producers can to go bankruptcy or less seriously the deficit is transferred to the proceeding budget period (i.e. a reduced budget in the following period). The funding system can be

[6]The following is mainly based on Mcguire et al. (1988).

expected to prefer a fixed budget since this should give incentives to minimise the costs by the producers such that the expenditures do not exceed the budget approved. However, this method can also imply negative consequences. Producers can react to a potential deficit by cutting back the activity level. This response to the potential deficit problem can be very undesirable since the cut-back of the activity level can hurt socially valuable services and thus involve welfare losses. Of course the unexpected increase in activity could be the result of unnecessary services, but the possibility that the increase represents some welfare-enhancing services is indeed present. Another possible response from the producers, in order to avoid the possibility of deficits, could be to secure that their group of patients are low resource-demanding. This means that high resource-demanding patients (e.g. chronical sick persons) can have difficulties of getting treated. An analysis of the consequences of activity based budgets following the DRG system is provided in Torup (1991).

In general, payment methods based on budgeting are well-suited to ensuring certainty that what the third-party financing system intends to spend on health care is actually reached. On the other hand, this implies that the emphasis is on securing that the expenditure does not exceed the obtained budget, which, as mentioned, above can result in undesirable activity cut-backs. Moreover, if the budgets have no direct link with the activity level then this payment method does not provide incentives for high production levels but for producers to obtain budgets by with the smallest possible effort.

The alternative to budgeting is to base the payment method on a fee-for-service principle, that is the producer is paid a charge for each service supplied. The size of the charge will most often differ with respect to the kind of service supplied. A main distinction concerning fee-for-service payment methods is whether the producers are paid in advance or retrospectively. If an advance payment is made it implies that charges are fixed before the service is carried out. Otherwise, with retrospective payment, the charge will correspond to the size of the bill. In relation to the third-party financing system control there is an important difference between a retrospective and advance payment method. The possibilities in the latter method for cost-control can be quite limited; the financing system is simply presented with a bill whereas the advance payment method enables the financing system to influence the size of the charges. Although budgets and fee-for-service payment methods are presented as alternatives, note the correspondence between an activity-based budget and a advance fee-for- service payment method. The main difference is that a fee- for-service payment method implies, in principle, that the producer receives a charge for each service, whereas a payment method based on budgets implies that the producer receives an amount of funds to cover the expenditures related to all services supplied.

The main advantage of a fee-for-service payment method is that such a system induces the producers to provide a high activity level (for the services for which they actually receive fees) and to provide as much as possible of the health care services themselves, that is not to send the patients to other producers. On the other hand, a fee-for-service payment method involves budget uncertainty on the part of the third-party financing system. Unless restrictions/controls are imposed on the number of services provided by the producers this payment method corresponds to a situation where the financing system effectively offers a blank cheque to each producer.

The producers of health care in Denmark receive payments according to different methods. General practitioners are paid according to a mixture of per-capita method and fee-for- service method. They receive a fee per patient attached to the practice and a fee corresponding to the type of service delivered[7]. Specialised practitioners are paid entirely by a fee-for-service method with different charges for different services. Hospitals are paid through politically determined budgets which have a weak and indirect expressed relationship with the activity level. The exceptions are the single state hospital and other large hospitals which, for specialised treatments, receive payments from the user counties corresponding to a fee per patient day. The same holds for the few private hospitals; see Torup (1991). These budgets are negotiated between the hospitals and the county.

4.4 The structure of the Danish health care sector.

Previously, it was noted that the primary source of funding for the health care sector in Denmark was public. Moreover, almost all the hospitals are publicly owned, that is each county owns a number of hospitals at different facility and speciality levels. On the contrary the community health sector (i.e. general practitioners and specialised practitioners) is funded publicly but consists of private producers[8].

This difference in ownership between the hospital sector and the community health sector creates a clear distinction between the two sectors. In fact, the community health sector has a clearly defined role in the Danish health care system, a clarity which is not typical in most countries[9]. Each person is connected to a particular general practitioners, the so-called list system. Normally the general practitioners is the patient's first contact with the health care system. The general practitioner then decides which course of treatment is called for including decisions with respect to whether the treatment can be provided by the general practitioner, by a specialised practitioner or in a hospital. In contrary to the case with the general practitioner where on their own initiative patient can gain contact this possibility does not exist in general with specialised practitioners or hospitals - patients need to be referred by the general practitioner. In this sense, the general practitioner has the role as gateway to the rest of the health care sector in Denmark. It should be noted that this part

[7]The size of the fees are negotiated between the practitioners' organisation and the county organisation.

[8]However, the counties control the number of practitioners so that physicians are not allowed to open a practice.

[9]In Sweden the general practitioners are publicly employed. In the US hospital practitioners are not a direct part of the hospital but they utilise the hospital facilities. See Harris (1977) for an analysis of physicians' behaviour with this relation to the hospital.

of the health care sector (should) represent(s) the relatively cheapest one within the overall health care sector. Therefore this gateway system provides means for cost control within the health care sector. Moreover the general practitioner have incentives to provide as much of the treatment as possible due to the payment system with a mix of per-capita and fee-for-service system. Their income depends positively on the number of services provided. In this way the payment structure supports the general practioners role as gateway to the rest of the health care sector.

In general the structure of the health care sector follows the so-called principle of lowest effective caring and treatment level such that, as mentioned above, the general practitioner is the patient's first contact with the health care system followed by specialised practitioners and hospital departments at low speciality level and, finally, the hospital departments at high speciality level. I will return to the specific problem of the highly specialised hospital departments (the so-called national hospitals and large regional hospitals) in section 4.5.

Before 1993 in some counties general practitioners could transfer their patients to all hospitals in that county. In other counties, this option was not available for the practitioner. However, it was the case in particular for hospitals outside of the patient's home county that there were restrictions in terms of bails, i.e. in order for the general practitioners (or specialised practitioner and hospitals inside the county) to transfer a patient to hospitals outside the home county it was necessary for the home county to be willing to pay for the treatment costs. Now, this system has been changed such that the practitioners can almost without restriction transfer patients to inside as well as outside hospitals except for the case of some highly specialised hospital departments.

The main argument for this reform of the Danish hospital sector was to facilitate treatment for patients in border regions of counties, where the situation can be that an outside hospital is closer than the alternative, inside hospitals. With the reform it is now possible for the patients to be transferred to the closest hospital even if it is placed in a neighbouring county. Another, more dubious, line of argument is that by this reform the health care sector is moved towards a situation where consumer sovereignty will prevail and is moving closer to the standard free competition model. This argument is, however, not very valid. First of all patients are not given the free choice of hospitals but it is the practitioner who choose and for it to be valid in this case it is necessary that the practitioner acts as a perfect agent for the patient, i.e. the practitioner should only be concerned with maximising the patient's utility without taking into account any other goals[10]. In addition, the whole theory of second-best (see Lancaster & Lipsey (1957)) applies in this case: in a situation with several market imperfections and with a non-optimal situation removing one of the imperfections it does not necessarily provide an optimum but can create a situation even farther from the optimum.

The reform can indeed create problems by inducing hospitals to compete on facilities in

[10]See Arrow (1963) for an analysis of the consequences of the delegation of demand from demanders to suppliers with respect to the organisation of health care.

order to attract patients, where it is not desirable that all hospitals have these facilities. Another possible effect could be that the non-observable quality elements of hospitals will be reduced in order to use resources on observable ones, e.g. the previously noted effect on facilities. Finally the reform could also cause huge transportation costs if patients in one county do not choose the neighbouring county hospitals but hospitals in counties further away; see Enemark (1992) for an economic analysis of these problems.

However, there are still some restrictions in the sense that the free transfer possibility concerns hospitals at the same speciality level. Thus if two hospitals (one inside the county the other outside the county) can provide the same treatment at the same speciality level then the choice of hospital is open. Though if an inside county hospital (or outside county hospital) can provide the treatment at a low speciality level relative to a hospital at a higher speciality level then the low speciality hospital should be chosen. Thus the course of treatment for patients is sought to be kept at the lowest effective caring and treatment level.

One problem concerning the relationship between the different health care suppliers is who should provide the treatment. In relation to the lowest effective caring and treatment level there could be cases where treatment is provided more effectively in the community health sector in place of hospitals. Some treatments traditionally provided by hospitals could be provided by other health care suppliers, e.g. general practitioners or specialised practitioners. This issue corresponds to the debate concerning treatment by hospitalisation vs. without hospitalisation. The tendency in Denmark concerning these interactions between the health care suppliers have, as described in section 4.2, been moved towards more out-patient treatment in hospitals, and the hospital sector still provides a significant share of the provision of health care goods. In the following section the structure of the Danish hospital sector is examined, i.e. how the size and type of the hospitals are.

4.5 The structure of the Danish hospital sector.

The Danish hospital sector consists of 17 different hospital systems, i.e. the 14 counties, Copenhagen Municipality and Frederiksberg Municipality and the state. Except for the state which runs a single hospital (Rigshospitalet) the 16 other hospital systems can each be regarded as a multi-hospital system, i.e. common management of several hospitals of different types. According to the hospital classification system from the National Board of Health in Denmark the following groups can be defined with the number of hospitals in each group indicated in parantheses for the year 1990:

1. Large specialised regional- or national hospitals, (6).
2. Other large specialised hospitals, (20).
3. Local hospitals with at least 3 clinical departments[11], (12).
4. Local hospitals with 2 clinical departments, i.e. a department of medicine and a department of surgery in addition to the departments of anaesthesia and x-ray, (28).
5. Local hospitals with one mixed medicine-surgical department, (9).
6. Non-psychiatric specialised hospitals[12], (15).
7. Psychiatric specialised hospitals[13], (16).

Table 2. The distribution of Danish hospitals.

In each county there is at least one hospital from the second group and a number of local hospitals from groups 3-5. The basic hospital type in the county structured hospital sector is this second group which can provide treatment for all the county's patients with the exception of around 10 per cent of the patients which have to be treated in the more specialised hospitals from the first group. The group 3-5 of local hospitals is the most typical hospital type among which the hospital type with 2 clinical departments is the most common. The speciality level decreases from the group of local hospitals with at least 3 departments to the group of local hospitals with a mixed department of medicine and surgery, such that the local hospitals with only 1 clinical department (mixed department of medicine and surgery) can only provide treatment related to the most common diagnoses.

Large specialised regional- or national hospitals and non- psychiatric specialised hospitals are not present in each county since these types of hospitals serve several counties or the whole country. The former provide treatment that cannot be provided at the county level in the other large specialised hospitals (or the local hospitals). Each of these does not normally provide all types of specialised treatments not available at the other large specialised hospitals but only some of them. Therefore what differentiates thems from other large specialised hospitals is that in addition to the treatments available at the other hospitals they also offer more specialised treatment. Moreover, it should be noted, that the group of large specialised regional- or national hospitals and other large specialised hospitals in addition to their more specialised treatment possibilities also function as local hospitals for the local population. Hospitals in the group of non-psychiatric specialised hospitals (6.) are normally small local

[11]The number of clinical departments is determined by the number of departments at the hospital, which have an autonomous management headed by a senior physician.

[12]This hospital group contains a number of different types of hospitals: recovery hospitals, specialised hospitals for treatment of diabetes, orthopedic hospitals etc.

[13]Some of the larger hospitals also have a psychiatric department, but this group contains only those hospitals which treat solely psychiatric patients.

hospitals which have changed from the status of general hospital to only providing specialised treatment within a limited field of diagnosis, e.g. orthopedic hospitals.

This multi-hospital structure enables each county to apply unified procedures with respect to administration and acquisition of supplies and material, to plan the distribution of specialisation of each hospital taking into account the other hospitals in the county and thus to avoid an undesirable duplication of specialities in the county.

A problem with the Danish hospital structure can be how to coordinate the hospital structure between the 16 (17) hospital systems. This coordination problem is present in connection with the large specialised regional- or national hospitals. The functions/specialities available require a high population basis, larger than that of a single county. The coordination problem arises when deciding at which hospitals in which county the specialities should be placed, which also concerns the provision/acquisition of facilities and equipment. The situation to be avoided is that high resource specialities with a need for a high population basis are spread to all counties, since this could involve an inefficient way of spending the resources. In addition spreading such specialities could imply that the personnel at each hospital do not achieve a high level of experience and thus run the risk of providing lower treatment quality. A way to reduce the probability that excessive spreading occurs is to construct the payment system for these specialities such that it gives incentives to the user counties not to establish them in their county. This is not the case with the existing payment method which is based on costs per patient day which include variable and fixed; see Petersen (1990).

In addition and related to the distinction between hospitals by type (as described by table 2) is the distinction according to the size of hospitals, e.g. where size is measured by the number of beds. In table 3 I have distributed all general[14] hospitals to 5 categories according to the number of beds, where the number of beds is based on the stock of beds in the end of the year 1989: Table 3 shows that the most typical size of hospital is one with 100-199 beds and more than half of the Danish general hospitals have less than 200 beds. Therefore the majority of Danish hospitals are rather small, normally small local hospitals with 1 or 2 departments and small non- psychiatric hospitals. Although this indicates a clear decentralisation of the Danish hospital sector it concerns only part of the structure. If the number of beds are distributed according to the 5 bed size categories and the shares calculated in the second column, as appears in table 3, it indicates that hospital production is concentrated at the largest units, since the hospitals with more than 200 beds have circa 80 per cent of total bed capacity. The average number of beds at a general hospital is circa 280. In addition the distribution of hospitals according to bed size is the result of a movement towards centralisation during the preceding years. While the total number of hospitals has been declining from 1980 to 1990 most of the hospitals that have been closed were small one, whereas the closing of larger hospitals have been very few. Therefore the structure of the Danish hospital sector in 1990 is characterised by a higher fraction of

[14]The group of psychiatric hospitals is not included in table 3.

	No. of hospitals	Share of total number of beds
0-99	25	5.49
100-199	29	15.30
200-399	15	17.29
400-599	13	22.89
600 +	10	39.03
Total	92	100.00

Table 3. The distribution of Danish general hospitals according to the number of beds in the end of the year 1989.

Source: National Board of Health (1989)

large hospitals compared with earlier years. Moreover, within the group of small hospitals, structural changes have also appeared such that this group in 1990 includes relatively more small specialised non-psychiatric hospitals and relatively fewer small local hospitals.

This development in the group of small hospitals has taken place mainly by changing the status of the small local hospitals to small specialised non-psychiatric hospitals; instead of closing them such hospitals are retained by providing them with a single speciality. The larger share of small specialised non-psychiatric hospitals within the group of small hospitals is not the result of building new small hospitals but mainly the result of the restructuring of existing small hospitals. Small general hospitals have less importance corresponding to an increased importance of larger hospitals and, to some extent, that of small specialised hospitals. These observations indicate a movement towards both a higher degree of centralisation and a higher degree of specialisation in the Danish hospital sector.

This development is mainly the result of the increased sophistication of treatment techniques involving advanced technical equipment and high levels of specific knowledge. Centralisation of such facilities should imply advantages in terms of expenditure and in terms of the quality of treatment, thus rendering the small general hospitals less attractive relative to the larger ones. Related to this issue is the subdivision of specialities, which makes it difficult for the small local hospitals with a very limited number of departments to keep up with the medicinal/diagnostical development in all special areas. This is true particularly for the local hospitals which only have a single mixed department.

However, small hospitals have also advantages. At small hospitals there should be better possibilities to construct entities where the patients can obtain more individual care and therefore they can feel less uncertain during the period of hospitalization. At the same time the possibilities for the patients to have more constant and close contact with the hospital personnel (especially relevant concerning doctors and nursing personnel) could be expected to be better in smaller hospitals. In this way small hospitals can compensate to some degree for the lower level of sophistication regarding treatment possibilities by providing a more individualised treatment. In addition there could also be advantages with respect to the returns to scale. The hospital size can be expanded to a level where the positive scale effects are exhausted and substituted by negative effects, e.g. the difficulties with running a large and complex entity in order to have consistent decisions in the different segments of the organisation; see Danish Health Institute (1989) for an analysis of small hospitals in Denmark.

4.6 Problems and advantages in the Danish hospital sector.

The structure of the Danish hospital sector described above indicated a number of problems and advantages which will be analysed below. It should be noticed that some of the issues concern both hospitals and other health care producers while others are relevant only for hospitals.

The main advantages of the Danish hospital structure seem to be the following:

1. The uniform financing system.
2. The multiple hospital system.
3. The practitioners as gateway to the rest of the health care system.

These characteristics have provided a satisfactory structure with respect to cost control and which offers the most effective treatment level at the lowest possible costs. These characteristics seem to provide important explanations for the modest growth in health care expenditures during the 1980's. The uniform system of financing where each county is provided with the possibility of obtaining funds by taxing its inhabitants and that hospitals receive resources only through the county budgets puts the county in a relatively strong position with respect to the hospital. The county can be characterised as a monopsonist (on behalf of the population) buyer of hospital services.

This construction could be compared with a system based on a number of buyers of hospital services. For instance, the US health care system has several independent agents which buy services from hospitals: different insurance companies, the federal government, the local government, patients. In such an environment it is much harder to secure cost control because there is no formal cooperation between the different buyers in order to put pressure on the hospitals.

Although this tight link between the county and the hospital is an advantage with respect to cost control it can create the risk of leading to too restrictive policies towards hospitals' resource requests and thus causing a decline in the quality of health care production. I will return to this problem, when analysing the problems within the Danish hospital structure.

The system with each county owning a number of different hospitals makes it possible within a county, to avoid the problems related to a system where the hospitals are functioning as separate units. First this multi-hospital system can be utilised to let some of the hospitals be responsible for some treatment while leaving other types of treatment for others. In this way a multi-hospital system can avoid the presence of duplication of investment and ensure that the personnel responsible for a certain type of treatment, have a higher level of experience. A problem with the Danish hospital structure is, as mentioned above, how the several county-based multi-hospital systems can establish cooperation concerning those treatment, which require a larger population basis than available in single counties. I will return to this issue when discussing the problems within the Danish hospital sector.

Finally, the practitioner's role as gateway to the rest of the health care system has restricted the access to other (and more expensive) parts of health care system and therefore has implied that the course of treatment for a patient can be accomplished at an appropriate level and not involving treatment by a higher level health care producer unless it is viewed necessary in terms of effects on the patient's health status. This element provides an instrument for containing costs. In addition the practitioner has an incentive to provide the treatment because her income is dependent on the number of services she has accomplishes. It should be noted that the practitioner's decision about the transfer of a patient to another segment of the health care system must, not be taken in principle on economic grounds but entirely on those for patient's welfare. For these two reason, this "gateway" system should reduce the use of the other parts of the health care system. However there is a risk that this system can imply an over-optimal use of the practitioners with an under-optimal use of the other parts of the health care system including hospitals.

Overall, the Danish health care sector has appropriate characteristics to restricting resource use. This does not imply that the production of health care services is carried out without inefficiencies and therefore there are reasons for studying efficiency within this sector. Being able to contain costs/resources is not the only important goal for a health care system. Other important goals are to provide treatment and care at a high quality level with high degree of efficiency in terms of positive effects on health status and that the patients have an easy access to the health care system. Below I will analyse some of the problems within the Danish hospital sector.

The main problems of the Danish hospital sector seem to be the following:

1. A too restrictive resource control by the county council.
2. The information advantage of the hospital personnel over the political decision makers regarding resource needs.

3. The potential conflicts with respect to the cooperation between the multiple hospital systems.
4. The combination of financing hospitals by budgets and practitioners by fee-for-service payments.

Actually 1. and 2. conflict to some degree, because if the political decision makers represented by the county councils are able to adhere to a restrictive resource policy then, despite of the hospital personnel's information advantages with respect to what they consider as necessary resource increases, they are not able to obtain more resources due to the restrictive resource control. Thus 1. can be viewed as a qualification of the implications of 2.

The problem of 1. is the possible effects a restrictive resource policy can have with respect to the quality and quantity of hospital production. The reason why 1. can appear as a problem is that the county has a very important role regarding the hospitals: the county is the only unit that supplies the hospitals with resources. Therefore a narrow-minded policy can have severe a effect. A restrictive resource policy can imply a number of effects but will reduce overall the ability of the hospitals to adapt to the demands of the patients, e.g. difficulties with introducing newly developed treatment techniques, queues for planned operations, excessive reductions in the length hospital stay etc. However the exact effects will depend on how the restrictive resource policy is implemented. If the political decision makers have a very detailed resource control of the hospitals then the effects are relatively easy to determine, that is close to the intended political decision. But with only a general control of the total budgets the effects are harder to infer.

It should be added that it is indeed difficult to conclude whether a resource policy is too restrictive or not. Obviously the personnel, patients and other related groups have interests concerning the resource flow to the sector and can point out areas which need additional resources. However these demands will be evaluated together with other resource requests in a political decision making process where the decisions will be taken according to preferences towards the different requests. Therefore it should be emphasized that even health spending with its more essential effects on the demanders than with other types of spending, are limited and must be evaluated. All requests/demands cannot be fulfilled.

The second problem refers to general characteristics of public sector institutions and is not specific to hospitals and refers to the public choice theory; see e.g. Lane (1987) or Mueller (1979). The possible information advantage can be used by the hospital to obtain suboptimal resource levels.

The third of the listed problems in the Danish hospital sector refers to the county based health care provision. This structure is well-suited as long as the activity can take place at the county level. The problems occur when the activity involves higher levels than the county, because several separate counties have to organise a cooperation concerning these activities.

The last of the listed problems considers the problem of using different payment methods for the health care producers: the community health sector, mainly through fees, and the hospitals, mainly through budgets, see e.g. Nielsen (1993). The former method is much more difficult to control than the latter since based on the agreed fees the county can do little when presented with a bill from a practitioner. On the other hand budgets can be controlled and changed by the county. This difference in payment methods should be seen in relation to the overall budget financing of public health care expenditure. In case of budget cuts on the public health care expenditures it is likely that the budget cut will be achieved mainly through cuts in hospital budget since these can be controlled, while leaving the community health sector free. This has indeed been the case in the Eighties, where the restrictive budget policy has been concerned primarily with the hospitals, see Nielsen (1993).

Chapter 5

Defining data of hospital production.

5.1 Introduction.

The analysis of the efficiency measurement methods in the preceding chapters showed that an important issue in relation to the application of these methods is the concern with the provision of data, that is the definition, inclusion and measurement of inputs and outputs in the models. This point is emphasized by Epstein & Henderson (1989) who state that:

"The set of candidate variables tends to be quite large. At the same time DEA makes no a priori distinction between the relative importance of any two outputs or of any two inputs. A variable is either in or out of the model, and all variables that are in have an equal opportunity to influence reported efficiency. The definition and selection of variables to include in the model is therefore critical".

The results from an efficiency analysis will depend crucially on the data included in the model. Therefore the usefulness of efficiency analysis results is determined mainly by the data quality, the quality of the results can never exceed the data quality or otherwise stated "garbage in, garbage out". Moreover in order to analyse technical efficiency measures of inputs and outputs in physical units are necessary, i.e. is measures expressed in money units are not sufficient[1].

This chapter will consider the problems related to obtaining data on inputs and outputs. Since the main problems are related to output measurement I will focus on these problems. In the following section the general problems of obtaining output data are analysed. Then attention is paid to the problems with specific reference to hospital production considering both quantity and quality aspects.

[1]However, as shown by Färe et al. (1988) it is possible to obtain technical efficiency measures on the basis of inputs and/or outputs measured in money units (both revenue and costs) provided that prices are equal and allocative efficiency prevails.

5.2 Common problems in the measurement of outputs.

In general, problems arise when valid outputs are to be defined and measured for sectors with service industries, such as hospitals, where the difficulties are due to several conditions. First of all quantification of the outputs from service producers is difficult, because services are more intangible than normal goods, which is caused by the problem of determining the time in the production consumption process where the service is delivered by the producer to the consumer since services cannot be stored. As a partial consequence it can be difficult to move from direct outputs to final outputs (or effects) since the imperfect output measures cannot be given a price (or value) in a proper way. The difficulties with quantifying outputs can imply that the production will be described by a few, or even a single, indicators, which will be based on those parts of the production which are more visible and measurable excluding the rest of the production. The risk of such indicators being internally heterogeneous can be high, leading to problems when output and efficiency differences have to be explained. Finally, since service producers often have less clearly defined goals the appearance of output measures can induce service producers to pay too much attention to these measures and not to consider the non-measurable outputs, implying that the output measures might not be incentive compatible with optimal behaviour.

In the following all the above mentioned problems will be analysed, where section 5.2.1 will deal with possible procedures to quantify outputs from service producers, section 5.2.2 will describe the homogeneity and aggregation problems and section 5.2.3 will conclude by examining the incentive problems involved with introducing output (productivity) measures for hospitals.

5.2.1 Intermediate output measures vs. final output measures or productivity analysis vs. effectiveness analysis.

Two main groups of output measures are available in relation to performance analyses of a given sector: intermediate (or direct) outputs and final (or effective) outputs, see e.g. Goudriaan et al. (1985). Intermediate output measures are concerned with the direct result of the production process whereas the effective output measures concern the end result of the production process. Examples of intermediate outputs for hospitals could be admissions or patient days and as example of final output measures could be change in health status or mortality rate. In table 1 a number of different output measures are shown (including both intermediate and final output measures).

Below the most important measures from table 1 are briefly considered:

I. Intermediate Measures

- A. Simple Counts
  - Inpatient days
  - Weighted patient days
  - Cases per day
  - Admissions
  - Discharges
  - Square footage
  - Meals served
- B. Labour Based Measures
  - Staff per bed
  - Staff per patient
  - Labour hours per case
  - Number of full time employees
- C. Cost Based Measures
  - Cost per day
  - Cost per case
  - Cost per procedure
- D. Population Based Measures
  - Expenditure per capita
  - Expenditure as % of GNP

II. Utility Based Measures

- Satisfaction with health care
- Desire to expand or change the system
- Number of complaints

III. Final Output Measures

- Life expectancy
- Mortality rate
- Infant mortality
- Severity of illness
- Number of acute conditions (attended and unattended)
- Appropriateness of treatment
- Quality adjusted life years (QALY)
- Monetary value

Table 1. Output measures for hospital production.

- *Discharges*: A discharge is counted when the patient leaves the hospital, that is when the patient is transferred to his home or for additional treatment on another hospital[2,3].
- *Admissions*: Equivalent to discharges except for the point of time at which a patient is counted, (i.e. at the time of arrival).
- *Patient days*: The period a patient has been hospitalised in a given department. In the Danish hospital statistics the day a patient arrives in the department is counted as a patient day, but not the day when a patient leaves.
- *Outpatient visits*: An outpatient visit is one which does not involve an overnight stay. As with discharges a patient can make several visits to the same department during a course of treatment. Danish hospital statistics on outpatient visits have been much criticized because of the very rough classification system, see Andersen (1989). It is thus not possible to distinguish between a given department's outpatient visits.

[2]Notice that a discharge is also counted if a patient is transferred to her home for short periods but returns to the same department afterwards. Thus the number of discharges will be greater than the number of patients since a patient can be readmitted to the hospital several times during one treatment course.

[3]Notice that discharges are counted differently in Denmark where a discharge also is counted when a patient leaves a department for additional treatment at the same hospital but in another department.

- *Quality adjusted life year, QALY*[4]: A comprehensive measure of the effect (outcome) of a hospital treatment, which combines the expected increase in "life expectancy" and the expected improvement of quality of life throughout the patient's lifetime.

- *Patient satisfaction*: Normally related to user surveys, where the patients are asked to evaluate different aspects of the "hospital production process" like the time spent by the doctor, the length of waiting time, the quality of hospital meals etc.

To the extent that it is possible to obtain valid data on final outputs efficiency analyses based on final outputs should be preferred to analyses based on intermediate outputs since what should matter in the performance evaluation is the effects of output rather than the direct outputs, i.e. a given unit can be efficient in the production of intermediate outputs but this is of little use if the production of final outputs is inefficient. This holds particularly for the activity of hospitals (and for other kind of service production) since care provided by hospitals does not increase utility for the user as such , it is the improvement in health status upon which the user's utility is dependent[5]. However the effect of health care on health status is indeed very difficult to detect. The difficulties arise mainly because health care is not the only factor that influences the health status for a patient and therefore it can be a problem to isolate the effect of hospital treatment on a patient's health[6]. These problems of controlling for the influence of other variables on health cover both whether a hospital treatment has effect on a patient health (internal validity) and whether a given effect can be generalised with other conditions (external validity) like another place or another period. Moreover the outcome of health care is multi facet ranging from prolonging life to improving the quality of life (including the absence of pain), see Mcpherson (1990, OECD). In most empirical applications[7] it has been impossible to obtain data on final outputs and therefore to base the analysis on intermediate outputs. The use of intermediate outputs can, however, be problematic since the relationship between these measures and the benefits that the patients receive from the hospital treatment can be rather weak. This can imply imperfect or biased inference concerning the efficiency evaluation of a group of hospitals if the intermediate output measures are used to evaluate the overall performance of the hospitals and not just the efficiency of the the production of intermediate outputs. Therefore the results from such analyses should be interpreted with much care.

A possible method to establish a link between intermediate and final output measures could

[4]See Culyer (1989) or Drummond (1989) for an analysis of QALY in relation to the extra-welfarist approach to health economics.

[5]The utility of health can be said to come in a direct and an indirect way; directly in the sense of well-being and indirectly are the sources of welfare of which a better health enables one to take advantage, Culyer (1989).

[6]Some of these factors are exogenous to the hospital treatment like the social group belonging and the state of national economy, other are endogenous like behavioural, cognitive characteristics,see Tatchell (1983).

[7]See e.g. Pedersen et al. (1987), Cowing et al. (1983), Banker et al. (1986), Valdmanis (1992).

be to include quality related variables along with the quantitative intermediate output measures. In principle this could be managed by increasing the number of intermediate outputs corresponding to the different quality levels, since outputs identical except for different quality can be viewed as different outputs. In practice it can be very difficult to obtain information concerning the quality aspects of outputs due to the subjective character of quality and the requirement of detailed data. One way to obtain such data on intermediate outputs is through user questionnaires for a sample of patients. This corresponds to the patient satisfaction concept. By including patient satisfaction measures the output vector obtained will be somewhere in between intermediate and final outputs. However, as noticed by Thanassoulis et al. (1991), this is only one part of the quality of intermediate hospital outputs. The second part refers to the quality of the outcomes of hospital treatment and therefore demands final output measures, QALY measures (see table 1). If the quality aspect of intermediate outputs can be completely covered then the production description could have included final output measures directly rather than the cumbersome indirect way: inclusion of quantitative intermediate outputs, quality of service and quality of outcome.

5.2.2 The problem of heterogeneous output measures.

In empirical studies of efficiency the problem of heterogeneous output categories often appear. An output unit can be different over time and place. Tatchell (1983) lists how a patient day of care can differ between hospitals due to: "changing technology, varying qualities of care, varying case- mix, varying case complexity or case severity, varying institutional characteristics like size, teaching status, location, composition, ownership and so on". Consider the following example with respect to hospitals. Two hospitals, A and B, are to be evaluated. The output of the hospitals is assumed to be measured by total number of patient days. Hospital A and hospital B have identical yearly expenditures (10 mill. dkr) and have the same number of patient days (1000 patient days per year). However the distribution of patient days according to diagnosis differ between the two hospitals: Hospital A has 100 patient days originating from patients with broken legs and 900 patient days originating from patients with heart diseases, while hospital B has the opposite distribution with 900 patient days from broken legs and 100 patient days from heart diseases. In an efficiency analysis based on output measured as total number of patient days (and inputs measured as total expenditure), hospital A and hospital B will be evaluated as having the same efficiency level. However, a priori, it is expected that treating heart diseases is more resource demanding than treating broken legs. Therefore it is expected hospital A to be more efficient than hospital B since hospital A, with the same resource level as B, achieves the same output level with a more difficult case-mix. The higher efficiency level of A is not measured since the output measure used covers in fact two distinct outputs (patient days related to broken legs and patient days related to heart diseases) which differ with respect to resource demands[8]. Thus

[8]Notice that this appears as a problem only because the outputs have different resource demands and the hospitals have different proportions of the sub-outputs.

one unit of a non- homogeneous output measure will differ from another unit of this measure and, as shown, creates problems with respect to the interpretation of efficiency results.

The example can illustrate why heterogeneous output measures can appear; the problem of heterogeneity is caused by aggregation of dissimilar outputs. Therefore a solution is disaggregate outputs until each output category is homogeneous. The reasons not to disaggregate to such a level are twofold. First it should be noticed that although DEA (or FDH) are designed to handle multiproduct situations there is a limit to the number of outputs (and inputs) that can be included in the analysis. Otherwise the results will be meaningless due to the loss of discrimination power (too many units will receive an efficiency score equal to 1). The second reason is simply that the availability of data can restrict the possibilities for disaggregating output categories. Therefore some aggregation is necessary in order to reduce the number of the outputs, where the important problem is to make the aggregation such that the resulting output(s) are to a large degree internally homogeneous. The following methods allow for output representation by an aggregated number of output categories which aim to be roughly homogeneous, see Barer (1982):

1. Stratification of the observations.
2. Grouping of output categories.
3. Grouping and intra-group weighing of output categories.
4. Grouping and inter-group weighing of output categories.

The first method involves grouping the complete set of hospitals into sets of similar hospitals and then making separate efficiency analyses for each set. Returning to the example it implies that separate efficiency analyses should be made for type A hospitals and for type B hospitals. In this case the way the output categories are constructed is not important since the hospitals are assumed to be similar. However two criticisms of this method can be raised. First, it seems difficult to obtain a set of hospitals with very similar production structures. Even if an analysis compares hospital departments the possibilities of getting a set of departments at different hospitals with an identical output structure can be quite limited. Moreover many interesting comparative results are lost by restricting the analysis to groups of homogeneous hospitals.

The second method implies grouping a larger number of output categories into a smaller number of categories by aggregating those which are similar, where the new groups consist of categories with the same weight. Although appealing the method has shortcomings if used alone since the requirement of DEA to obtain a limited number of production categories might demand the combining of output categories which are quite dissimilar and thereby result in heterogeneous groups. However, if the grouping is combined with weighting the groups or categories (method 3 and 4) to create an even smaller number of categories, the method is relevant.

Method 3 constructs homogeneous output groups by grouping output categories but the output categories within an output group are assigned different weights. The output weights are introduced in order to make the output groups internally homogeneous. Returning to the same example, if a single output measure is applied in the efficiency analysis then method 3 will imply that patient days from broken legs and patient days from heart diseases obtain different weights with the latter receiving a relatively higher weight reflecting the differences in resource requirements. The problem of this method is how to define the weights for each output category. In the absence of market-determined output prices it is not without problems to define reliable output weights. This lack is the norm for publically produced services and hence also for hospital production. Often, as an alternative to price-determined weights, the weights are based on differences in resource need, e.g. construction of unit costs. This requires, however, that it is possible to define and distribute the resources used for each output, which can be very difficult and is further complicated by the presence of integrated production processes[9].

Method 4, as presented by Barer (1982) assumes that the construction of homogeneous output groups takes place without internal weighting but that these constructed groups are weighted such that a single scalar output measure is obtained. The weighting procedure is applied to secure homogeneity, but the same problem of weight definition appears as with method 3. Returning to the example the same interpretation applies as related to method 3, because the example is restricted to include two initial output categories[10]. It is possible to interpret method 4 as the final stage in an output standardisation process in order to obtain a single homogeneous output measure starting by grouping a large number of output categories into a smaller set of homogeneous outputs by weighting the included output categories, then weighting the homogeneous groups to obtain an overall measure.

A problem not considered in the description of the methods of output standardisation is that, often, empirical data is already in a aggregated form which does not secure homogeneity of the outputs. Thus in practice the data sets are not in a form which immediately allows for the use of the 3-stage procedure for output standardisation; heterogeneity can be present in the initial output categories and therefore application of standardisation techniques on such data may seem meaningless. For instance in the empirical applications on hospital efficiency measurement (see chapter 6) in one of the data set for Danish hospitals the original outputs are the number of discharges in 473 diagnosis related groups. These groups are based on the

[9]The extent of integrated and unified production can be expected to exist for hospitals. However Ankjær-Jensen (1990) sketches a procedure for a distribution based on single hospital outputs.

[10]In order to illustrate a difference between method 3 and 4 the example needs to be extended with further subgrouping of patient days, such as broken legs with secondary diagnosis and broken legs without secondary diagnosis.

so-called DRG-system which was constructed in the late Seventies and early Eighties as a hospital output classification system for US hospitals, see section 5.3. and Fetter & Freeman (1985). These groups are already aggregated from more disaggregated diagnosis information. In the empirical analysis it is not possible to take into account possible heterogeneities in the aggregation to the 473 groups. Only once aggregated further is it possible to standardise the output groups with the above described methods. However they can be used to avoid that an eventual further aggregation implies even more heterogeneity.

5.2.3 Hospital output measures and incentives.

An important problem with respect to establishing output measures is how they affect the incentive structures for the producers. In many cases such measures included in productivity measures will be used to reward or at least to judge the performance of the producers by a third-party agent (e.g. public authorities or private insurance companies) and this will influence producer incentives. The essential question in this case is whether incentive compatibility will prevail or not, that is, whether the implied reward structure will provide optimality for both the agent who has established the productivity based reward structure (in the case of hospital production, the public authorities or private insurance companies) and the producers (the hospitals). The reward structure is not incentive compatible if the behaviour of the hospitals leads to a sub-optimal position for the public authorities. This optimality concept can be extended to concern the society as a whole versus the hospitals.

In relation to the hospital output measures described in table 1 such as discharges, patient days or out-patient visits, incentive problems can easily appear. In particular this concerns the intermediate output measures since they focus only on part of hospital production; they are not comprehensive. This implies that producer attention will be put on that part of the production which is included in the measure leaving other, but important, outputs aside. Moreover, if output measures for hospitals are used as part of a productivity measure to judge the performance of the whole health care system then incentive problems can appear if there are interdependencies between the different sub-sectors (e.g. GP's and hospitals). Only if the sub-sectors are not correlated (and this seems unlikely) such analyses will not distort the incentive structures. Otherwise the hospitals (and the other sub-sectors) could have incentives to maximise their own performance evaluation by shifting its burden to other parts of the health care system, e.g. by not admitting resource-demanding patients in order to leave such patients for the other components of the health care system (see Weisbrod (1992)). This behaviour is sub optimal for the society but optimal for the hospitals.

An example of the effects of using intermediate output measures in the evaluation of hospitals is the following. Suppose hospitals are evaluated by the number of patients discharged. If the evaluation is related to benefits for the hospitals then the hospitals have incentives to change behaviour towards maximising the number of discharges through all kind of procedures, e.g. splitting the period of hospitalisation into a small periods, moving patients from outpatient care to in-patient care, moving patients from department to department etc. If the increased number of discharges were due to improved treatment techniques then this would be an advantage for the society, but in the above example the increased number of discharges are caused by the reward structure. At best this situation represents neither an advantage nor a disadvantage, but it could be expected that the society will be disadvantaged since the consequences for the patients are adverse in this very unstable period of hospitalisation. Only the hospitals have an advantage of the increased number of discharges by the link between discharges and hospital benefits. In this case the hospital producers can be characterised as sub-sector optimisers. What has happened is that a discharge before and after the introduction of the reward structure are not identical in terms of quality. Similar responses can be expected with respect to the other intermediate output measures, such as admissions, out- patient visits, surgeries, patient days etc.

The important issue is that this will not happen in the case of a final output measures like QALY, since, in this case, the outputs are quality-adjusted. The problem with final output measures, apart from being difficult to obtain and very costly to measure, is their uncertain validity: does the effect come from the hospital treatment or from other sources.

Utility-based output measures such as patient satisfaction also contain adverse incentive effects if used to judge and reward performance. In order for a hospital to obtain a large proportion of satisfied patients it has incentives to pay attention to activities which are highly visible for the patients and are related to patient satisfaction such as well- organised visiting periods, single bedrooms etc. The activities less visible and less oriented towards patient satisfaction will, although important, receive less attention.

Therefore when hospital output measures are established attention has to be paid to the possible incentive effects for the hospitals.

5.3 Issues related to intermediate hospital output measures.

Given the cumbersome problems related to the provision of valid final output measures for hospital activity I will now turn to issues related to the provision of intermediate output measures since the possibilities for getting output data of this type seem to be more likely. As noted in section 5.2 an important problem of measuring output is to secure homogeneity within each output category, that is where a unit from one output category remains identical over time and between hospitals. Otherwise analyses based on such heterogeneous output data will give biased results and, if used for steering, create incentive problems. In order to reach homogeneity for intermediate output measures two main approaches have been followed (see Tatchell (1983)): service-mix adjustments and case-mix adjustments. These approaches will be described in the following sections (5.3.1 and 5.3.2) with the main focus on the case-mix approach. Related to the problem of heterogeneity of outputs is the problem of choosing the unit of output measurement since the extent of heterogeneity is dependent on this choice. This issue will be analysed in section 5.3.2.2.

5.3.1 Intermediate hospital output measures.

The two approaches, service-mix (see Berry (1967)) and case-mix (see Cowing et al. (1982)), to measuring hospital production by intermediate outputs are different. The service-mix approach has as a basic unit the facility or service available[11][12], while the case-mix approach is based on the hospitals' case load[13]. Moreover there is a difference between the two approaches regarding whether the facilities and services available determine the hospital's case load or vice versa. The service-mix approach assumes that the presence of facilities and services determine the hospital's case mix. On the other hand, the case-mix approach takes the opposite position by assuming that a hospital's case load is determined by the needs of the patients and this case load then determines the facilities and services available at the hospital. In this way the case-mix approach can be characterised as a demand-side analysis of hospital activity, whilst the service-mix approach has a supply-side basis. The service-mix approach has been used more in earlier hospital analyses, while hospital analyses based on the case-mix approach appear later due to the appearance of improved data on the hospitals case load. However the crucial question in relation to the basis for output measures is what characterises output variation between hospitals. If it can be determined that output measures based on service/facility availability provide a valid description of output variation then such measures certainly can be used.

The choice between case-type and service-type measures depends much upon whether case load determines (or is determined by) service availability, i.e. whether demand or supply has the dominating role with respect to the structure of hospital production. This in turn depends on the institutional structure of health care provision. Less regulated health care systems corresponds to greater emphasis on the supply side since establishing facilities and services becomes important in order to attract patients. In this way it is possible to characterise the situation as one where the availability of facilities and services determine the hospital's case load. On the other hand a more regulated health care system should put more

[11] As examples of services/facilities used as output measures the following can be listed: the availability of: operating room, nursing school, clinical laboratory; the provision of: diagnostic x-ray, radium therapy.

[12] The output measures related to the service-mix approach have not only been based on availability of services/facilities (see e.g. Berry (1967)), but also on the number of services performed, see Jenkins (1980). By using the number of services performed the problem that the availability of services does not give an indication of the actual utilisation of those services is solved.

[13] The description of a hospital's case load in relation to output measures is in general based on a number of more or less aggregated diagnosis groups with patient day or discharge as examples of measurement units. In section 5.3.2.2 the problems related to the choice of measurement unit from a case-mix approach are discussed.

importance on constructing hospitals with facilities such that it can satisfy the needs of the population, thereby imposing a relationship between case load and facilities, in which case load determines the availability of facilities. Overall it appears, that using only service or facility variables is more problematic than using only case load variables since some hospital output variation is due to case-mix variation with no relation to service-mix variables. A main reason for this is that facilities services can be used in the treatment of different types of cases. In this way the service outputs can be viewed as inputs for the treatment of some case-types. Only by focusing on the case load is such output variation visible.

However, even if case-type variables provide a better description of hospital production than service-type variables, some aspects of the variation in hospitals output due to service outputs are not taken into account if only case-type variables are used. For instance this can be the case with out-patient visits, which should be a proxy for the outputs from the hospital's out- patient department. If only case-type variables (such as discharges or patient days) are included, which concern the in-patient hospital activity, then an important part of hospital production is excluded from the analysis (dependent on the extent of correlation between in-patient and out-patient activity). This point indicates that a more thorough description of hospital production would have to include service/facility variables along with case-type variables and thus take a more mixed position in the choice of intermediate output measures. In this way both the supply-side and the demand-side are assumed to influence the hospital production, which seems to provide a more realistic description of the determinants of their activities hospitals.

Although some hospital efficiency studies have focussed solely on case-type variables as output measures, see e.g. Banker et al. (1986), the majority of studies have included both case-type and service-type variables, see e.g. Fāre et al. (1989)[14] and Morey et al. (1990)[15]. The focus on only case-type outputs is therefore not so present in efficiency measurement analyses but in the estimation of hospital cost functions, see e.g. Cowing et al. (1982) and for criticisms of this Jenkins (1980) and Breyer (1987). Most often the case-type variables are used regarding to the in-patient care, while service-type variables are used for out-patient care and other activities indirectly related to the in-patient care, e.g. to include the outputs from the hospitals teaching department. In the empirical applications in chapter 6 both types of variables will be included in order to obtain a more realistic description of hospital production.

In the proceeding section I will examine problems related to the use of case-type variables

[14]Fāre et al. (1989) analyse efficiency within a sample of Michigan hospitals with at least 200 beds based on the following outputs: acute care patient days, intensive care patient days, inpatient surgeries and outpatient surgeries, and outpatient visits.

[15]Morey et al. (1990) examine public and not-for-profit Californian hospitals with at least 200 beds included the following output variables: acute patient days, intensive patient days, inpatient and outpatient surgeries, outpatient visits, residents per attending physician.

not due to the fact that only case-type variables should be included as outputs, but because in-patient care constitutes an important part of hospitals activity and because the problems related to establishing case-type variables have been discussed widely in recent years. In particular the diagnosis-related groups system, the DRG system, will be discussed in relation to its provision of output measures for hospitals.

5.3.2 Output measures based on patient classification systems.

Most of the case-related output measures for hospital production are based on some kind of patient classification system. The general principle behind these systems is to distribute the hospital's patients over a number of different groups according to some predefined criteria, often the patient's diagnosis[16]. The description of the production from hospitals is then provided by the number of cases[17] in the different categories. These patient classification systems are derived from the patient classification system established by the WHO, International Classification of Diseases (ICD). This system contains more than 1000 disease groups, but these disease groups can be combined to form 17 broadly defined heterogeneous groups. The problem with the direct application of the ICD patient classification system is that the number of categories is too large to be manageable. Many categories will be empty or include a very small number of patients and therefore have little relevance in terms of information. In any case for efficiency measurement methods such as DEA and FDH the number of groups is too large. Below I will analyse in detail the so-called Diagnosis Related Groups system (DRG), which has been discussed intensively in recent years.

5.3.2.1. Diagnosis Related Groups system, DRG.

The Diagnosis Related Groups system is based on the ICD system, where the diagnosis groups are constructed such that the groups are homogeneous, clinically meaningful and mutually exclusive with a relatively small number of groups. It was developed in United States in the Seventies and early Eighties in order to obtain a description of the hospital production structure whitch was related to the use of resources, see Tatchell (1983). The relation in the DRG between production and resources is included by determining an average unit cost for each group. It is this element in the DRG-system that, in the US, has been used to define fixed predetermined fees for hospital treatment as means to controlling health care spendings (I will return briefly to this application of the DRG-system). The groups are defined using hospitals in New Jersey, Connecticut and South Carolina. In the construction of the groups the discharges from the patient data material are partitioned according to primary diagnosis into 23 main diagnosis groups, which are anatomically or clinically oriented. In table 2 these 23 main diagnosis groups are listed.

[16]Other possible criteria could be age, case severity or case complexity.

[17]For instance the unit of measurement can be patient days or discharges. This issue is discussed in section 5.3.2.2.

1 - DISEASES AND DISORDERS OF THE NERVOUS SYSTEM
2 - DISEASES AND DISORDERS OF THE EYE
3 - DISEASES AND DISORDERS OF THE EAR, NOSE AND THROAT
4 - DISEASES AND DISORDERS OF THE RESPIRATORY SYSTEM
5 - DISEASES AND DISORDERS OF THE CIRCULATORY SYSTEM
6 - DISEASES AND DISORDERS OF THE DIGESTIVE SYSTEM
7 - DISEASES AND DISORDERS OF THE HEPATOBILITY SYSTEM AND PANCREAS
8 - DISEASES AND DISORDERS OF THE MUSCULISKELETAL SYSTEM AND CONNECTIVE TISSUE
9 - DISEASES AND DISORDERS OF THE SKIN, SUBCUTANEOUS TISSUE AND BREAST
10 - ENDOCRINE, NUTRITIONAL AND METABOLIC DISEASES AND DISORDERS
11 - DISEASES AND DISORDERS OF THE KIDNEY AND URINARY TRACT
12 - DISEASES AND DISORDERS OF THE MALE REPRODUCTIVE SYSTEM
13 - DISEASES AND DISORDERS OF THE FEMALE REPRODUCTIVE SYSTEM
14 - PREGNANCY, CHILDBIRTH AND THE PUERPERIUM
15 - NEWBORNS AND OTHER NEONATES WITH CONDITIONS ORIGINATING IN THE PERINATAL PERIOD
16 - DISEASES AND DISORDERS OF BLOOD AND BLOOD FORMING ORGANS AND IMMUNOLOGICAL DISORDERS
17 - MYELOPROLIFERATIVE DISEASES AND DISORDERS, AND POORLY DIFFERENTIATED NEOPLASMS
18 - INFECTIOUS AND PARASITIC DISEASES
19 - MENTAL DISEASES AND DISORDERS
20 - SUBSTANCE USE AND SUBSTANCE INDUCED ORGANIC MENTAL DISORDERS
21 - INJURIES, POISSONINGS AND TOXIC EFFECTS OF DRUGS
22 - BURNS
23 - FACTORS INFLUENCING HEALTH STATUS AND OTHER CONTACTS WITH HEALTH SERVICES

Table 2. 23 main diagnosis groups.

Source: Health Systems International (1989).

The discharges within these 23 main diagnoses are partitioned with respect to whether the patient has been operated upon or not, i.e. a partitioning into surgical and medicine groups. The further subgroup construction is provided by including the patient's age as well as secondary diagnoses. These subgroups are constructed such that they are homogeneous to a high degree with respect to resource use. Resource use is measured in the DRG system by the average length of stay and thus a high degree of homogeneity within a DRG group is achieved when the average length of stay has a low variation. However, the average length of stay does not appear directly in the subgroup definition, rather it is used to determine those patient-related variables (age and secondary diagnosis) that minimise the variation in the average length of stay. In this way 473 DRG groups are defined.

Notice that this grouping is fixed and determined on the basis of patient data from 1975 for hospitals in 3 US states. When the DRG system is applied to analyse the production structure for a sample of hospitals the basis is these defined DRG groups where the patients are distributed accordingly, giving information about the hospital's case-mix. A comprehensive measure of a hospital's production can be obtained in principle by using the above mentioned unit costs[18] defined for each DRG group by weighting the number of patients in each group with its corresponding unit cost and summing over the DRG groups. Similarly an over-all case-mix measure for each hospital can be obtained by: (1) weighting each DRG group's discharges by its unit cost and summing over the DRG-groups, (2) dividing this weighted number of discharges with the un-weighted number of discharges (this is the same as assuming that discharges from different diagnoses have identical resource use). The ratio of the number of weighted to un-weighted discharges represents the case-mix measure, where a high value indicates that the hospital has a relatively high resource-demanding case-mix, while a low value indicates the contrary. Moreover the weights are essential for the application of the DRG system in efficiency measurement studies since 473 groups represents too large a number of groups to obtain meaningful results. Although it is not necessary to aggregate to a scalar output measure this aspect puts importance on the validity of the DRG weights[19]. Another possibility could be to use total un-weighted discharges and then include the case-mix measure as an additional output in the efficiency measurement analysis.

Although the DRG system is relevant for the analysis of hospital production structures it is not without disadvantages. One type of criticism is related to the general aspects of the DRG system (see e.g. Hatting (1989) or Pedersen (1988)), other studies have emphasized the problem of transferring the DRG system to other countries (see Bak (1989)) and, finally, the use of DRG to control and contain hospital expenditure has been criticized, see Weisbrod (1992), Blomquist (1992) and Torup (1991). Below I will look more closely at the various criticisms of the DRG-system.

A frequently mentioned problem of the DRG is that it does not include out-patient treatment, long term care or the activity in psychiatric hospitals or departments. In order to base efficiency analyses on DRG it is necessary therefore to add separate indicators for these areas of hospital production corresponding to a mixed approach in the choice between inclusion of service-type variables and case-type variables. Doubt has also been cast on the homogeneity property of the DRG groups in particular when other countries apply the US-based DRG system; homogeneity in US does not imply homogeneity in other countries (Hatting (1989)). In relation to this problem it has been examined whether the DRG group structure could be reestablished by using patient data from other periods or other places, but in general the group structures were not identical (see Pedersen (1988)). Moreover it has been criticised that the homogeneity criteria with respect to resource use is defined in terms of the variation in length of stay of the discharged patient in a DRG group. Factors other than resource use have an influence on the length of stay, e.g. how the coordination between the hospital and the rest of the health care sector is functioning. Homogeneity problems also appear due to an imperfect allowance for case severity, which can take forms other than those represented

[18]These unit costs are calculated on the basis of data from US hospitals. However attempts have been made in other countries (e.g. Norway) to establish country specific weights.

[19]I will return to this issue in the following paragraph, where the disadvantages and limits of applying the DRG system are discussed.

by the DRG group definition and the presence of secondary diagnoses. Case severity also depends on the general health status of the patient, the patient's history of treatment. In addition the homogeneity problems can be accentuated by 2 properties of the DRG system: the final choice of DRG group is very dependent on the initial choice between the 23 main diagnoses. If there are several competing diagnoses the actual DRG group can be difficult to forecast. Second the order in which the diagnoses are listed influences the DRG group chosen. This property introduces variation in the DRG group chosen for discharges that are very similar. The final criticism of this system is the missing omission of the output quality in the construction of the patient classification system. In this way an important aspect of the hospital's production is left aside, since quality measures could provide the link between intermediate and the final outputs - the outcome of hospital treatment as stated in section 5.2.

As mentioned previously, the DRG system has, in addition to its use as a patient classification system, been applied as an instrument for hospital cost control in the United States. In the beginning of the Eighties a new system for hospital payments was introduced, where the hospitals before were reimbursed retrospectively the new payment method was in advance in the sense that the hospital's fees were fixed before the treatment. The fee that a hospital receives for treating a given patient depends on the DRG group to which the patient is classified and can be derived from the cost weights attached to the DRG group. The idea behind this construction was to introduce incentives for cost-savings in the hospitals. If a hospital is able to provide the treatment connected to a given DRG group with less resources than the fixed fee it will receive the difference as a revenue. On the other hand, if the hospital needs more resources for a given treatment in a DRG group than the fixed fee it has to cover the difference. This should induce hospitals to attempt to search for less resource-demanding ways of providing the treatment related to a given DRG group. However a number of possible reactions from the hospitals can prevent the control of health care spendings. One very likely hospital reaction is to respond to the incentives for resource reduction by reducing the average length of hospital stay[20] or other steps that reduce the number and level of services performed for each patient. This response will shift the treatment costs from the hospitals to other parts of the health care system, e.g. providers of home-care services and nursing homes, see Weisbrod (1992). This response is only optimal for the hospitals, but not optimal for the health care sector as a whole. It should be noted that the incentives imposed by the DRG-based advance payment method can induce the search for more efficient shifting from new forms of treatment to more efficient methods of existing ones, see Blomquist (1992)[21]. Another possible reaction to the use of a DRG based payment method is to attempt to classify patients in those DRG groups which have a large fee and where the hospital have

[20]Note that the DRG-based fees are related to discharges and not days of care. That is the revenue from treating a patient does not increase if the hospital stay is prolonged.

[21]See Weisbrod (1991) and Newhouse (1992) for analyses of the relationship between the technological innovations in the health care industry and the payment method for the health care providers.

cost advantages, the so- called DRG creep, see Tatchell (1983). This also occurs when hospital shift treatment from in-patient care to out-patient care. They have an incentive to do this since outpatient care is not included in the DRG-based advance payment method, therefore the hospital will shift from fixed fee treatment to that where the fees are adjustable.

5.3.2.2. The unit of measurement for hospital outputs.

The question of choosing the unit of measurement for hospital production given the focus on intermediate outputs is important and has been much debated (see e.g. Feldstein (1967), Lave & Lave (1970), Evans (1971), Frank (1988), The Danish Home Office (1986) and Magnussen (1992)). This importance is caused mainly by the influence the choice of unit of measurement can have on the efficiency variation, that is the relative size of efficiency measures are dependent on which unit is used to measure the hospitals outputs.

In general the choice of the unit of measurement is between days of care and cases (admissions, discharges, separations[22]). A priori, it is not possible to decide which measure is superior in the description of hospitals outputs, the decision has to be based on which measure is the most homogenous in the actual application between hospitals. Frank (1988) seeks to close the debate between using cases or days in favour of cases by arguing that the cost per day might behave in a peculiar way, e.g. if a hospital allocates resources in order to decrease the average length of stay the cost per day will increase (in contrast to cost per case where this pattern is not present). However, this pattern can be related to the possibility that those days that can be removed are the days in the last part of a patient's stay which are the cheapest ones. However both cases and days are indeed imperfect measures of hospital production and both raise problems. In fact a hospital can be considered as producing both cases and days, without a case no day of care can be produced and a case demands days of care.

The problems related to using case as the output measure are mainly that the magnitude of hospital production will depend on how the course of patient treatment is organised. For instance, if a hospital sends a patient home in the weekend this will be registered as a case and the patient will appear as a new case on the following Monday. In fact no additional production has taken place, but the hospital will show up with a higher production level compared with a hospital without this practice. The same situation holds for both transferring a patient from one department to another within the same hospital and for other types of breaking up the course of treatment for a patient. To the extent that such procedures differ between hospitals they will induce different (and erroneous) efficiency evaluations of the hospitals[23]. In this case it will be more appropriate to use complete patient treatment

[22] The total number of separations is equal to the number of discharges plus the number of deceased.

[23] In The Danish Home Office (1986) this problem is the reason for using patient days rather than discharges as the output measure.

period as the basic unit, but this measure can be very difficult to obtain. In addition, the unit case might not be homogeneous between hospitals.

However, the conclusion is however not to use days of care as the basic unit since, as noted above, using the number of days of care is also problematic. Normally an increase in output given a certain input level will be interpreted as an indication of productivity (efficiency) increase, but this might not be the case with the number of days of care. An increase in a hospitals number of days of care can be the result of slower treatment procedures[24]. Similarly a decrease in a hospital's number of days of care can be the result of the better organisation of patient treatment[25] and thus not reflect a decrease in productivity. Part of the problem with days of care as the output measure unit is the composition of the number of days of care. The number of days of care is a combination of the number of cases and the time each case is hospitalised, that is the number of days of care is equal to the sum of days each case is hospitalised or is equivalent to the number of cases times the average length of hospital stay.

The problem relates to the inclusion of the time component in output measures. On one hand, it can be argued that the time component takes into account differences between cases in terms of complexity as well as other types of case differences. The implicit assumption is that more complex cases take more time and therefore represent an output. However, complex cases do not necessarily appear as cases with a long length of hospital stay. Moreover as pointed out above the time component can cover up inefficiencies[26] in the organisation of patient treatment.

Therefore, in relation to efficiency analyses it seems more relevant to base the output measure on cases rather than on days of care. There can be practical problems related to the application of cases which, as mentioned, depend on the case definition. An output unit which is less dependent on case definition is a completed course of treatment. However, as noted in the beginning of this section the choice of output unit is very ad-hoc and has to be decided for each application.

[24]This could be caused by a retrospective payment structure which pays the hospital per day of care; in this case the hospital has an incentive to increase the number of days of care. In section 5.2 the issue of the relationship between hospital output measures and incentives is analysed in detail.

[25]However, a decrease in the number of days of care can also be the result of incentives that are present for the hospital. For instance, if the hospital is paid according to the number of cases then it will have an incentive to discharge the patients more quickly but before they are fully recovered.

[26]See Magnussen (1992).

Chapter 6

Empirical applications of efficiency measurement models to hospitals.

6.1 Introduction.

This chapter includes different types of empirical applications of the DEA and the FDH methods to different hospital data sets. The aim of these empirical applications is partly to illustrate the functioning of the theoretical models and partly to show the kind of information that can be obtained through the use of these methods. In addition, information concerning efficiency structures within different hospital samples is provided, with possibilities for testing hypotheses on hospital performance. The last point should be qualified slightly since the input-output data, although at the hospital level, are still in a highly aggregated form. In order to have more relevance for policy implementation it is necessary to obtain more disaggregated data, e.g. at a departmental level. However the analyses show that it is possible to apply the efficiency measurement methods to hospitals and the information resulting from these analyses can give an indication of which aspects to examine with more disaggregated data.

The rest of the chapter is disposed as follows. Section 6.2 includes an efficiency analysis of 80 Danish hospitals where case-mix differences between hospitals are ignored. This is followed up in section 6.3-6.6 where a sample of Danish hospitals with case-mix dependent output measures is used to analyse the extent of efficiency. In section 6.3 the effects on efficiency with and without case-mix adjustment is analysed. Section 6.4 uses the efficiency measurement models to analyse to what extent Danish hospitals are producing at an optimal scale. In section 6.5 the efficiency results from 6.3 are compared with efficiency measures from a parametric frontier model in order to examine whether the two sets of results are similar. Section 6.6 includes an analysis of the capacity utilisation in the hospitals based on the efficiency measurement models. Section 6.7 contains an efficiency analysis of a sample of private British hospitals.

6.2 An efficiency analysis of 80 Danish hospitals without case-mix adjustment

6.2.1 Introduction

This section deals with an empirical analysis of the Danish hospital sector, hopefully illustrating some of the mechanisms behind the theoretical framework put forward in chapter 3. Hence, the primary aim is not to make a thorough and complete analysis of the efficiency of the Danish health care sector but merely to illustrate some of the difficulties involved in efficiency evaluation with Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH) methods. In doing this I also show the extent to which the efficiency information provided by DEA and FDH methods is usable and sufficient in relation to overall performance evaluation of production units.

My approach is the following. First, I specify a standard model which, at first sight, seems to be appropriate with respect to hospital activities. Obviously this is only a crude model and hence alternative specifications are presented all of which are related to the standard model. Through looking at these various models I obtain an indication of robustness, that is of possible factors which may effect the efficiency variation in the standard model. Moreover, I analyse this aspect in a parametric way i.e. by a regression approach. The variables from the various models which proved to be of some importance with respect to the efficiency variation all enter as explanatory variables in one form or another. In this way a statistical basis for testing the significance of the included variables is obtained. Hence there is an interface between the non-parametric and the parametric approach to efficiency evaluation.

The Danish hospital sector has been chosen for two main reasons. First, because this area is characterised by political attention partly due to the fact that it constitutes a large fraction of total public expenditure at the regional or county level. Second, because the activity in the hospital sector is one of the few public areas well covered by highly disaggregated production statistics at a micro level. In Denmark, such statistics are made available through the annual publications "Virksomheden ved sygehuse" (Statistics on Hospital Activity) and "Personale- og Økonomistatistik for Sygehusvæsnet" (Statistics on Hospital Employment and Expenditures) from the Danish Ministry of Health.

In general, the health care sector has been a rather popular area for applied studies of efficiency measurement. Table 1 offers a survey of earlier DEA studies of health care activity.

The rest of this section is organised as follows. Section 6.2.2 briefly describes the main characteristics of the Danish hospital sector. In section 6.2.3 the hospital data set is described; the source and the selection of data. The results of the application of FDH and DEA methods on the selected hospital data will be examined in section 6.2.4. A range of results from models with different variables are interpreted with respect to a standard model. Section 6.2.5 attempts to put the information obtained from these different models together by regressing the efficiency scores from the standard model on variables reflecting the different models. Section 6.2.6 concludes with final remarks.

Author	Number of units	Type and period of data	Number of outputs/ inputs
1) Banker, Conrad & Strauss (1986)	114	Cross-section of North Carolina hospitals 1978	3 outputs 4 inputs
2) Bogetoft, Olesen & Petersen (1987)	96	Cross-section of Danish hospitals 1983	6 outputs 1 input
3) Bruning & Register (1989)	1254	US hospitals 1985	6 outputs 5 inputs
4) Fare, Grosskopf, Lindgren & Ross (1989)	17	Paneldata of Swedish hospitals 1970-85	3 outputs 4 inputs
5) Grosskopf & Valdmanis (1987)	22 public 60 private	Cross-section of Californian hospitals 1982	4 outputs 4 inputs
6) Sherman (1984)	7	Cross-section of Massachusetts university hospitals 1976	4 outputs 3 inputs

Table 1. A survey of earlier studies of productive efficiency for hospitals.

Remark: DEA-C, DEA-D and DEA-V denote, respectively, a DEA model with constant, decreasing and variable returns to scale. COLS (Corrected Ordinary Least Squares) is a parametric efficiency measurement method, where the production frontier is estimated in two steps: First the parameters are estimated by OLS, secondly the intercept is shifted up until all residuals are non-positive. A * indicates that information on this category was not available.

Author	Types of outputs/ inputs	Methods	Inefficient units	Units pr. category
1) Banker, Conrad & Strauss (1986)	Patient days/ Working-hours, beds	DEA-C DEA-V COLS	77 (67.5%) 69 (60.5%)	16.30
2) Bogetoft, Olesen & Petersen (1987)	Patient days, emergency visits/ Net expenditures	DEA-C DEA-D DEA-V	78 (81.3%) 69 (71.9%) 65 (67.7%)	13.70
3) Bruning & Register (1989)	Patient days/ Physicians, nurses, other personnel and beds	DEA-C	1128 (90.0%)	114
4) Fare, Grosskopf, Lindgren & Ross (1989)	Discharges, patient days, emergency visits/ expenditures	DEA-C (Malmquist)	14 (82.4%)	2.42
5) Grosskopf & Valdmanis (1987)	Patient days, surgeries, outpatient visits/ Physicians, non-physicians, netplant assets and admissions	DEA-C DEA-V	* *	10.25
6) Sherman (1984)	Patient days, students/ Full-time employed doctors, available beds, expenditures	DEA	2 (28.6%)	1.00

Table 1 (continued).

6.2.2 A brief description of the Danish hospital sector

Almost all of the 100 or so hospitals in Denmark are in the public sector. A few hospitals (6) are organised privately but are publicly financed and a single hospital is purely private. Public expenditure on hospitals accounted for 6.8 per cent of total public spending in 1990.

The Danish health care system is not organised as a national health service. However there are 14 county councils (the regional public authorities) which by law are responsible for health care delivery within their geographical boundaries (the system in Copenhagen is different since the municipalities in that region in combination with the only existing state hospital (Rigshospitalet) are responsible for the provision of health care). In general, the structure of the hospitals at the county level consists of one large specialised regional- or national hospital and a number of smaller local hospitals with a maximum of three departments. The almost total absence of direct consumer charges constitutes a further characteristic of the provision of health care within the hospitals. The county system is based on the possibilities of each council imposing a proportional income tax on its residents, and budgets for the different hospitals are negotiated and allocated in advance on a one year basis by the county council administration. These budgets are the result of a political decision process. If a hospital spends more than the budget allows the management concerned is criticised by the higher level authority when there is no special reason for the overspending, but normally no further sanctions are imposed on the hospital. Surpluses at the end of the budget period are, as a rule, returned to the county council (however in recent years some possibilities for transfering money from one fiscal year to another have been allowed). This procedure for hospital resource allocation indicates that there is indeed a need for control of hospital activities.

6.2.3 The data

Data for the present study of productive efficiency for Danish hospitals are based on hospital statistics published yearly by the Danish Ministry of Health. The available statistics from this central source consist of records of the activity of individual hospitals as well as employment and budget information on a hospital level.

The activity statistics are provided from a central patient data base updated every year under the Ministry of Health to which each hospital is obliged to give information concerning every discharge. Therefore for every discharge the hospital, and department from which the discharge originates, is registered as well as the patient-identification, date of arrival and date of leaving the hospital, the course of treatment, diagnoses, operations etc. From the discharge records it is possible to construct measures indicating the level of activity for each hospital such as the number of discharges, the number of patient days etc. These measures are only indications of the total production from each hospital, but they contain basic information about the demand for resources arising from the demand for hospital services that are satisfied (not considering rationing).

6.2.3.1 A description of data

The above-mentioned statistics from the Danish Ministry of Health contain data on the number of full-time employed personnel divided into 57 job-categories. Furthermore, data indicating the activity level i.e. the number of discharges, the number of out-patient visits, the number of patient days and the number of beds (the latter indicates capacity rather than activity) exist. These activity data are divided into emergency versus non-emergency cases on a departmental level depending on the medical condition of the patient. The total number of e.g. discharges for each hospital is found by aggregating over all hospital departments. These activity statistics are purely quantitative and therefore completely ignore the quality dimension. Moreover, statistics on total expenditure consist of expenditure on wages, goods, services and materials and finally hospital earnings arising from transactions between county funds. These groups are aggregates made by the Ministry and from more disaggregated information supplied by the hospitals.

6.2.3.2 The choice of data set

In this study I use a reduced set of the data supplied by the Ministry of Health. The main reason for using a smaller data set is that the methods I intend to use for analysing productive efficiency (DEA and FDH methods) require a small number of inputs and outputs compared with the number of observations. Otherwise a large part of the observations will become non-comparable and thus will be classified as efficient, making the whole exercise meaningless.

The reduction of the data set takes place at two levels. One level concerns the construction of the aggregates to obtain categories of inputs and outputs representing the activity of each hospital. The other level concerns the choice of hospital sample where the included hospitals have to be similar given the aggregated information on activities.

The employment data on 57 job-categories offer a good base for the aggregation of inputs since all categories are measured by the same units, i.e the number of individuals. Notice that in aggregating the data one implicitly assumes internal homogeneity of the aggregated groups, e.g. that all the personnel within a particular group have the same productivity. For the standard model I have chosen to measure the personnel, or input, by aggregating the job-categories into the following 4 groups (capital letters represent the name of the variable in the models):

1. Doctors (DOCTOR)
2. Nurses (NURSE)
3. Other types of health care personnel (OCARE)

4. A residual group (OTHER).

The first group (DOCTOR) contains the number of doctors and other types of academic health care personnel e.g. dentists etc. NURSE is an aggregate of nurses and other types of nursing personnel. OCARE contains other types of non-academic health care personnel. Finally, OTHER includes administrative personnel, cleaning personnel etc. These four categories of employment do imply that some allowance for differences in quality between the employment categories is considered. However the four employment categories are mainly constructed in order to indicate differences in the way each group interacts in the production process rather than to indicate differences in quality. Realising that these four categories may seem incomplete since data on capital as well as consumption of goods are excluded I will operate with alternative models. As a proxy for capital earlier studies of hospital efficiency (e.g. Banker, Conrad & Strauss (1986) and Sherman (1984)) have used the total number of beds. This is an incomplete measure for capital, however following this tradition I include the variable BED (see model C2 in section 6.2.4.1). Moreover, another alternative input measure may be the total current net expenditure[1] (in this study called EXP) as an aggregated variable which takes into account the consumption of all goods (including labour).

The outputs in this study are chosen as aggregates from the activity statistics. For the standard model the outputs are represented by the total number of discharges and the total number of outpatient visits[2]. I aggregate over emergency cases and non-emergency cases for discharges and outpatient visits. Implicitly this kind of aggregation assumes that emergency cases and non-emergency cases are homogeneous regarding the amount of required resources. That is:

a. Total number of discharges (DISCH)

b. Total number of outpatient visits (AMBULA)

This basic model is incomplete in the sense that the above partitioning does not consider the fact that patients are different with respect to length of stay. Therefore I will operate an alternative model using the total number of patient days (PDAY) instead of DISCH (see model B1 in section 6.2.4.1).

In 1989 the total number of hospitals included in the statistics from the Ministry of Health is 111. However, not all of these 111 hospitals are comparable in terms of the above stated variables. Psychiatric specialised hospitals and physiotherapeutic hospitals

[1]Total current net expenditure is defined as total current expenditure minus earnings obtained from patients from foreign counties. The reason for not using total current expenditure is that the activity data concern patients from the county in which the hospital is placed. Therefore total current expenditure overestimates the costs of these patients since they also involve costs incurred on patients outside the county.

[2]Notice that in Denmark there is no official system such as Diagnosis Related Groups (DRG) for aggregating outputs.

must be excluded due to their highly specialised activities and this concerns some of the somatic hospitals as well. This reduces immediately the number of hospitals to 80 which form the basic sample.

Table 2 shows some descriptive statistics of this sample of 80 hospitals. Notice that the mean and the median of each category is almost identical indicating that the distributions are symmetric. In the sequel I will only consider data from 1989.

	DOCTOR	NURSE	OCARE	OTHER	DISCH	AMBULA
Max.	17.78	60.89	27.27	41.54	64.50	96.76
Min.	4.11	24.00	3.90	16.62	3.24	35.5
Mean	10.22	49.15	12.91	27.72	30.12	69.88
Std.	2.30	6.00	3.38	4.28	11.29	11.29
Med.	10.28	49.33	12.54	27.54	29.59	70.14

Table 2. Descriptive statistics of the distribution of inputs and outputs in the standard model, per cent.

Note: The numbers are based on data from 85-89, e.g. the mean is calculated over the whole range from 85-89.

6.2.4 The choice of modelling approach

The major disadvantage when turning from parametric approaches towards non-parametric ones such as DEA is the lack of foundation for statistical analysis. Hence recent developments in applied DEA point towards the introduction of statistical methods in the form of statistical tests, in order to reestablish the robustness of the results obtained. In particular there has been a search for the "true model" describing the observed production relation. One such way to obtain a "true model" is considered through the so-called stepwise DEA (Kittelsen (1992)). The idea of this procedure is to extend a basic model with a number of new variables included on the basis of an F-test for the relative difference in average efficiency. The inclusion of new variables stops when these variables become insignificant. However, in its present form, this procedure seems to have some drawbacks. Firstly, one has to assume that the efficiency distribution is half-normal or exponential which, as such, are rather strong assumptions. Secondly, and more importantly, the way the F-statistic is defined seems unfortunate in relation to the way in which it is used. According to the F-test, a variable is significant when average efficiency increases due to the inclusion of the variable. In the worst case this implies that irrelevant variables may be included if they result in the non-comparability of the units since non-comparability means high efficiency scores i.e. a higher F-value. Furthermore the inclusion of new variables depends crucially on the variables in the basic model. If these variables have been wrongly chosen the final model will also be erroneous.

Also the non-parametric Mann-Whitney test[3] have been used, though not in order to determine the "true model". Briefly, the Mann-Whitney test is used to analyse whether two subgroups can be assumed to have been drawn from the same population. Consider the following example. Through a criteria of geographical location the hospital sample can be separated into two groups. If these groups fail to pass the Mann-Whitney test then they have not been drawn from the same population i.e. they differ in the distribution of efficiency scores due to geographical location. Recently Valdmanis (1992) and Magnussen (1992) have applied the Mann-Whitney test to hospital data from USA and Norway respectively. In both papers the overall idea is to analyse the sensitivity of the DEA efficiency scores obtained with respect to different criteria.

As part of a larger framework I intend to apply yet another kind of test, the Spearman rank-order correlation coefficient,[4] testing the degree to which two rankings are associated. In the proceeding sections I will follow the approach outlined below:

1. A standard model is defined which, at first hand, seems representable.

2. Considering the results of the standard model a number of alternative models are used concerning e.g. the aggregation of some variables, the inclusion of new variables or the replacement of old ones etc.

3. The association of the obtained rankings are tested through Spearman's correlation coefficient.

4. In order to explain the efficiency variation I have chosen to follow a regression approach where the efficiency scores from the standard model are regressed on a vector of explanatory variables which include environmental factors. This kind of analysis can be seen as a parallel to the above mentioned Mann-Whitney test and as such it may have policy implications.

The focus is on input efficiency in the following, since the hospitals are assumed to take output as given i.e. to act as cost minimisers.

6.2.4.1 Measuring input efficiency by the radial Farrell index

Applying FDH and DEA-C[5] methods to the standard model of the four job-categories and the two output categories on the hospital sample of 80 Danish hospitals (as described in section 6.2.3.2) yields the efficiency results depicted in figure 1. Both models are based on Farrell's

[3] See e.g. Siegel & Castellan (1988)

[4] See e.g. Siegel & Castellan (1988)

[5] The extentions on DEA, V, D, C means, respectively, a DEA with variable returns to scale, decreasing returns to scale and constant returns to scale, see e.g. chapter 3.

radial index of input efficiency. The immediate impression of these results corresponds with the intuitive expectation, since there is a relatively small number of efficient hospitals under the DEA-C technology and a very large number of efficient hospitals under the FDH technology. For 1989 the average efficiency score under DEA-C is 0.71 but 0.99 under FDH. In DEA-C, 6 out of 80 hospitals received a score of 1, whereas there are 75 out of 80 under FDH. Among these 75 efficient hospitals, 68 were undominated but non-dominating units. This large number corresponds to the findings in Tulkens (1990). Even though this difference seems large, it was partly to be expected since, theoretically, it is known that the free disposal hull technology is a subset of the constant returns to scale technology. Moreover, the efficiency scores from DEA-C and FDH constitute the range of variation in technical efficiency where DEA-C provides the lower bound and FDH provides the upper bound. As an example the largest hospital Rigshospitalet can be mentioned. In 1989 it obtained the score 0.49 by DEA-C but 1.0 by FDH. This large variation is due to the fact that Rigshospitalet is "uncomparable" under FDH since it is the largest in the sample and, by definition, it cannot be dominated by the other units in the sample.

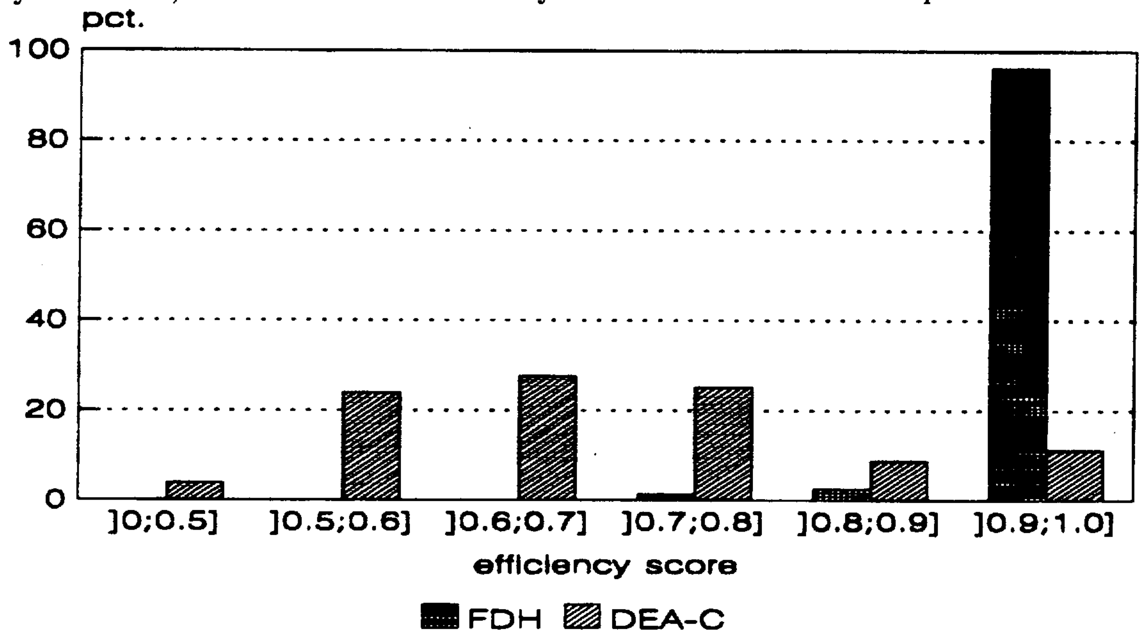

Figure 1. DEA-C and FDH efficiency scores for 1989.

Does this make FDH meaningless and DEA-C preferable when ranking the sample? The answer is classical in the sense that no direct conclusion can be drawn. At first sight the most interesting analysis seems to be DEA-C because it provides a usable ranking of the sample. However, the large variation in efficiency scores between FDH and DEA-C may indicate that constant returns to scale is too strong an assumption on the observed technology in favour of the FDH technology. At first hand it is not possible to conclude whether the variation in efficiency scores is caused by convexity or constant returns to scale (or both). However, by calculating the efficiency scores under the technological assumption of variable returns to scale it is possible to examine this aspect. The results for 1989 are illustrated by figure 2. As could be expected the variable returns to scale technology is relatively close to the free disposal hull technology, but not identical - that is the scores obtained under variable returns to scale seem to indicate that convexity is in fact of importance.

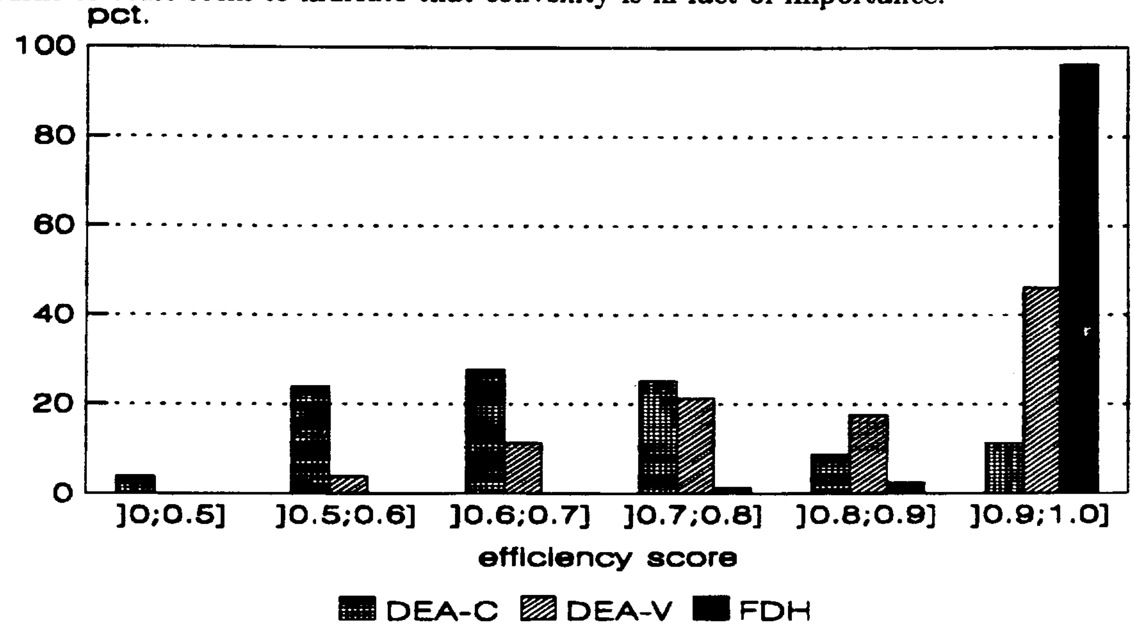

Figure 2. DEA-C, DEA-V and FDH efficiency scores for 1989.

Therefore at this early stage, there seems to be a real difference between choosing a DEA model or a FDH model - a difference which will be further examined in the following.

By introducing the DEA-V model, my results seems to indicate that unit size is negatively correlated with the DEA-C efficiency scores. This turns out to be a fact as illustrated by figure 3 where beds are used as a proxy for size, and it can be further confirmed through regression analysis[6].

The DEA-C efficient hospitals are mainly very small and specialised local hospitals which, due to the assumption of constant returns to scale can be argued to set unfair performance standards for the large regional hospitals (as indicated by a low score of around 0.5 for the group of largest hospitals). Hence I introduce:

MODEL A: *Altering the hospital sample.* Since it can be argued that very small, specialized and hence efficient hospitals are setting unfair standards, it seems obvious to try to exclude such hospitals from the sample.

[6]It is worth noticing that the negative correlation is not found among the DEA-V scores.

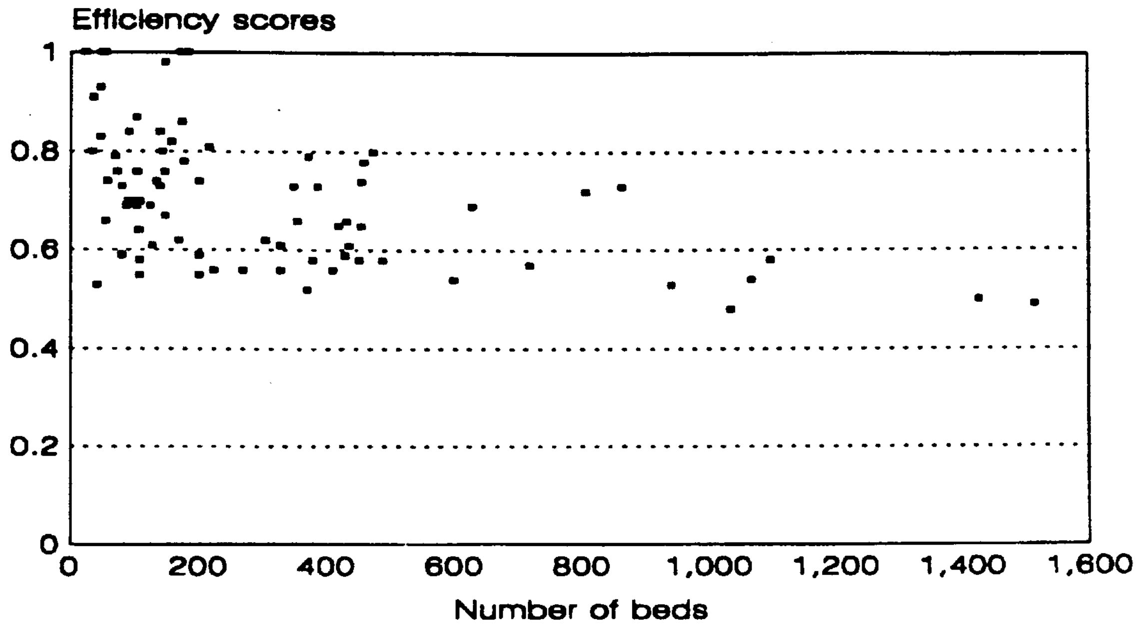

Figure 3. Hospital size and DEA-C score 1989.

As a measure of size I have chosen to represent the hospitals by the total number of patient days and set a threshold at 2.5 per cent of the observed maximum. Thus hospitals with a total number of patient days smaller than that are excluded. Furthermore the excluded hospitals must be efficient in a DEA-V sense.

This reduces the sample to 75 hospitals where the results obtained are illustrated by figure 4.

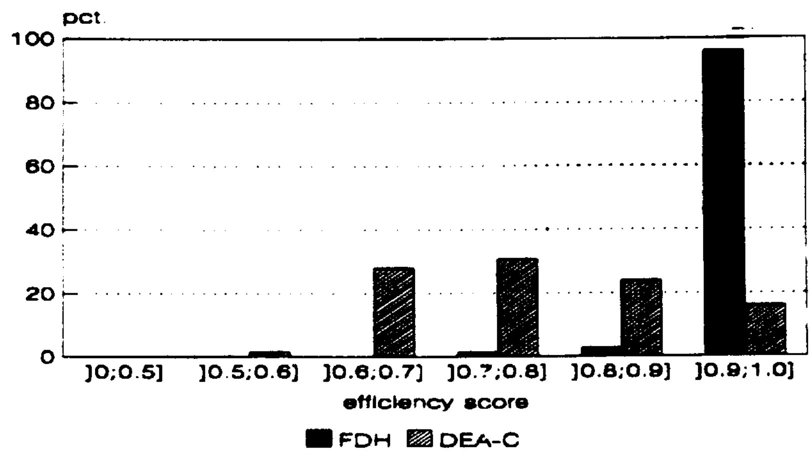

Figure 4a. FDH and DEA-C efficiency scores with reduced sample in 1989.

Notice that the variation betweeen FDH and DEA-C efficiency scores is reduced, in particular for the largest hospitals. This follows from the above-mentioned fact that excluding

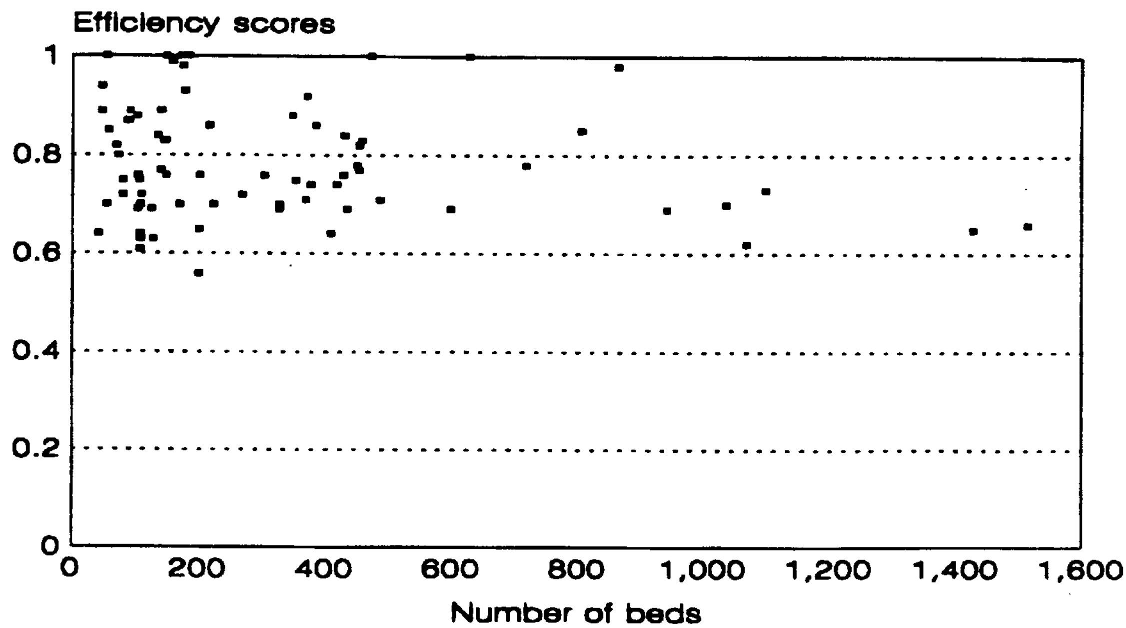

Figure 4b. DEA-C efficiency scores and size with reduced sample in 1989.

the five small and specialised hospitals from the constant returns to scale technology has the highest relative impact on the largest hospitals. Obviously there is no effect of the exclusion on the FDH technology. The previous negative correlation between DEA-C efficiency scores and size has also disappeared. Therefore, the previous difference between DEA-C and FDH efficiencies seems to have been exaggerated by the inclusion of "outliers" in the sample. The next natural step is, therefore, to analyse the impact of alternative representations of the activity.

MODEL B: *Changing the output categories.* Returning to the standard model with a sample of 80 hospitals we will analyse the effect of changes in output categories. As mentioned in section 6.2.3.2 the output category DISCH consists of the total number of discharges, but this variable does not cover the fact that patients may differ according to length of stay. Hence, an obvious alternative will be to include the total number of patient days (PDAY) as a replacement of DISCH (model B1). These results are illustrated by figure 5. In general the efficiency scores tend to increase in both DEA-C and FDH by the introduction of PDAY. 14 hospitals were efficient under DEA-C and 78 under FDH. Among these 78 efficient hospitals 76 were undominated but non-dominating. A possible explanation can be that the small hospitals, which were efficient with the variable DISCH, may be characterised by a relatively large number of uncomplicated cases. This would make the large hospitals with complicated cases dominated by the smaller hospitals. Such a feature may be revealed by the introduction of the variable PDAY to the extent that complicated cases are reflected in the length of stay. Furthermore, it is possible that aggregating emergency and non-emergency discharges implies biased efficiency results since emergency cases may interfere with hospital planning. Hence I try to include the emergency aspect explicitly by disaggregating both discharges and outpatient visits (model B2). As a result I obviously get a higher level of average efficiency (0.83) as well as more efficent hospitals. More interesting, though, is the fact that some hospitals with extreme emergency ratios have above average increases in efficiency. For

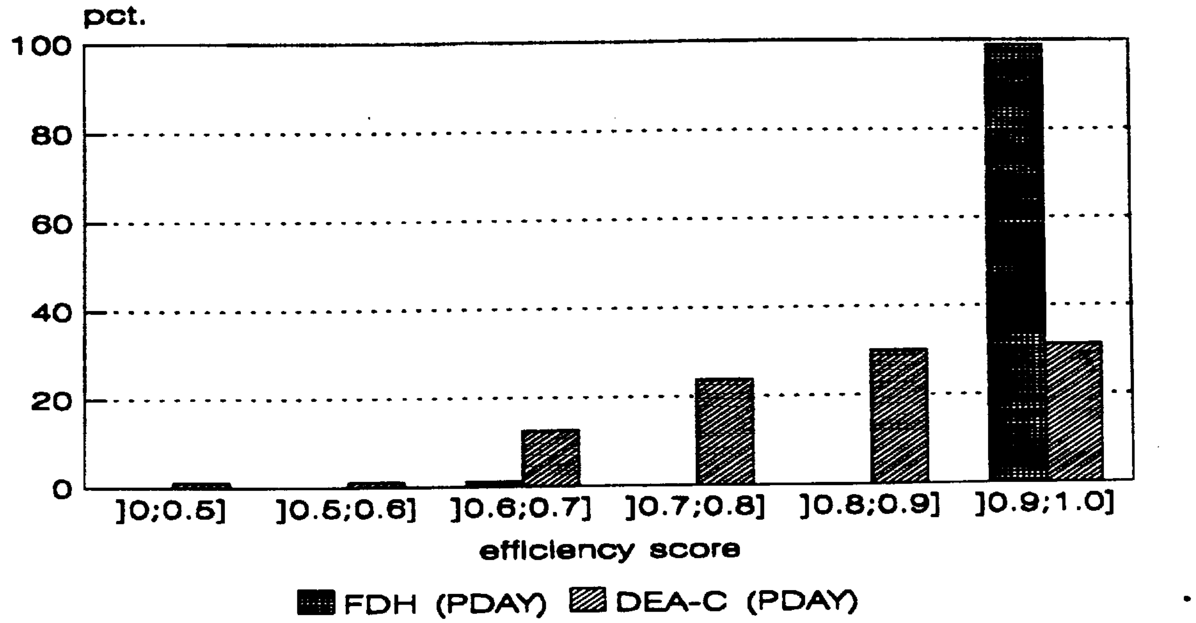

Figure 5. DEA-C and FDH efficiency scores for model B1 in 1989.

example the hospital Sundby, which has a large number of emergency discharges, changes from 0.52 in the standard model to 0.88 if outputs are disaggregated.

MODEL C: *Changing the input categories.* As one possible change in input categories of the standard model I have chosen to aggregate inputs by prices into a single variable "total current net expenditure" (EXP), model C1. In this case, it is worth noticing that the interpretation of the efficiency scores as purely technical, to a certain extent, may be misleading. Introducing EXP causes the efficiency index to represent an indication of some sort of cost-efficiency. The results from this model are illustrated by figure 6. First, it is worth noticing that aggregating inputs reduces the number of efficient hospitals in both DEA-C (where the number is 2) and the FDH model (where the number is 64). Among the 64 efficient hospitals 46 are undominated but non-dominating. This is due to the fact that reducing the number of production categories obviously makes the units more comparable, since fewer dimensions cause less specialisation. Secondly, being labour-efficient does not necessarily imply that the units are "cost-efficient". A possible difference may have several explanations. Obviously the hospitals could have an excessive use of input factors other than labour. Furthermore, measuring labour by the number of employees does not take into account either the actual hours worked or the "price" of these hours. This argument could be further analysed by using the total salary bills as an aggregated input and comparing the efficiency results with the results obtained in the standard model. As illustrated by figure 6, differences between labour and cost efficiency do, in fact occur, in my case.

If the expenditure of each hospital is multiplied by a factor defined as 1 minus the obtained efficiency score we obtain a proxy for excess spending, that is the amount which could have

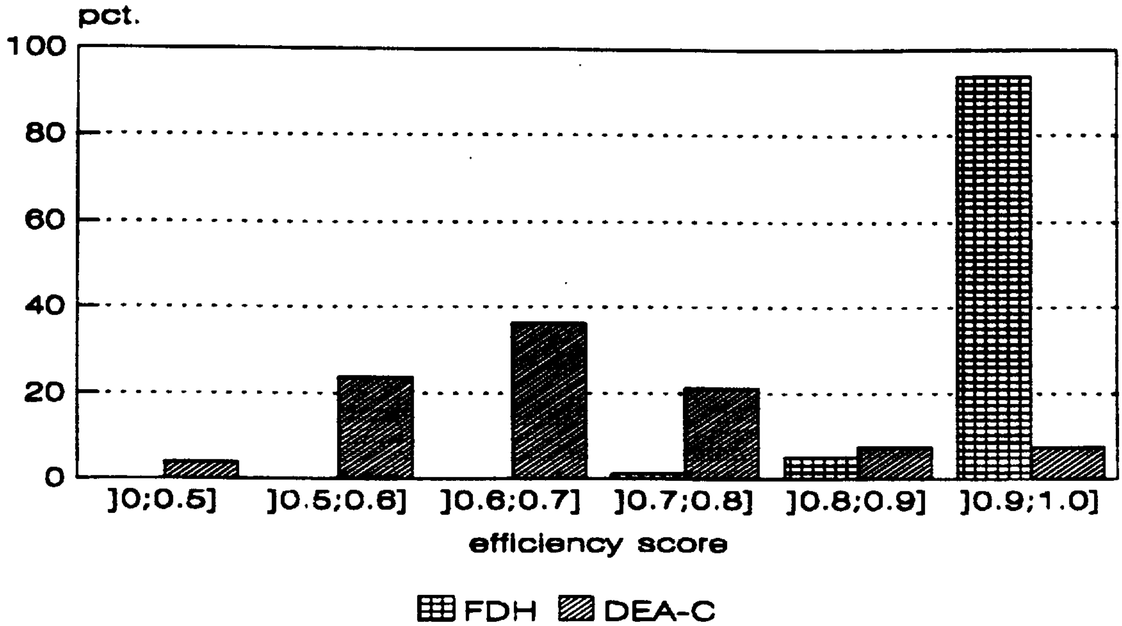

Figure 6. DEA-C and FDH efficiency scores for model C1 in 1989.

been saved if the hospital had been cost efficient. Table 3 shows the proportion of total expenditures which are due to excess spending for DEA-C and FDH models.

	FDH	DEA
Excess spending	1.7	40.7

Table 3. Excess spending as a percentage of total current net expenditure.

Obviously the proportion of excess spending is largest under DEA-C since fewer hospitals are declared cost efficient. In fact this might be a practical argument in favor of the FDH-method since, from an empirical point of view, it seems unrealistic that the hospitals should be able to reduce their expenditure by 41 per cent as indicated by the DEA-C model. A point also emphasized by Vanden Eeckaut, Tulkens & Jamar (1993).

Furthermore, as mentioned in section 6.2.3, one can consider the capital factor through the proxy "total number of beds", model C2. If the number of beds is added to the standard model, the degree of capacity utilisation becomes important when the efficiency variation is to be explained. Typically the small hospitals have a relative bad utilisation of beds, but these hospitals are normally "labour"-efficient and hence they continue to be efficient when the standard model is extended. However, among the largest hospitals, which are in general "labour"-inefficient, there is a relatively good utilisation of beds and hence these hospitals all have above average increases in efficiency scores if BED is included. This is illustrated by figure 7. The figure shows a positive correlation between the change in efficiency score and BED, that is the larger the number of beds for a hospital the higher

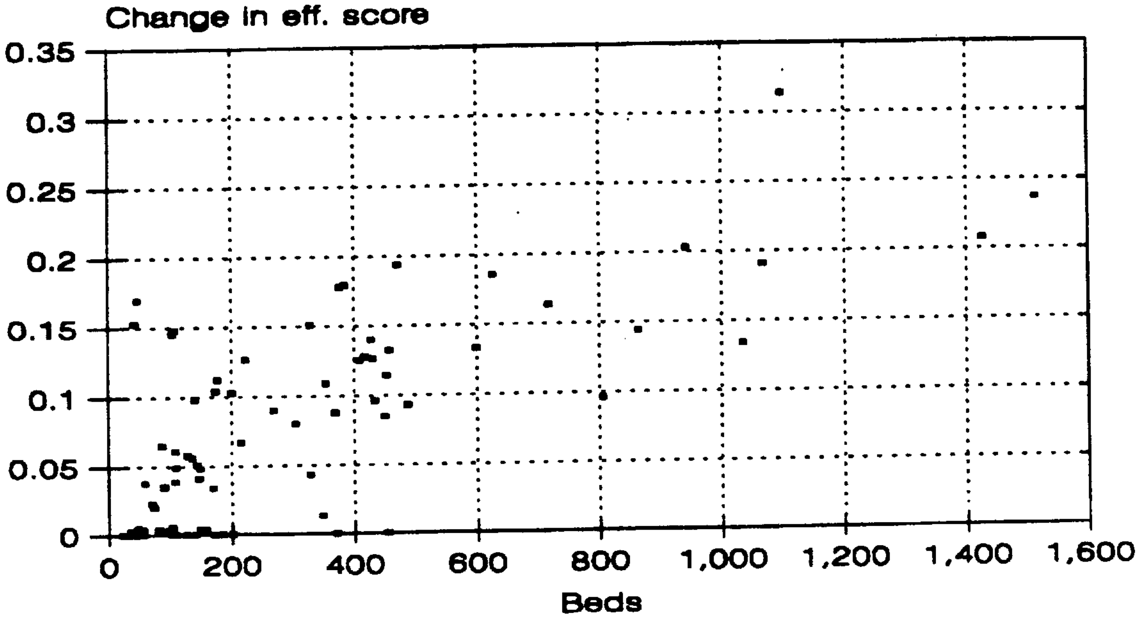

Figure 7. Changes in efficiency scores from the standard model to model C2.

is the change in efficiency score. Thus the hospital's utilisation of beds is of importance in the efficiency evaluation. This conclusion is further confirmed if, instead, the average time a bed is empty (EMPBED)[7] is added to the standard model, model C3. EMPBED measures more directly the capacity utilisation. Therefore the model with EMPBED results in above average increases in the efficiency scores for the largest hospitals. This is illustrated in figure 8. Figure 8 indicates that the correlation between efficiency score change and the number of

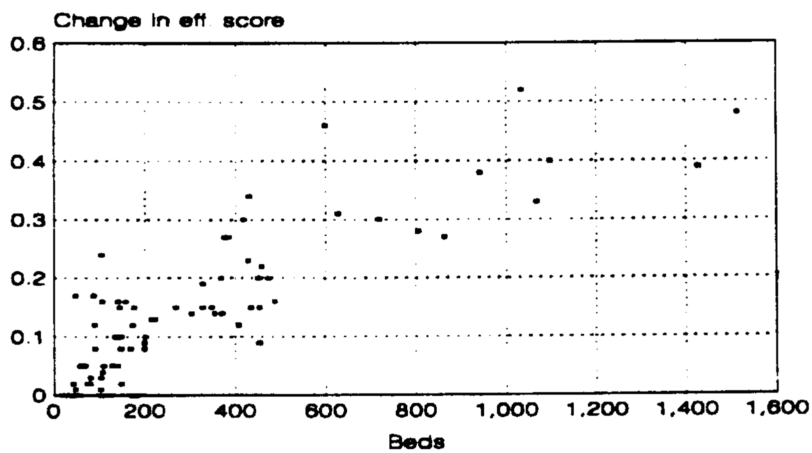

Figure 8. Changes in efficiency scores from the standard model to model C3.

beds is more significant with EMPBED included than in the model with BED included. This

[7]The average time a bed is empty measures how much the average length of stay could be increased if the hospital utilised the bed capacity completely.

can be confirmed through regression analysis: If the efficiency score changes are regressed on the number of beds then the R^2 for the model with EMPBED is 0.723, while the R^2 with BEDS is 0.514. This characteristic is related to the more direct modelling of the degree of capacity utilisation with EMPBED than with BED. The largest hospitals do not only need a relatively smaller number of beds to generate discharges, but they utilize their capacity to a higher degree.

6.2.4.2 Rank-order correlation coefficients

To test whether the alternative rankings obtained from the models mentioned above are associated, the Spearman rank-order correlation coefficient can be computed (see e.g Siegel & Castellan (1988)). This non-parametric measure makes pairwise comparisons and results in a correlation coefficient, as well as a test statistic, in relation to the null hypothesis of no association.

	Coefficient	Test statistic
Standard DEA-C vs. standard FDH	0.298	2.645
Standard DEA-C vs. standard DEA-V	0.499	4.438
Standard FDH vs. standard DEA-V	0.389	3.459
Standard DEA-C vs. model A	0.830	7.143
Standard DEA-C vs. model B1	0.650	5.778
Standard DEA-C vs. model B2	0.878	7.804
Standard DEA-C vs. model C1	0.818	7.275
Standard DEA-C vs. model C2	0.830	7.379
Standard DEA-C vs. model C3	0.544	4.832

Table 4. Corrected Spearman rank-order correlation coefficients.

Note: A value of the test statistic is significant at a 1 per cent level if it exceeds 2.326.

In fact I use the corrected Spearman coefficient because of the presence of "tied observations" which are those with identical ranking positions (in this particular case e.g. when observations have efficiency score equal to 1). In table 4 the Spearman rank-order correlation coefficients are tabulated. The test statistic follows a standardised normal distribution.

For all pairwise comparisons in table 4 the Spearman correlation coefficient is significant at a 1 per cent level. Thus I can conclude that all the efficiency rankings are associated to some extent. This holds in particular for the comparisons: standard DEA-C vs. model A, standard DEA-C vs. model B2, standard DEA-C vs. model C1 and standard DEA-C vs. model C2. Obviously FDH has a low association to DEA-C due to the relatively large number of efficient units and likewise FDH vs. DEA-V has a low association. The high association between standard DEA-C and model A indicates that removing the five small hospitals does not alter the ranking of the remaining hospitals in a significant way. However, it is worth noticing that the average efficiency increases in model A, an aspect which the Spearman coefficient cannot take into account. If model B1 is considered, i.e. change the output category DISCH for PDAY, we obtain a moderate degree of association, that is, it does seem to have an impact on the over all efficiency ranking result. On the other hand disaggregating DISCH and AMBULA to emergency and non-emergency does not seem to have a significant effect on the ranking of the hospitals, although the average level of efficiency increases in this specification. If the model where inputs are aggregated to total current net expenditure is compared to the standard model there is a very high degree of association indicating that the aggregation is justified in the sense that information is preserved. Here it is worth noticing that around 70 per cent of the costs is composed of salary. Similar there is a high association between the standard model and the standard model extended with the number of beds (model C2). Thus, although the average level of efficiency increases and especially the largest hospitals increase their efficiency score, the ranking is preserved to a high extent. This is not the case for the standard model extended with the average empty bed time (model C3) which is rather weakly related to the standard model.

Both model C2 and model C3 were constructed with respect to concerns about the bed capacity utilisation, but they have indeed very different effects on the ranking from the standard model. Therefore it seems that including the variable EMPBED in the model does provide additional information about the hospitals' performance, whereas the variable BED does not include any significant new information about performance.

6.2.4.3 The non-radial Fãre-Lovell index

Replacing Farrell's radial efficiency index with the non-radial Fãre-Lovell index provides additional and useful information about partial performance. Consider the following specific result concerning a single hospital (Rønne hospital) obtained with respect to the standard model under FDH. The scores are depicted in table 5.

Hospital	θ_1	θ_2	θ_3	θ_4	E_{FL}
Rønne	0.88	0.87	0.87	0.54	0.79
Bispebjerg	0.47	0.73	0.39	0.34	0.48
Avg. eff.	0.68	0.78	0.62	0.57	0.66

Table 5. Examples of partial Fãre-Lovell input efficiency scores 1989.

Note: $\theta_1, \theta_2, \theta_3, \theta_4$ refer to partial efficiency score, for DOCTOR, NURSE, OCARE and OTHER respectively. $E_{FL} = (\theta_1 + \theta_2 + \theta_3 + \theta_4)/4$.

Notice that the obtained Fāre-Lovell score of 0.79 is a mean of the four partial input efficiency scores, and hence difficult to interpret. The important information consists rather of the partial input scores themselves because they measure the ability to utilise each input category separately. From table 5 it is noticed that Rønne hospital seems to have a bad utilisation of the input factor OTHER, whereas the utilisation of DOCTOR, NURSE and OCARE are relatively good. In the case of Rønne the largest partial efficiency score is 0.88 for the input factor DOCTOR, which incidently is identical to the Farrell efficiency score. In general this property will not be present. As a special feature of the FDH model the actual amounts of input weighted by their related partial efficiency scores yield the input vector characterising the best-practice reference hospital. In the case of Rønne this is found to be Frederikssund hospital.

Consider table 5 again where the efficiency scores for Bispebjerg hospital are given. These results are obtained using the DEA-C model on the standard model and the sample of 75 hospitals. First, I notice that the largest partial score (NURSE = 0.73) is higher than the radial efficiency score 0.70. Secondly, weighting the input vector by the partial scores does not necessarily result in the input vector of an actually existing hospital. This follows from the assumption of convexity. The scores of Bispebjerg hospital are quite interesting because they illustrate the consequences of radial efficiency evaluation. Notice that except for NURSE all other factors are badly utilised by Bispebjerg, but if efficiency is evaluated radially the relatively good utilisation of NURSE covers up this fact. Though, the radial efficiency score and the largest partial score are not completely identical, the latter determines the former.

The last line in table 5 shows the partial Fāre-Lovell input efficiency scores for the 80 hospitals on average. The average score for DOCTOR is probably underestimated since output does not account for teaching activities. The relatively good utilisation of DOCTOR and NURSE are mainly caused by fixed settings of the number of employees per bed. The bad utilisation of OCARE and OTHER are partly a result of the fact that some hospitals have privatised cleaning which obviously interferes with OTHER.

6.2.5 The significance of non-included variables on the efficiency results of the standard model

As observed in the various models there are several factors which have an impact on the efficiency ranking of the hospitals. There are variables which are not directly included in the standard model although they influence the efficiency variation obtained. In an attempt to estimate and systematize the relative significance of such non-included variables I have chosen to follow a regression approach where the efficiency scores of the standard model are regressed on a vector of different variables of which some of them are so-called environmental factors. Obviously I could have chosen to include all such variables directly in the DEA-model but this would often result in too many dimensions causing meaningless results as it was partly illustrated by the different models.

In the following I apply the standard procedure for regression analysis, that is:

1. Based on the efficiency variation from the standard model I specify a regression model and define the variables to be included.
2. Presentation of regression results.
3. Examination of some econometric issues related to whether the assumptions from the OLS estimator are satisfied.
4. Residual analysis and model implications.

6.2.5.1 A regression model

I assume that the variation in efficiency scores can be approximated by the following log-linear model:

$$\ln(\theta) = Z\beta + e$$

where θ is the vector of efficiency scores, Z is a matrix of explanatory variables which includes both controllable organisational features and non-controllable characteristics for each hospital and e is a disturbance term with mean 0 and standard deviation σ. Variables in the matrix Z will be defined below. The above specification is chosen in order to obtain consistent estimates of β. The problem of consistency arises because the efficiency scores are restricted to take values between 0 and 1. This generates a dependency between the Z-variables and the disturbance term e. In the above specification consistent estimates of β can only be obtained if the efficiency scores are allowed to take values without an upper bound. Hence I will use the procedure for ranking the efficient observations described in chapter 3. Inefficient observations obtain the same efficiency score, but the efficient observations ($\theta = 1$) can obtain scores above 1. These efficiency scores indicate how much an efficient observation could increase its inputs and remain efficient. This procedure is applicable only when constant returns to scale is assumed and thus the model is restricted to DEA-C. $\ln(\theta)$ is not defined for $\theta = 0$, but since all efficiency scores are greater than 0 this case does not represent a problem.

From the models in section 6.2.4.1 it is known that at least the following excluded factors play a role in the efficiency ranking obtained in the standard model: specialisation in outputs, hospital size, the number of patient days, capacity utilisation approximated by the number of beds and the proportion of emergency cases. As it turns out, in the regression analysis, the following explanatory variables provide a satisfactory estimation of the model:

1. OUTPROP: The ratio of the number of outpatient visits to the number of discharges, i.e.

$$OUTPROP = \frac{\text{AMBULA}}{\text{DISCH}}$$

This variable describes an aspect of the output structure related to the case-mix and in this sense it also reflects the extent to which the hospitals are specialised with respect to the output specification in the standard model. From the standard model it was known that this type of output specialisation influences the efficiency ranking. The expected sign in the regression analysis is positive since outpatient cases should be less resource demanding than inpatient cases. Obviously, this variable is not under the control of the hospital management.

2. RCMSHARE: The inverse of the ratio of the number of beds at the hospital to the total number of beds in the county of the hospital, i.e.

$$RCMSHARE = \frac{\text{Total beds in county } j}{\text{beds at hospital } i \text{ in county } j}$$

RCMSHARE is included as an indicator of the degree of centralization. A low value for this variable implies that a hospital is a major provider of health care in the county. As such this variable offers a better representation of the degree of centralization compared to the number of beds, for example, since the size of the other hospitals in the county is taken into account. Since the aspect covered by RCMSHARE is not explicitly included in the model specification it can be considered as an environmental factor and therefore outside the control of the management. The expected relationship between the efficiency scores and RCMSHARE is positive since larger hospitals (which have a low value of RCMSHARE) may have a more complicated case-mix, use more resources on teaching[8] and there is a possibility of scale effects as indicated by the DEA-C model.

3. MTIME: The average length of stay, defined as the number of patient days divided by the number of discharges, i.e.

$$MTIME = \frac{\text{PDAY}}{\text{DISCH}}$$

MTIME is also an environmental factor in relation to the actual model specification and can be taken as a proxy for the complication of the case-mix. Moreover inefficiencies might appear through longer length of stay since it could be an indicator for slow treatment procedures or unnecessary tests, although other factors may influence the magnitude of the average length of stay. Therefore the expected relationship between the efficiency scores and the average length of stay should be negative. In principle MTIME can be controlled by the management since it, to a certain extent, is able to control the number of patient days. It should be noticed that MTIME is one of the traditional indicators used in the hospital sector as performance measure and was considered, through the variable PDAY only, as an output in model B1.

4. EMPBED: The average time a bed is empty, defined as the average length of stay divided

[8] Unfortunately it has not been possible to include directly a variable measuring the resources devoted to teaching in the regression model due to lack of centrally-collected data.

by the occupancy in percentage terms (CAPPCT) minus the average length of stay, i.e.

$$EMPBED = \frac{\text{MTIME-CAPPCT*MTIME}}{\text{CAPPCT}}$$

where

$$CAPPCT = \frac{\text{PDAY}}{\text{365*BEDS}}$$

EMPBED concerns the capacity utilisation by measuring the average time a bed is empty, that is the average time between one patient leaving the hospital and the next one arriving. A high value of EMPBED could indicate that the planning of patient flow is bad such that the hospital could increase the number of patient days without capacity consequences. For two hospitals which only differ with respect to EMPBED the hospital with the highest value of EMPBED will other things being equal, obtain the lowest production level. Therefore a negative relationship between the efficiency scores and EMPBED is assumed. Notice that EMPBED is one of the key variables used as a traditional performance measure in the hospital sector. BED and EMPBED were considered as alternative inputs in model C2 and C3.

5. EPROPPD: The proportion of the patient days which is made up of emergency cases, i.e.

$$EPROPPD = \frac{\text{Emergency patient days}}{\text{PDAY}}$$

The variable EPROPPD is an environmental factor, and hence uncontrollable, which concerns the output structure with respect to patient days by measuring the proportion of emergency cases. A negative relationship between the efficiency scores and EPROPPD should be expected due to restricted possibilities of planning.

6. NEPROPAM: The proportion of the outpatient visits which is made up of non-emergency cases, i.e.

$$NEPROPAM = \frac{\text{Non-emergency outpatient visits}}{\text{AMBULA}}$$

The variable NEPROPAM is also a non-controllable environmental factor which describes a characteristic of the output structure concerning outpatient visits, AMBULA. If NEPROPAM is equal to 1 it implies that no outpatient visits are emergency cases since the hospital does not have an emergency clinic. As such this variable gives a better representation of the presence of emergency outpatient visits compared to a dummy variable indicating whether the hospital has an emergency clinic or not, since the proportion of emergency cases is taken into account. In general, emergency cases tend to lower the efficiency of a hospital since they restrict the possibility of planning and the emergency clinic has to be with staff even when there are no patients. Therefore the expected relationship between NEPROPAM and the efficiency scores is positive. The emergency aspect was considered in model B2.

7. LABPRBED: The ratio of the total number of employees to the number of beds, i.e.

$$LABPRBED = \frac{\text{Total number of employees}}{\text{BEDS}}$$

The variable LABPRBED is indeed controllable and represents a form of inefficiency namely too many employees per bed and, as such, it characterises the organisation of the hospital production process. It indicates the intensity of the health care production which might cover certain quality aspects. However, the expected relationship between LABPRBED and the efficiency scores is negative since low values of LABPRBED will, other things being equal, give high efficiency scores through labour resource savings.

In table 6 the signs of the relationship between the above listed explanatory variables and the efficiency score θ are shown.

	$\ln\theta$
OUTPROP	+
RCMSHARE	+
EMPBED	-
MTIME	-
EPROPPD	-
NEPROPAM	+
LABPRBED	-

Table 6. The expected relationship between the explanatory variables and the dependent variable $\ln\theta$.

The dependent variable, $\ln\theta$, is based on the efficiency scores[9] from the standard model with DOCTOR, NURSE, OCARE and OTHER as inputs and the number of discharges (DISCH) and the number of outpatient visits (AMBULA) as aggregated output variables. The model is estimated for 1989 by ordinary least squares OLS (although other estimation techniques could have been applied). Table 7 shows some descriptive statistics of the variables in the regression model.

[9]The efficiency scores are as mentioned allowed to take values above 1.

MEANS of VARIABLES

Ldeac	constant	outprop	rcmshare	empbed	mtime	eproppd	nepropam	labprbed
-.3461	1.0000	3.2629	16.1129	2.1406	6.8191	0.6780	.6823	2.6475

STANDARD DEVIATIONS OF VARIABLES

Ldeac	constant	outprop	rcmshare	empbed	mtime	eproppd	nepropam	labprbed
.2487	.0000	3.6318	22.1890	1.2767	1.3418	.1432	.1663	.5012

DURBIN-WATSON TESTS

Ldeac	constant	outprop	rcmshare	empbed	mtime	eproppd	nepropam	labprbed
1.2587	.0000	1.3282	1.1063	1.3220	1.7228	1.8678	1.8805	1.1977

CORRELATION MATRIX

	Ldeac	constant	outprop	rcmshare	empbed	mtime	eproppd	nepropam	labprbed
Ldeac	1.0000								
constant	.0000	1.0000							
outprop	.4061	.0000	1.0000						
rcmshare	.5845	.0000	.5102	1.0000					
empbed	.3018	.0000	.4830	.6812	1.0000				
mtime	-.3230	.0000	.2640	-.0161	.0123	1.0000			
eproppd	-.2996	.0000	-.4051	-.3394	-.4481	.0774	1.0000		
nepropam	.2011	.0000	.3585	.2218	.2334	.1474	-.3751	1.0000	
labprbed	-.5210	.0000	.1050	-.3427	-.4304	.0505	.0080	.1251	1.0000

Table 7. Some descriptive statistics for the variables in the regression model.

6.2.5.2 Regression results

The regression results are displayed in table 8.

The regression model appears as (ignoring the disturbance term):

$$\ln(\theta) =$$

1.22 + 0.04OUTPROP + 0.004RCMSHARE - 0.13EMPBED - 0.08MTIME - 0.19EPROPPD + 0.25NEPROPAM - 0.36LABPRBED

EQ(1) Modelling Ldeac by OLS

VARIABLE	COEFFICIENT	STD ERROR	H.C.S.E.	t-VALUE	Par r^2
constant	1.21842	.12947	.30399	9.41079	.5516
outprop	.04255	.00414	.00739	10.27777	.5947
rcmshare	.00428	.00070	.00117	6.14493	.3440
empbed	-.12798	.01314	.02012	-9.74275	.5687
mtime	-.08366	.00848	.01324	-9.87052	.5750
eproppd	-.19151	.08969	.25501	-2.13535	.0596
nepropam	.24531	.07140	.07136	3.43576	.1409
labprbed	-.36465	.02645	.03553	-13.78742	.7253

R^2 = .8721101 σ = .0931671 F(7, 72) = 70.14 [.0000] DW = 2.048
RSS = .6249677471 for 8 Variables and 80 Observations
Information Criteria: SC = -4.413879; HQ = -4.556580; FPE = .009548
R^2 Relative to DIFFERENCE+SEASONALS = .89500

Table 8. Regression results.

This model can explain a high proportion of the variation in the dependent variable, $\ln(\theta)$, as reflected by $R^2 = 0.87$ ($\bar{R}^2 = 0.86$). This is confirmed by the F-test where the null-hypothesis $\beta_{outprop} = \beta_{eproppd} = \beta_{rcmshare} = \beta_{labprbed} = \beta_{mtime} = \beta_{nepropam} = \beta_{empbed} = 0$ is rejected at the 1 percent level (the F-statistic is equal to 70.14 which is much greater than the critical value $F_{0.99}(7,72) = 2.9$). As could be inferred from the different models in section 7.2.4.1 a large part of the differences in efficiency is related to the included explanatory variables. It seems unlikely that other (excluded) variables should prove significant in relation to the overall sample. Influences of other variables (e.g. whether the cleaning at the hospital is privatized or not) on the efficiency scores are covered by the included variables. The remaining variation in the efficiency scores is mainly due to statistical noise. Furthermore the significance of each variable examined by the t-test implies that for all variables the null hypothesis $\beta_i = 0$ is rejected at a 5 percent level. Therefore the included variables seem to be relevant for the explanation of efficiency differences in the standard model.

Looking at the sign of the parameter estimates these have all obtained the expected signs. Therefore the estimation confirms that hospitals with high proportions of non-emergency outpatient visits (NEPROPAM) and low proportions of emergency inpatient cases (EPROPPD) have higher levels of efficiency. In addition hospitals with high numbers of outpatient visits compared with the number of discharges tend to have higher efficiency scores. Similar hospitals with small proportions of the total county bed supply (RCMSHARE) tend to have higher efficiency scores. The sign for LABPRBED is positive, meaning that hospitals with a high number of employees per bed tend to have lower efficiency scores. The coefficient of the average length of time (MTIME) is negative indicating that longer length of stay implies lower efficiency scores. Moreover hospitals with long average empty bed time (EMPBED) have lower efficiency scores.

6.2.5.3 Econometric issues

Before proceeding to a more detailed analysis of the estimated model concerning examination of the residuals and evaluation of the implications of the model I will turn to some econometric issues related to the estimation. First, I will analyse the presence of heteroscedastic errors because if heteroscedasticity is present the desirable properties of the OLS estimator with respect to the minimum variance of the parameters fails to hold. I consider one possible source of heteroscedasticity namely the hospital size measured by the number of beds. Heteroscedasticity from this specific variable could arise from a larger variation in the patient flow as hospital size increases. This increased variation in the patient flow could be the result of higher proportions of patients from foreign counties, where this number could be more difficult to forecast than patients from the county where the hospital is situated. This hypothesis is tested with the Breusch-Pagan test for heteroscedasticity of a particular form. In this case the test consists of regressing the squared residuals on hospital bed size. The test statistic, l, is equal to the product of R^2 from this regression and the number of observations. l is asymptotically χ^2 distributed where the degrees of freedom is equal to the number of regressors. R^2 is equal to .018 and the number of observations is 80 giving a value of $l = 1.44$. The critical value for $\chi^2_{0.95}(1) = 3.841$. Threfore I accept the null hypothesis of homoscedastic errors for this particular form, i.e. hospital bed size does not influence the errors in any systematic way.

The consequences of using OLS when errors are autocorrelated are the same as with heteroscedastic errors, namely unbiased but inefficient estimates and problems with inference procedures. In the case of cross-section data the autocorrelation stems from other observations at the same time. One a priori explanation for autocorrelated errors in the present model is mainly due to the hospital data structure. Data are listed such that the county structure is preserved and one county's hospitals are followed by hospitals from a neighbouring county. Moreover, in general, the largest hospitals in a county are listed before the smaller hospitals in the county. This data structure could clearly result in dependencies between the errors. I have applied the Durbin-Watson test for 1st. order autocorrelation and obtained a $DW = 2.05$. DW values higher than 2 mean that the null hypothesis of non-autocorrelated residuals has to be compared with the alternative hypothesis of negative first-order autocorrelation. The values of the upper and lower bounds indicate that the null hypothesis of non-autocorrelated residuals can be accepted at the 1 percent level.

However this is only testing for 1. order autocorrelation. In order to test for higher order autocorrelation I have employed a Breusch-Godfrey test with the test statistic

$$l = n(r_1^2 + r_2^2 + ... + r_p^2)$$

where r_i is the i'th autocorrelation of the OLS residuals and n is the number of observations. The test considers p. order correlation as the maximum. l is asymptotically χ^2 distributed with p degrees of freedom. We have chosen to use a model for autocorrelation where 10. order correlation is the maximum, i.e. $p = 10$, since the largest number of hospitals in a county is 10. The test statistic, l can be computed as:

$$l = 80(r_1^2 + ... + r_{10}^2) = 6.755$$

Since $\chi^2_{0.95} = 18.307$ the null hypothesis of no autocorrelation is accepted. Possible dependencies among the errors due to the data structure can be rejected to influence the errors in a systematic way.

The assumption of normally distributed errors is only important with respect to inference procedures, e.g. the possibility of using F- and t-tests. Therefore non-normal errors as such do not change the attracting properties of the OLS estimator. In order to test for the normality of the residuals I tested for whether skewness and excess kurtosis are jointly zero (since both skewness and excess kurtosis will be zero if the population of residuals has a normal distribution. The null-hypothesis is that the skewness and excess kurtosis are jointly zero compared with the alternative hypothesis that skewness and excess kurtosis are not jointly zero. The test statistic c is defined as: $c = \frac{(n-k)}{6}(SK^2 + \frac{1}{4}EK^2)$, where SK is skewness, EK is the excess kurtosis and k is the number of regressors. c is asymptotically χ^2 distributed with 2 degrees of freedom if the null-hypothesis is true. For n=80 and k=8 c becomes equal to 8.296 and the critical value for $\chi^2_{0.99}(2) = 9.2103$. With $8.296 < 9.2103$ I conclude that the null hypothesis can be accepted, i.e. the population of residuals can be approximated as normal distributed.

Finally I will consider the presence of significant multicollinearity. It is possible a priori that some of the independent variables are highly correlated and therefore can result in more uncertain parameter estimates. The possibility arises because some of the independent variables can be structurally related, e.g. hospitals with small proportions of emergency outpatient visits could be expected to have small proportions of emergency patient days as well. A crude indicator for multicollinearity is a high R^2 combined with many insignificant coefficients, but in the present case a fairly high R^2 is combined with significant coefficients for all variables. Another indicator for multicollinearity is if the parameter estimates do not have the expected signs, but this is not the case. Moreover, looking at the correlation matrix reveals only few highly correlated variables. Thus only two correlation coefficients are higher than +0.5 or -0.5. The problem with this procedure for detecting multicollinearity is that it only considers pairwise dependencies but not more complicated, collinear patterns. However it should be noted that the problem of multicollinearity is dependent on the intended application of the regression results. If the purpose is to examine the sign of single coefficients then multicollinearity represents a problem. On the other hand if the purpose is to explain as much of the variation in the dependent variable as possible then multicollinearity is of less importance. In the present case both applications are interesting but based upon the correlation matrix and the other crude indicators multicollinearity does not seem to be significant.

6.2.5.4 Residual analysis and model implications

The size of the residuals provide information concerning how well the model explains the dependent variable for each hospital – a positive (negative) value of the residual implies that the actual efficiency is larger (smaller) than predicted by the model. Obviously there may be local explanatory variables for the residual of a single hospital, but such variables are not included

since they are not of general significance with respect to the chosen regression model. The (un-scaled) residuals take values in the range from -0.153 to 0.275. All residuals are quite small although the largest residual of 0.275 covers a difference between observed and estimated efficiency score of 0.22. Overall a good fit of the estimated efficiency scores compared with the actual efficiency is present scores which is indicated by the low standard deviation of the residuals equal to 0.09 around the mean of 0.

As an example of how the model functions I consider the largest positive residual of 0.275 obtained by Ærøskøbing hospital. This residual is the difference between an efficiency score of 0.91 and the estimated score of 0.69. The estimated efficiency score is obtained from the following Z values: OUTPROP = 1.40, RCMSHARE = 68.58, EMPBED = 3.72, MTIME = 8.44, EPROPPD = 0.80, NEPROPAM = 0.84 and LABPRBED = 2.22. The most important contribution to the variation stems from LABPRBED followed by MTIME and EMPBED. The model underestimates the actual efficiency score due to an unusual output structure at Ærøskøbing hospital. Normally small hospitals have high proportions of non-emergency cases for both inpatient treatment and outpatient treatment. In the case of Ærøskøbing hospital the high proportion of non-emergency outpatient visits is accompanied by a high proportion of emergency patient days.

If I consider the average percentage contribution of each explanatory variable to the overall explanation of the model, I obtain the results as listed in table 9[10].

[10] The average percentage contribution is calculated by multiplying the average values of the explanatory variables by the estimated parameters and then calculating the ratio of the absolute value of each pair over the sum of the absolute values of all pairs.

OUTPROP	5.8
RCMSHARE	2.8
EMPBED	12.2
MTIME	24.0
EPROPPD	5.7
NEPROPAM	7.5
LABPRBED	42.0
Total	100.0

Table 9. Average percentage contribution of the explanatory variables to the overall explanation of the model.

As noticed previously the included explanatory variables differ with respect to the possibilities of control from the point of view of the hospital management. In general, variables related to the output structure are being out of the hospital management's control, that is OUTPROP, EPROPPD and NEPROPAM which cover a total of 19 % of the variation. These variables are determined mainly by the patient flow although they are also influenced by the hospital facilities. Moreover the proportion of beds for a given hospital to the total county bed supply is not controlled by the hospital but by the regional authorities. In addition, assuming the patient flow to be exogenous, reducing MTIME in order to improve efficiency can only be obtained by an increase in EMPBED and, therefore, leaving the efficiency unchanged.

Only LABPRBED seems to be adjustable by the hospital management but this variable is very influential with respect to the variation in efficiency. It has the highest numerical parameter and can account for around 40 per cent of the variation. Although part of the 40 per cent could be caused by non-excessive labour usage (e.g. hospitals with teaching commitments or hospitals providing high-quality care) it still indicates that decreasing the labour per bed ratio could be a possible way to obtain efficiency improvements.

The variables included in the estimated model were chosen according to the information obtained from the models described in section 6.2.4.1. This information compressed by the Spearman correlation coefficients indicated that including the number of patient days or a measure for the capacity utilisation had especially strong effects on the efficiency ranking. This is indeed confirmed by the estimation of the regression model as illustrated in table 9, since MTIME and EMPBED are those variables which, next to LABPRBED contribute on average most to the explanation of the efficiency variation. Excluding these variables from the standard model implies that part of the measured inefficiency is caused by the exclusion of these variables from the model. The low percentage contribution of RCMSHARE indicating the relative size of the hospital is surprising due to the clear correlation between beds and the efficiency scores from the standard model. However a part of the potential explanation of RCMSHARE is taken over by EMPBED since this variable has a higher influence the larger the hospital. The relatively low average contribution to the model by the two emergency-related variables corresponds to the relatively high Spearman correlation coefficient between the standard model and model B2. Disaggregating the output categories into emergency and non-emergency cases did not induce a significantly changed ranking. The same holds for the variable OUTPROP (indicating the extent of output specialisation) which has a low contribution to the model explanation and a high Spearman correlation coefficient for the association between the standard model with 80 hospitals and the standard model with 75 hospitals. In large, the information from the non-parametric analysis and the parametric analysis correspond to each other.

The validity of the results described above is examined by using the reduced data set with the 5 non-comparable hospitals excluded. This analysis can be viewed as testing for possible influence of outliers or extrem observations on the estimation results. The dependent variable $\ln\theta$ is based on the DEA-C efficiency scores from model A with 75 hospitals except for the efficient observations where the scores obtained from the procedure for ranking the efficient units again is used. In table 10 the parameter signs from these two data sets are compared:

	80 hosp.	75 hosp.
OUTPROP	0.04	0.11
RCMSHARE	0.004	0.004
EMPBED	-0.13	-0.10
MTIME	-0.08	-0.09
EPROPPD	-0.19	-0.02
NEPROPAM	0.25	0.16
LABPRBED	-0.36	-0.28
R^2	0.87	0.81

Table 10. Parameter estimates for the regression model from the complete data set and the reduced data set.

In general, identical signs for the estimates is obtained and, moreover, the size of the estimates are approximately the same. However, the variable EPROPPD is an exception since the estimate is much lower and insignificant in the reduced data set. This insignificance is caused by a very strong relation between the 5 hospitals and EPROPPD. Reducing the data set lowers the standard deviation which drops from 0.14 to 0.10. But from an overall point of view the chosen model seems to be robust with respect to changes in the hospital sample.

6.2.6 Concluding remarks

The above efficiency results can be summarised by the following table:

	Standard	DEA-V	A	B1	B2	C1	C2	C3
DEA								
Min	0.48	0.57	0.56	0.45	0.54	0.41	0.55	0.55
Mean	0.71	0.86	0.79	0.84	0.83	0.68	0.78	0.84
Std	0.14	0.13	0.11	0.12	0.14	0.12	0.12	0.13
FDH								
Min	0.79	-	0.79	0.64	0.87	0.78	0.79	0.88
Mean	0.99	-	0.99	0.99	0.997	0.982	0.995	0.997
Std	0.03	-	0.03	0.04	0.02	0.05	0.03	0.02

The FDH-method is fairly uninteresting since almost nothing can be concluded while DEA offers a usable ranking but imposes strong restrictions concerning the production technology. The standard model was changed in a number of different ways, which revealed that other excluded variables did have an effect on the efficiency results. The non-parametric analysis was supplemented in a second step with a parametric regression analysis where the efficiency scores of the standard model were regressed on explanatory variables chosen according to the different models. This approach implies an interesting link between parametric and non-parametric analyses and provides a procedure for introducing statistical evaluation of the findings obtained.

The practical relevance of this kind of analysis concerning hospital planning and management can be contested due to the fact that the analysis includes a wide range of hospitals, which may not appear as similar as demanded by the theory. However, the basic procedures seem applicable. If one finds it unrealistic to compare hospitals it is possible to restrict the analysis to cover departments etc. Furthermore, DEA has, in general, been fairly successfully applied to Scandinavian hospital data[11]. Therefore DEA seems to be a promising tool for analysing the extent of inefficiency amongst health care producers.

[11] Roos (1993) examines productivity changes for Swedish hospital data and Magnussen (1992) considers efficiency differences between Norwegian hospitals.

6.3 Measuring input efficiency with allowance for case-mix variation in outputs.

6.3.1 Introduction.

In the previous section on measuring input efficiency for the set of Danish hospitals several interesting aspects concerning the extent and reasons for inefficiency were revealed. However, there are two important and related problems with the data set. First, the hospitals included in the data set are relatively heterogeneous at least in terms of output profile. Moreover, it is implicitly assumed that all cases require the same input level since cases from different diagnoses are summed to form the totals: total number of discharges, total number of patient days and total number of outpatient visits. No allowance is made for case-mix variation in the efficiency analysis. As described in chapter 5 it is important to adjust for case-mix differences in order to get a reliable product description and thereby to obtain valid efficiency results. In this section a revised data set will be introduced in order to analyse efficiency for Danish hospitals where it is possible to allow for case-mix variation adjustments. Unfortunately it is only possible to adjust for heterogeneity regarding in-patient hospital care represented by the number of discharges. Out-patient visits will be aggregated, as before, without adjusting for differences. Moreover the number of hospitals will be reduced such that the remaining set of hospitals are more similar. This modified data set will be used to measure the extent of inefficiency for Danish hospitals (the present section), calculating the returns to scale properties (sec 6.4), comparing the non-parametric efficiency results with corresponding parametric results (sec. 6.5) and calculating capacity utilisation rates (sec. 6.6).

6.3.2 The data.

The input data are the same as in the previous section. The output data regarding in-patient care represented by discharges are new. The primary data on discharges correspond to the number of discharges in each of the 470 DRG groups for each hospital for the year 1989 (the US-developed DRG system was described in chapter 5). These DRG-based discharges for Danish hospitals are not implemented in the official hospital statistics (from the National Board of Health), but have emerged as the result of research from the Danish Health Ministry in order to develop a Danish DRG system (see Bay-Nielsen & Olesen (1993)). Therefore, at present, there does not exist a Danish set of weights for each DRG group in order to construct aggregated outputs. This is necessary since the 470 disaggregated output categories cannot be used for efficiency analysis. As a rough alternative, weights developed for a Norwegian DRG system have been used in the Danish DRG-project and these weights will also be used in the present analysis as well. However there can be problems with this step, since these Norwegian weights are only valid if the Danish and Norwegian average length of stay are similar to a high degree, but this do not seem to be the case. Therefore, there can be problems in the weighting procedure which should be taken into account when the results are interpreted.

In the analysis I include two output measures are used: a weighted measure for discharges (WDISCH) obtained by weighting the discharges in each DRG group with the corresponding Norwegian DRG weight and then summing over the groups, and the number of outpatient visits (AMBULA). The number of hospitals included in the following analysis is smaller than in the previous analysis due to a more restrictive selection procedure[1]. Only somatic hospitals with an all-round production profile will be included; very specialised hospitals are excluded[2]. This concern, 13 hospitals and thus reduces the number of hospitals to 67[3]. In the following I will analyse which effects case-mix adjustment has on the efficiency results. This will be examined by calculating the efficiency scores for the reduced data set using the previous model (DISCH, AMBULA, DOCTOR, NURSE, OCARE, OTHER, BEDS) and the following model (WDISCH, AMBULA, DOCTOR, NURSE, OCARE, OTHER, BEDS). The two sets of efficiency rankings will be compared and the degree of association will be measured through the comprehensive non-parametric Spearman correlation coefficient.

[1] This also implies that the need for case-mix adjustment is smaller due to the more homogeneous hospital sample.

[2] This means hospitals that only perform a single or a few types of treatment, therefore specialisation does not refer to whether a hospital is able to carry out specialised treatments, which is the case for the large hospitals.

[3] However it can be discussed if further reductions of the number of hospitals included should take place concerning some of the smaller hospitals which could be expected to simply transfer difficult patients to larger hospitals. This is a problem since such patients will also be registered as discharges from the smaller hospitals even if they have only contributed to a small extent to the treatment of these patients.

6.3.3 Results.

The efficiency results obtained for the two models with DEA-C are shown in table 11.

	mix	No mix		mix	No mix		mix	No mix
RIGS	0.832	0.813	RONN	0.674	0.743	RIKO	1.000	1.000
SUND	0.698	0.725	SVEN	0.750	0.743	LEMV	0.976	1.000
BISP	0.781	0.781	ODEN	0.821	0.787	SILK	0.763	0.767
HVID	0.755	0.787	NYBO	0.856	0.880	AARH	0.949	0.929
FRED	0.785	0.786	FAAB	0.842	0.850	AHUS	0.961	0.961
GENT	0.998	0.981	SOND	0.771	0.794	RAND	0.785	0.791
GLOS	0.829	0.878	HADE	0.789	0.790	ODDE	0.937	1.000
ELIS	1.000	0.994	TOND	0.899	0.958	GREN	0.942	0.972
HERL	1.000	1.000	AABR	0.897	0.954	TRAN	1.000	1.000
HILL	0.778	0.828	ESBJ	1.000	1.000	VIBO	0.829	0.842
HORS	0.998	0.996	RIBE	0.879	0.899	SKIV	0.771	0.723
FRSU	1.000	1.000	VARD	0.749	0.710	THIS	0.869	0.977
HELS	0.927	0.930	GRIN	0.946	0.981	MORS	0.765	0.707
ROSK	0.896	0.924	BROE	0.930	0.979	KJEL	0.738	0.668
KOGE	0.734	0.733	BRAA	0.887	0.853	LBOR	0.842	0.768
HOLB	0.784	0.798	FRCI	1.000	1.000	JORR	0.703	0.726
SLAG	0.714	0.741	GIVE	0.992	0.933	HOBR	1.000	1.000
KALU	0.820	0.840	HOOR	0.858	0.904	ARSO	0.905	0.906
RING	0.719	0.746	KOLD	0.848	0.838	DRON	0.998	1.000
NAST	0.967	0.952	VEJL	0.797	0.793	HAVN	0.964	0.995
NYFA	0.887	0.962	HOLS	0.820	0.821	SKAG	1.000	1.000
FAKS	1.000	1.000	HERN	0.692	0.729	Mean	0.866	0.876
NAKS	0.838	0.831	TARM	0.910	0.967	Stdev	0.100	0.105

Table 11. DEA-C efficiency results for Danish hospitals with and without case-mix adjustment.

Note: "Mix" are the efficiency results with case-mix adjustment and "No mix" are the results without such an adjustment.

Although the size of the two efficiency scores are different for many hospitals, the general impression is that these are not significant. However even with minor changes in the level of the efficiency score from a model without case-mix adjustment to a model with case-mix adjustment changes can still occur with respect to the efficiency ranking. Analysing the association between the efficiency ranking with and without case-mix adjustment using the

Spearman rank-order correlation coefficient shows that these are indeed strongly associated, since the correlation coefficient is 0.95 which is clearly significant at the 1 per cent significance level.

Thus with the chosen procedure for adjusting output for case mix differences no significant differences between adjusting for case-mix and not adjusting can be revealed. However it should be noticed that this is only one possible way to adjust output for case-mix differences. An alternative could be not to aggregate to a single weighted output measure, but to base the analysis on a larger number of output measures, e.g. the total number of weighted medicine discharges and the total number of weighted surgical discharges. On the other hand the need for case-mix adjustment is less present in the preceding efficiency analysis than in the analysis in section 6.2 due to the more restrictive selection of hospitals giving as such a more homogeneous hospital sample.

6.4 Returns to scale and the most productive scale size (MPSS) in the Danish Hospital Sector.

6.4.1 Introduction.

In many studies of empirical applications of efficiency measurement models a frequent subject has been to characterise the scale properties of the estimated production relations (see e.g. Banker et al. (1986) with US hospitals, Byrnes, Färe, Grosskopf, Lovell (1988) with US coal mines, Färe, Grosskopf & Logan (1985) with US private/public electricity utilities, Ferrier & Lovell (1990) with US banks, Field (1990) with British building societies). Information about the scale properties is of importance because such knowledge can be used in connection with the proper choice of size for a given unit or the resources required if demand increases occur. Examples related to hospitals could be a situation with exogenous increases in demand for hospital treatments where the problem is to determine how many more resources are needed. This depends crucially on the scale properties of the hospital production correspondence.

The advantage of DEA in relation to an analysis of scale properties is the possibility of examining these for different segments of the production relation, that is local scale properties. In contrast standard parametric production frontier models estimate an overall production function, where the returns to scale properties will be the result of averaging over local properties (see Banker et al. (1986))

As described in chapter 3 a range of DEA LP-problems can be used to detect the returns to scale for each observation. This is illustrated in figure 9.

If input efficiency for observation A is measured relative to the DEA-V frontier (a technology with variable returns to scale) then the input efficiency score is equal to $E_1 = \frac{OM}{OA}$.

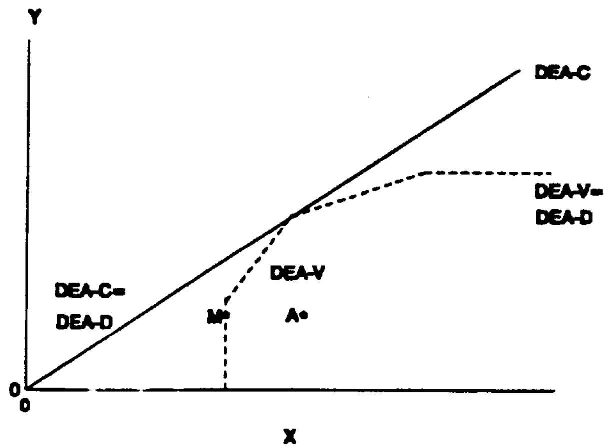

Figure 9. The measurement of scale efficiency for different DEA-models.

Next consider input efficiency measures calculated with respect to the DEA-D frontier (non-increasing returns to scale technology) then the resulting efficiency score for observation A, E_{D1} is equal to or smaller than E_1. If $E_{D1} = E_1$ then non-increasing returns to scale (NIRS) is the case for observation A, otherwise observation A produces at increasing returns to scale (IRS). In order to determine whether an observation with non-increasing returns to scale has decreasing returns to scale (DRS) or constant returns to scale (CRS) a third LP problem has to be solved, where the efficiency measures are calculated with respect to the DEA-C frontier (constant returns to scale technology). If $E_{D1} = E_1 \geq E_3$ then the observation A produces at decreasing returns to scale, otherwise if $E_{D1} = E_1 = E_3$ observation A produces at constant returns to scale, i.e. it is has an optimal scale which maximises the average productivity after any possible technical inefficiency has been removed. In the case of observation A only two steps are needed to determine the returns to scale since $E_{D1} < E_1$ implies that observation A produces at increasing returns to scale (IRS). This three stage procedure can be carried out for all observations and in this way it is possible to divide the hospital data set into 3 groups consisting of hospitals with CRS, hospitals with DRS and hospitals with IRS.

Related to the returns to scale characterisation is the so-called Most Productive Scale Size (MPSS), which has two characteristics. It is a frontier point on the VRS frontier and it maximises the average productivity. This is similar to state that a MPSS is technically efficient with respect to DEA-V, (DEA-D) and DEA-C, that is $E_{D1} = E_1 = E_3 = 1$. The concept MPSS was introduced by Banker (1984) and Banker et al. (1984) within a DEA framework. In these papers a procedure is described for calculating the MPSS for each observation (that is the value the inputs and outputs should have in order for that observation to be a MPSS) based on the sum of the intensity variables δ_k obtained as part of the solution from a DEA problem assuming constant returns to scale. If $\sum_k \delta_k < 1$ then the observation has increasing returns to scale, if $\sum_k \delta_k > 1$ then it has decreasing returns to scale and finally if $\sum_k \delta_k = 1$ then it has constant returns to scale. The MPSS can then be calculated for the k'th observation as $(\frac{E^1}{\sum_k \delta_k} x_k, \frac{1}{\sum_k \delta_k} y_k)$, with the interpretation that the actual input levels should be adjusted for technical inefficiency and the inputs and outputs changed such that optimal scale is achieved. If an observation produces with increasing (decreasing) returns to scale the inputs and outputs should be increased (decreased) to a scale where the production will be with constant returns to scale.

However an important problem of this procedure is that the $\sum_k \delta_k$ is not necessarily unique, making the returns to scale characterisation and MPSS measures uncertain and difficult to interpret, e.g. that for a given observation $\sum_k \delta_k > 1$ while the observation in fact produces with constant returns to scale, see Chang & Guh (1991). This can happen if multiple observations has the same maximum average productivity level with a different MPSS. This case is illustrated in figure 10.

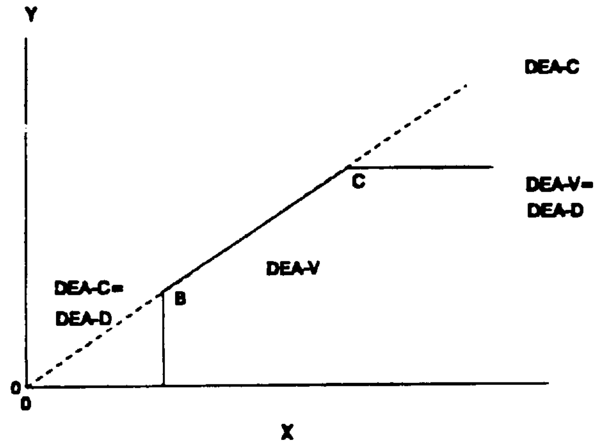

Figure 10. Multiple MPSS's.

All points belonging to the segment BC represent a MPSS. Therefore even if $\sum_k \delta_k \neq 1$ for a given observation, this observation can still be a MPSS. In Banker & Thrall (1992) it is shown however that if an observation which is a MPSS one of the solutions to the DEA problem will be such that $\sum_k \delta_k = 1$. Moreover they consider unlikely the possibility of multiple solutions.

In the following I will examine the returns to scale properties for the Danish hospital sample by analysing which hospitals belong to the sets with constant returns to scale (CRS), decreasing returns to scale (DRS) and increasing returns to scale (IRS). Furthermore I will use $\sum_k \delta_k$ (recalling the uncertainty about this measure) to analyse the relation between returns to scale and hospital size (measured by the number of beds) and thereby determining an indication for the optimal hospital size.

6.4.2 Results.

In the previous section 6.3 the consistency between case-mix adjusted and non-adjusted outputs was examined. In this section I will analyse in more detail the efficiency results obtained with case-mix adjusted outputs, i.e. case-mix adjusted discharges.

Table 12 includes pure technical efficiency measures in inputs (E_1), pure technical efficiency measures in outputs (E_2), gross scale efficiency measures (E_3), the input scale efficiency measure E_4[1]. The $\sum_k \delta_k$ and the returns to scale characterisation for each hospital are shown in table 12a.

The average gross scale efficiency level is equal to 0.866, mainly due to purely technical inefficiency rather than scale inefficiency. The average level of purely technical efficiency in terms of input is 0.906, and is equal to 0.914 in terms of output. The average level of scale efficiency is 0.958 in terms of input and 0.948 in terms of output. Therefore the observations are, in general, closer to the optimal scale than to the DEA-V frontier. The number of gross scale efficient observations is 10 (that is 14.9 per cent of the hospitals are DEA-C efficient). The number of purely technically efficient hospitals is 22 in terms of inputs (corresponding to 32.8 per cent of the hospitals). The number of input scale efficient hospitals is 12, while the number of output scale efficient hospitals is 13 (the output scale efficiency measure is not tabulated). Three hospitals are only scale efficient in terms of either inputs or outputs. This can happen when an observation is not placed on the DEA-V frontier but in the interior of the DEA-V production possibility set since such observations will be projected to different parts of the frontier with possible different returns to scale properties, e.g. an observation can, in terms of input be projected to a segment of the frontier with constant returns to scale, while, in outputs, be projected to a frontier segment with decreasing returns to scale.

The returns to scale characterisation of the hospitals (table 12a) shows that 12 hospitals have CRS, 35 hospitals have DRS and 20 hospitals have IRS. A majority of hospitals (52.2 per cent) produce under decreasing returns to scale, implying that these hospitals should decrease their inputs and outputs in order to reach an optimal scale. If the returns to scale characterisation is related to the number of beds it appears that CRS appears in the range from hospitals with 34 to 627 beds, with an average level of 191.6 beds. Hospitals with IRS are in the range from 73 to 223 with an average equal to 138 beds. Hospitals with decreasing returns to scale appear in a wide range from 269 beds to 1514 beds with an average equal to 531.4 beds. It should be noticed that the reason for the overlapping intervals is because the returns to scale type is related to one input and have not taken into account the level of the other inputs. In general it seems that decreasing returns to scale is the case for hospitals with quite a small number of beds, implying that with the present specification the optimal scale is at a low input level. This aspect is analysed further by relating each hospital's $\sum_k \delta_k$

[1] E_1 and E_2 are the technical efficiency measures obtained with respect to a DEA-V frontier while E_3 is the technical efficiency measure obtained with respect to a DEA-C frontier. See chapter 3 for details.

to the number beds, as shown in figure 11.

The figure shows a positive relationship between the $\sum_k \delta_k$ and beds which is confirmed by the correlation coefficient equal to 0.72. This implies that as the number of beds increases the probability of decreasing returns to scale increases. Moreover, the figure indicates that the optimal scale with $\sum_k \delta_k = 1$ seems to be around 200 beds, which is consistent with the average number of beds for hospitals with CRS equal to 191.6. Although there is a problem with the uniqueness of the $\sum_k \delta_k$ it can be used at least in the present case as an indicator for optimal hospital size.

	$E_{1,DEA-V}$	$E_{2,DEA-V}$	E_3	E_4		$E_{1,DEA-V}$	$E_{2,DEA-V}$	E_3	E_4
RIGS	1.000	1.000	0.832	0.832	VARD	0.750	0.749	0.749	0.999
SUND	0.702	0.704	0.698	0.994	GRIN	0.973	0.974	0.946	0.972
BISP	0.987	0.992	0.781	0.791	BROE	0.958	0.952	0.930	0.971
HVID	0.776	0.819	0.755	0.973	BRAA	0.887	0.887	0.887	1.000
FRED	0.959	0.973	0.785	0.819	FRCI	1.000	1.000	1.000	1.000
GENT	1.000	1.000	0.998	0.998	GIVE	1.000	1.000	0.992	0.992
GLOS	0.978	0.984	0.829	0.848	HOOR	0.965	0.974	0.858	0.889
ELIS	1.000	1.000	1.000	1.000	KOLD	1.000	1.000	0.848	0.848
HERL	1.000	1.000	1.000	1.000	VEJL	0.807	0.836	0.797	0.988
HILL	0.828	0.867	0.778	0.940	HOLS	0.925	0.940	0.820	0.886
HORS	1.000	1.000	0.999	0.999	HERN	0.778	0.834	0.692	0.889
FRSU	1.000	1.000	1.000	1.000	TARM	0.916	0.915	0.910	0.993
HELS	0.935	0.930	0.927	0.991	RIKO	1.000	1.000	1.000	1.000
ROSK	1.000	1.000	0.896	0.896	LEMV	0.988	0.987	0.976	0.988
KOGE	0.736	0.768	0.734	0.997	SILK	0.787	0.859	0.763	0.970
HOLB	0.788	0.817	0.784	0.995	AARH	1.000	1.000	0.949	0.949
SLAG	0.751	0.805	0.714	0.951	AHUS	1.000	1.000	0.961	0.961
KALU	0.821	0.821	0.820	0.999	RAND	0.810	0.848	0.785	0.969
RING	0.724	0.719	0.719	0.993	ODDE	0.946	0.942	0.937	0.990
NAST	1.000	1.000	0.967	0.967	GREN	0.953	0.949	0.942	0.988
NYFA	0.924	0.928	0.887	0.960	TRAN	1.000	1.000	1.000	1.000
FAKS	1.000	1.000	1.000	1.000	VIBO	0.859	0.890	0.829	0.965
NAKS	0.912	0.893	0.838	0.919	SKIV	0.794	0.783	0.771	0.971
RONN	0.684	0.683	0.674	0.985	THIS	0.869	0.892	0.869	1.000
SVEN	0.753	0.785	0.750	0.996	MORS	0.814	0.797	0.765	0.940
ODEN	1.000	1.000	0.821	0.821	KJEL	0.786	0.758	0.738	0.939
NYBO	0.863	0.869	0.856	0.992	LBOR	1.000	1.000	0.842	0.842
FAAB	0.848	0.850	0.842	0.993	JORR	0.750	0.812	0.703	0.937
SOND	0.855	0.889	0.771	0.902	HOBR	1.000	1.000	1.000	1.000
HADE	0.896	0.929	0.789	0.881	ARSO	0.906	0.906	0.905	0.999
TOND	0.902	0.900	0.899	0.997	DRON	1.000	1.000	0.998	0.998
AABR	0.899	0.898	0.897	0.998	HAVN	0.974	0.980	0.964	0.990
ESBJ	1.000	1.000	1.000	1.000	SKAG	1.000	1.000	1.000	1.000
RIBE	0.961	0.955	0.879	0.915					
Mean	0.906	0.914	0.866	0.958					
Stdev	0.098	0.090	0.100	0.056					

Table 12. Efficiency results.

	$\sum_k \delta_k$	RTS		$\sum_k \delta_k$	RTS		$\sum_k \delta_k$	RTS
RIGS	2.288	DRS	RONN	0.840	IRS	RIKO	1.000	CRS
SUND	0.784	DRS	SVEN	0.896	DRS	LEMV	0.880	IRS
BISP	2.431	DRS	ODEN	2.265	DRS	SILK	1.147	DRS
HVID	1.329	DRS	NYBO	1.192	DRS	AARH	2.206	DRS
FRED	1.845	DRS	FAAB	1.486	DRS	AHUS	2.158	DRS
GENT	2.373	DRS	SOND	1.912	DRS	RAND	1.261	DRS
GLOS	3.537	DRS	HADE	1.878	DRS	ODDE	0.812	IRS
ELIS	1.000	CRS	TOND	0.794	IRS	GREN	0.738	IRS
HERL	1.000	CRS	AABR	0.852	IRS	TRAN	1.000	CRS
HILL	1.794	DRS	ESBJ	1.000	CRS	VIBO	1.275	DRS
HORS	0.891	IRS	RIBE	0.331	IRS	SKIV	0.514	IRS
FRSU	1.000	CRS	VARD	0.868	IRS	THIS	1.004	CRS
HELS	0.872	IRS	GRIN	1.407	DRS	MORS	0.420	IRS
ROSK	1.580	DRS	BROE	0.425	IRS	KJEL	0.488	IRS
KOGE	0.947	DRS	BRAA	0.996	CRS	LBOR	1.907	DRS
HOLB	1.441	DRS	FRCI	1.000	CRS	JORR	1.789	DRS
SLAG	1.994	DRS	GIVE	0.469	IRS	HOBR	1.000	CRS
KALU	0.774	IRS	HOOR	1.452	DRS	ARSO	0.726	IRS
RING	0.686	IRS	KOLD	1.841	DRS	DRON	1.633	DRS
NAST	1.449	DRS	VEJL	1.509	DRS	HAVN	1.884	DRS
NYFA	1.884	DRS	HOLS	1.972	DRS	SKAG	1.000	CRS
FAKS	1.000	CRS	HERN	1.346	DRS	Mean	1.277	
NAKS	0.545	IRS	TARM	0.518	IRS	Stdev	0.618	

Table 12a. Returns to scale and $\sum_k \delta_k$.

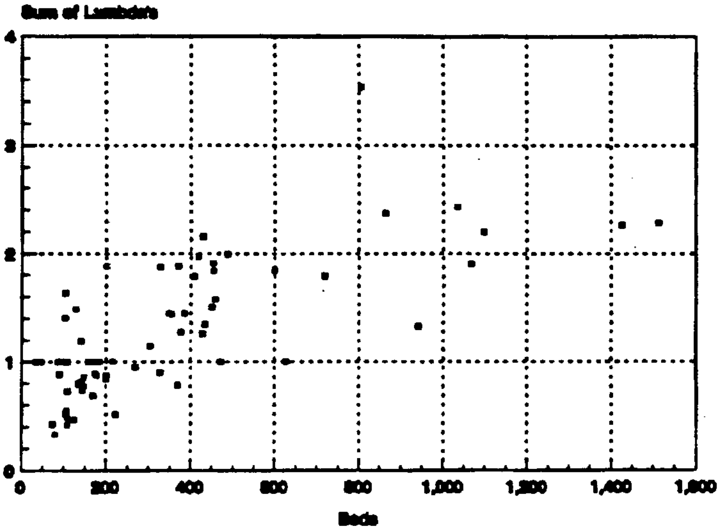

Figure 11. $\sum_k \delta_k$ plotted against number of beds.

6.5 Parametric versus non-parametric production frontier based efficiency measurement methods.

6.5.1 Introduction.

In order to examine the consistency of the efficiency findings based on the non-parametric Data Envelopment Analysis this section will compare these results with efficiency results calculated from a parametric production frontier. Many empirical studies which employ efficiency measurement methods have used either parametric or non-parametric methods whereas comparative studies are more seldom, see e.g. Førsund (1992), Bjurek et al. (1990), Banker et al. (1986) and Deprins et al. (1984) for a few exceptions. The purpose of this section is also related to more general properties of the non-parametric and the parametric efficiency measurement methods.

Chapter 2 included a survey of the parametric approach to efficiency measurement. In this section I will focus on one specific procedure from the parametric approach. I have chosen to examine the so-called Displaced Ordinary Least Squares (DOLS) which includes an explicit efficiency distribution but is deterministic in the sense that all observations are assumed to be on or below the frontier. The reason for this choice is mainly that this non-parametric approach also is deterministic. Furthermore, in order to be able to compare this method with the DEA-model using exactly the same data set I have to perform a new experiment with the DEA-model. This is due to the difficulty of including more than 1 output in the estimation of the parametric production frontier so I cannot use the set-up from sec. 6.3. Therefore, I have chosen to measure the hospital production by the total number of weighted discharges (WDISCH). The inputs are the same as used previously, that is (DOCTOR, NURSE, OCARE, OTHER, BEDS). The parametric frontier is estimated without restrictions on the values of the coefficients. Therefore this frontier is compared to a DEA model with variable returns to scale. If the DEA-C model had been used in the comparison with a parametric model sufficient parameter restrictions would be needed to provide a parametric model with constant returns to scale. Indeed the use of a restriction free parametric production function allows a much more flexible estimation.

The plan for the rest of the section is as follows. In the next section (6.5.2) the chosen parametric model is described in more detail. The results from the parametric model is compared with the DEA-model in section 6.5.3.

6.5.2 The DOLS model.

The introduction of the frontier approach was among, other reasons, the result of a critique of the traditional methods used for estimating production functions. These methods assumed that, on average, the observations are efficient and therefore are distributed randomly around the estimated function. In this case both positive and negative residuals are allowed for. However this property contradicts the theoretical interpretation of the production function showing the maximum attainable production level given the input levels. This inconsistency between the theoretical production function and the traditional empirically estimated production functions is removed in the frontier approach.

There are several methods within the parametric approach regarding frontier estimation (see chapter 2). I will focus in the following on one parametric approach the so-called DOLS method, Førsund et al. (1980). This method utilises the information provided by the average production function parameters but adjusts the constant term such that all residuals are non-positive, i.e. that all observations are on or below the frontier. Notice that this procedure implies that efficiency ranking from the DOLS frontier is identical to the efficiency ranking from the corresponding average production function from the OLS residuals. Only the magnitude of the efficiency measures differ between the two models, see Lovell (1993). Formally the method can be formulated in the case of a Cobb-Douglas production function as follows: in the first step the average CB-function is estimated:

$$\ln y = \ln A + \sum_j \alpha_j \ln x_j + \ln e$$

using standard regression techniques, e.g. OLS. This estimation implies that there are positive as well as negative residuals. In order to obtain only non-positive residuals the constant term is adjusted leaving the α_j unchanged. The constant term is shifted up until no residual is positive and one is zero. The shift factor, β, is determined by the largest positive residual, i.e.

$$\beta = \max \ln e_k, k = 1, ..., n$$

The final model can be written as

$$\ln y = \ln A + \beta + \sum_j \alpha_j \ln x_j + lne - \beta$$

Notice the transformed expression for the residuals, $\ln e - \beta$, which satisfies exactly that all residuals are non-positive. In figure 12 the DOLS method is illustrated.

In Greene (1980) it is shown that $\ln A + \beta$ is a consistent estimate of the constant term and, provided the traditional OLS- assumptions are satisfied, the α_j are estimated consistently. An output efficiency index[1] (in the Farrell tradition) can be constructed as follows:

$$\theta^{PAR} = \frac{\exp^{\ln y}}{\exp^{\ln y'}} = \exp^{(\ln e - \beta)}$$

[1]It should be noted that the above specification implies that the efficiency measure should be specified in terms of output. An efficiency measure specified in terms of input can be obtained in the parametric approach by using an aggregated input as a dependent variable and a set of outputs as independent variables.

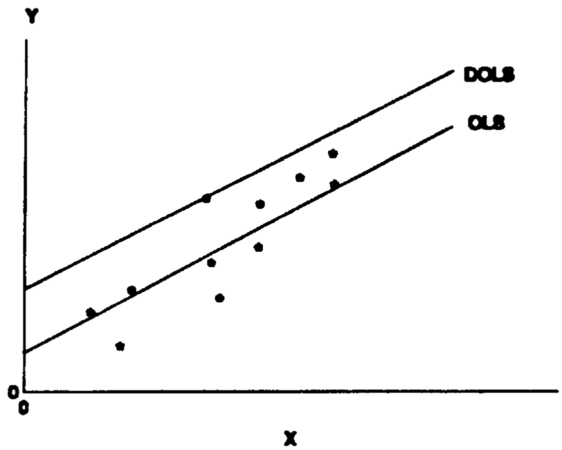

Figure 12. The parametric DOLS method.

where $\ln y'$ is the estimated value of ln y such that $\ln y' = \ln A + \beta + \sum_j \alpha_j \ln x_j$. In the following I will estimate a deterministic Cobb-Douglas production function using the DOLS approach based on the Danish hospital sample. The residuals obtained from the estimated function will be used to calculate efficiency measures for each observation, which will be compared with efficiency measures obtained from the non-parametric DEA-V model in order to examine the consistency between a parametric approach and a non- parametric approach. This comparison will be based on the non- parametric Spearman correlation coefficient.

6.5.3 Results.

The average parametric Cobb-Douglas production function appears as follows:

$$\ln WDISH = 4.724 + .3221 \ln DOCTOR + .558 \ln NURSE + .197 \ln OCARE$$

$$-.236 \ln OTHER + .021 \ln BEDS$$

The DOLS frontier is found by adjusting the constant term with the largest positive residual, which, in the present case, turns to be the residual for Esbjerg Hospital (ESBJ), equal to 0.219. The constant term is then equal to 4.943 = 4.724 + 0.219, while leaving unchanged the coefficients of the independent variables. The residuals are recalculated as $e_k - 0.219$. The Cobb-Douglas production frontier, appears as:

$$\ln WDISH = 4.943 + .322 \ln DOCTOR + .558 \ln NURSE + .197 \ln OCARE$$

$$-.236 \ln OTHER + .021 \ln BEDS$$

Before turning to the analysis of the efficiency results I will comment briefly on the estimated relation. The high $R^2 = 0.985$ is noted, but this property is usually found in estimation of production functions. Moreover, the parameter estimates have, in general, the expected positive sign. The exception is the coefficient of $\ln OTHER$, which is negative. This implies that a 1 per cent in the labour category OTHER will result on average in a .236 decrease in the weighted number of discharges. However, it can be due to this labour category including mainly cleaning personnel and administrative personnel, who are not directly involved in the production of discharges.

The efficiency results from this model will now be analysed. The output efficiency scores based on the adjusted residuals are given in table 13.

	DEA-V	DOLS		DEA-V	DOLS		DEA-V	DOLS
RIGS	1.000	0.800	RONN	0.683	0.711	RIKO	1.000	0.813
SUND	0.704	0.623	SVEN	0.785	0.710	LEMV	0.987	0.785
BISP	0.992	0.751	ODEN	1.000	0.818	SILK	0.859	0.780
HVID	0.819	0.749	NYBO	0.869	0.776	AARH	1.000	0.908
FRED	0.973	0.607	FAAB	0.850	0.738	AHUS	1.000	0.802
GENT	1.000	0.928	SOND	0.889	0.825	RAND	0.848	0.770
GLOS	0.984	0.902	HADE	0.929	0.847	ODDE	0.942	0.910
ELIS	1.000	0.866	TOND	0.900	0.866	GREN	0.949	0.952
HERL	1.000	0.899	AABR	0.898	0.865	TRAN	1.000	0.717
HILL	0.867	0.665	ESBJ	1.000	1.000	VIBO	0.890	0.797
HORS	1.000	0.926	RIBE	0.955	0.811	SKIV	0.783	0.692
FRSU	1.000	0.973	VARD	0.749	0.813	THIS	0.892	0.980
HELS	0.930	0.929	GRIN	0.974	0.815	MORS	0.797	0.773
ROSK	1.000	0.776	BROE	0.952	0.793	KJEL	0.758	0.786
KOGE	0.768	0.672	BRAA	0.887	0.734	LBOR	1.000	0.817
HOLB	0.817	0.823	FRCI	1.000	0.722	JORR	0.812	0.761
SLAG	0.805	0.777	GIVE	1.000	0.754	HOBR	1.000	0.928
KALU	0.821	0.832	HOOR	0.974	0.789	ARSO	0.906	0.867
RING	0.719	0.698	KOLD	1.000	0.771	DRON	1.000	0.774
NAST	1.000	0.911	VEJL	0.836	0.686	HAVN	0.980	0.968
NYFA	0.928	0.876	HOLS	0.940	0.887	SKAG	1.000	0.768
FAKS	1.000	0.764	HERN	0.834	0.737	Mean	0.914	0.808
NAKS	0.893	0.767	TARM	0.915	0.800	Stdev	0.090	0.087

Table 13. Parametric efficiency scores and non-parametric efficiency results.

The mean efficiency level is equal to 0.808 with a standard deviation of 0.087. The smallest efficiency score is equal to 0.607 obtained by Frederiksberg Hospital (FRED). Notice that the DOLS approach restricts the number of efficient observations to a single one, which is obviously identical to the observation with the largest positive residual in the estimated average production function, in the present case Esbjerg Hospital, ESBJ. Comparing these efficiency scores with the DEA-V efficiency scores (also in table 13), the following comments can be made. The mean efficiency levels are different for the non-parametric DEA-V model and the parametric DOLS model, the mean efficiency is 0.914 and 0.808 respectively. Moreover the minimum efficiency level is different with 0.683 for DEA-V (Rønne Hospital, RONN) and 0.607 for DOLS (Frederiksberg Hospital, FRED). FRED is close to being efficient in the DEA-V model with an efficiency score equal to 0.973. Not only are these differences in the efficiency levels, but also in the efficiency rankings. The association between the two rankings is analysed through the computation of the Spearman rank-order correlation coefficient. The Spearman rank-order correlation coefficient is equal to 0.419 which, although significant at a 1 per cent level, indicates a rather weak association between the two efficiency rankings. This clearly suggests that the use of a parametric frontier in efficiency analyses should be imposed with care, since it is likely that the chosen function will involve misspecification of the relation between inputs and outputs. Unless there are specific reasons to believe that the relation has a certain form, e.g. a Cobb-Douglas, a wiser procedure is to base the efficiency analysis on a non-parametric efficiency measurement model such as DEA.

6.6 A non-parametric measure of capacity utilisation.

6.6.1 Introduction.

The reasons for and extent of excess capacity within hospitals has been analysed in several studies. The demand for hospital services is uncertain and therefore gives reasons for maintaining excess capacity as a permanent situation. Even if excess hospital capacity related to costs there is also a benefit related to the social value of having beds available when needed. Excess capacity has been related to non-price competition among hospitals; see Joskow (1980). Excess capacity can be seen as a quality attribute; the hospital with excess capacity can be expected to be capable of meeting the demand for its services to a high extent. This quality approach is in accordance with Newhouse (1970) who proposed a hospital objective function including quantity as well as quality as arguments.

The actual extent of excess capacity[1] and the relation to hospital cost has been analysed in a number of studies but in all cases the analyses have been based on US hospitals, see e.g. Friedman & Pauly (1981), Gaynor & Anderson (1991) and Färe, Grosskopf and Valdmanis (1989). Friedman et al. (1981) and Gaynor et al. (1991) both estimate a parametric cost function and thereby obtain a measure for the cost of empty beds. In Gaynor et al. (1991) the proportion of total hospital costs related to empty beds is reported as 18 per cent. Färe et al. (1989) follow a different approach in a study on capacity utilisation for 39 Michigan hospitals with at least 200 beds. They use a non- parametric DEA model to obtain capacity utilisation measures based on best-practice capacity levels.

I will follow the approach from Färe et al. (1989) (in the following called the FGV procedure) in order to obtain capacity utilisation measures for Danish hospitals. Below I will describe the FGV procedure and then analyse the results obtained.

6.6.2 A DEA-model for measuring the capacity utilisation.

Consider a situation with n observations of outputs, y_k, and inputs, x_k, but the input vector can be divided into fixed and variable inputs, that is $x_k = (x_{kf}, x_{kv})$. The capacity concept is related to the Johansen concept of plant capacity, see Førsund (1987): "The maximum amount that can be produced per unit of time with the existing plant and equipment, provided that the availability of variable factors of production is not restricted". In relation to DEA this implies an output-based efficiency measure, which expresses the maximum proportionate increase in outputs given the technology and the amount of inputs, whereby the maximum output can be determined. In this way a multi output capacity measure has been defined. Therefore, the next step in this approach is related to the definition of fixed and variable inputs, where the capacity utilisation measures concern the fixed inputs. Färe et al. (1989) proceed by formulating two LP problems: one LP-problem where all inputs are assumed fixed, that is a standard DEA model with efficiency measured in terms of output

[1] Capacity is, in general, approximated by the number of hospital beds.

and one where some inputs are fixed and some inputs are variable with only the fixed inputs appearing in this LP-problem. The latter corresponds to the Johansen concept of maximum output with variable inputs that do not restrict outputs[2]. The LP- problem with all inputs fixed appears as:

$$(1)\ F_0(x_{k0}, y_{k0}) = \max \theta$$

$$s.t.$$

$$\sum_k \delta_k y_{ki} \geq \theta y_{k0i}, i = 1, ..., s$$

$$\sum_k \delta_k x_{kj} \leq x_{k0j}, j = 1, ..., m$$

$$\sum_k \delta_k \geq 0$$

where F_0 is the maximum proportionate increase in outputs given the fixed level of inputs. F_0 is the maximum proportionate increase in outputs given the fixed level of inputs. F_0 is greater than or equal to 1 with 1 implying that the unit is on the best-practice frontier while values larger than 1 imply that the unit can increase the level of outputs without increasing the inputs by $(F_0 - 1)$ per cent. The second LP-problem appears as:

$$(2)\ F'(x_{k0f}, y_{k0}) = \max \theta$$

$$s.t.$$

$$\sum_k \delta_k y_{ki} \geq \theta y_{k0i}$$

$$\sum_k \delta_k x_{kf} \leq x_{k0f}, f = 1, ..., F,\ F < m$$

$$\sum_k \delta_k \geq 0$$

$F < m$ insures that not all inputs are fixed, otherwise the two models would yield the same solution, see also footnote 19. The only difference between (1) and (2) is that the inequalities, with respect to the variable inputs, have been deleted since corresponding to the Johansen capacity concept these inputs do not restrict the maximum level of output. (1) and (2) are then solved for each observation. In general $F_0' \geq F_0$ since less restrictions are included in the maximisation of F_0'. FGV constructs a so-called plant capacity utilisation measure for

[2]The first LP-problem would also correspond to the Johansen concept if there are no variable inputs. But in that situation the FGV procedure needs to be reformulated in order to give useful results. In that case it is only necessary to run a single LP-problem in order to determine the capacity utilisation for a given observation.

each observation based on the output efficiency scores from the two LP-problems which is defined as follows:

$$PCU(x_{k0}, x_{k0f}, y_{k0}) = \frac{F_0(x_{k0}, y_{k0})}{F_0'(x_{k0f}, y_{k0})}$$

Since $F_0(x_{k0}, y_{k0}) \geq 1$, $F_0'(x_{k0f}, y_{k0}) \geq 1$ and $F_0' \geq F_0$, $PCU(x_{k0}, x_{k0f}, y_{k0}) \in]0;1]$. PCU measures the increase in outputs that can be obtained when not all inputs are fixed. The measure is adjusted for technical inefficiency and thus indicates the increase in outputs that is possible by increasing the capacity utilisation of the fixed inputs. This is also the reason for the constructed two stage procedure. Otherwise the effects of capacity utilisation on possible output increases will be confused with the presence of technical inefficiency. A low *PCU* value indicates a low capacity utilisation, that is the outputs can be increased by increasing the utilisation rate. If the *PCU* value is equal to 1 then $F_0' = F_0$. This implies that the increases in outputs are identically, independent of whether some inputs are variable or all inputs are fixed, that is the actual pattern of capacity utilisation cannot be changed in order to increase the outputs.

In addition, this procedure is used to calculate the optimal input levels of the variable inputs and to relate these to the actual input levels for each observation. The procedure is the following: the optimal input level for the v'th variable input for the k0'th observation ($v = 1, ..., V$ with $V + F = m$) is defined by multiplying δ_k from (2) with the input level for the k'th observation and then summing over the observations, that is:

$$x_{k0v}^{opt} = \sum_k \delta_k x_{kv}$$

The input utilisation rate for the v'th variable input, IUR, is then defined as:

$$IUR_{k0_v} = \frac{x_{k0v}}{x_{k0v}^{opt}}$$

A value lower than 1 implies that this variable input is underutilised, a value larger than 1 implies that the input is overutilised.

In the following I will analyse the distribution of capacity utilisation and input utilisation rates for the sample of Danish hospitals. I will assume that BEDS is the only fixed input while the different labour inputs are assumed to be variable. However I will follow a stepwise procedure letting more inputs and more inputs be variable. It is indeed possible that some of the labour groups are fixed or quasi-fixed, this seems very likely for the labour input DOCTOR. The sequence in which labour inputs are changed from a fixed to a variable status is the following: (1) OTHER (x4), (2) OTHER (x4), OCARE(x3), (3) OTHER (x4), OCARE(x3), NURSE (x2), (4) OTHER (x4), OCARE (x3), NURSE (x2), DOCTOR (x1). The outputs are, as before, WDISCH and AMBULA. In the following section the results from these models will be analysed.

6.6.3 Results.

In table 14 the output efficiency measures are shown for 4 of the 5 models where the difference between the models is which inputs are regarded as fixed and variable inputs as described above (the results for the model with OTHER and OCARE as variable inputs are not shown. The first column shows the efficiency measure F0 with all inputs included in the DEA problem, that is the results when all inputs are fixed (1). Columns 2-4 contain the efficiency results when some or all labour categories can vary, that is different computations of F_0', where column 4 shows the results when all inputs are variable with the exception of BEDS.

If all inputs are fixed (column 1) the hospitals can increase the outputs on average by 13.4 per cent corresponding to the average efficiency measure of 0.866. Allowing more and more labour input categories to vary implies that the possible output increases become larger and larger: 13.5 per cent if OTHER is a variable input, 15.6 per cent if OTHER and OCARE are variable inputs, 18.4 per cent if OTHER, OCARE and NURSE are variable inputs and 23.8 per cent if all labour categories are regarded as variable inputs. This indicates that excess capacity becomes more present as more inputs do not restrict the output levels. Moreover it seems that the possibilities of output increases due to excess capacity are mainly related to whether DOCTOR and, to a lesser extent, NURSE are allowed to vary, whereas the status of OTHER does not restrict the possibility of output increases. It is indeed open to criticism to regard DOCTOR as a variable input, since the doctors in Danish hospitals in many cases are employed on long-term contracts. The small effect of letting OTHER vary can be related to the type of personnel in this group. OTHER includes mainly non-health care personnel such as cleaning personnel and administrative personnel and these groups have a more indirect relation to the production of discharges. Based on the 5 sets of efficiency measures I have shown in table 15 different values of *PCU*, plant capacity utilisation rate, with each set of *PCU* values reflecting which inputs are regarded as fixed and variable inputs.

Obviously the capacity utilisation rates correspond with the findings in table 14, since these measures are computed as $\frac{F_0}{F_0'}$. Therefore if only OTHER is allowed to vary there is almost full capacity utilisation, 0.999. This reflects that OTHER does not have a large effect on restricting the output level. In this model only 6 hospitals out of 67 have capacity utilisation rates below 1 and all are above 0.97. However this property is changed when allowing additional labour inputs to vary. In the model with OTHER and OCARE regarded as variable inputs the average capacity utilisation is equal to 0.975 with 35 hospitals having rates less than 1. However only 5 of these have rates less than 0.9. Allowing NURSE to be regarded as a variable labour input factor results in the average rate decreasing to 0.943 and only 18 hospitals achieving full capacity utilisation. The major effects are related to letting DOCTOR be a variable labour input. The largest possible output increases are obtained when DOCTOR in addition to the other labour inputs, does not restrict output, see table 14. In this case the average capacity utilisation rate drops to 0.885 and only 6 hospitals have rates equal to 1. Moreover the minimum value drops from 0.972 in the model with OTHER as the only variable inputs to 0.499 in the model where all labour categories are regarded as variable inputs. However the standard deviation of the capacity utilisation rates are all quite low, only in the model where all labour categories are variable inputs is the standard deviation slightly larger than the others, equal to 0.122, but this is still not a very high number.

This kind of information from the DEA model seems to be quite interesting since it is possible to analyse which inputs are restricting the possibility of output increases and it can be computed in addition measures of capacity utilisation rates in cases with multiple outputs, which are adjusted and independent to the possible presence of inefficiency. Moreover Fãre et al. (1989) report a significant correlation between the standard measure for capacity utilisation, occupancy rate[3], and the plant capacity utilisation rate when BEDS is the only fixed input. This is also the case in my analysis of Danish hospitals, where the correlation between the occupancy rate and the plant capacity utilisation rate for the model with all labour inputs regarded as variable is equal to 0.511.

In addition to the computation of plant capacity utilisation rates I have also calculated input utilisation rates (IUR) for the variable inputs in the model with BEDS as the only fixed input. These are depicted in table 16.

A value larger than 1 implies that this input is overutilised. The opposite interpretation holds when a value is less than 1. On average DOCTOR is close to the optimal level, the input utilisation rate is equal to 1.03. NURSE and OTHER are employed on average at too high a level as they obtain input utilisation rates equal to 1.271 and 1.234 respectively. Finally OCARE is underutilised since the average input utilisation rate is equal to 0.886. Only two hospitals (Herlev Hospital, HERL, and Esbjerg Hospital, ESBJ) have all input utilisation rates equal to 1, that is only these hospitals are employing the inputs corresponding to the optimal levels. A further property of these input utilisation rates is that the variation in the rates are much larger for all inputs than when the plant capacity utilisation rates were considered. This was also the case in Fãre et al. (1989). The input utilisation rate for DOCTOR has values ranging from 0.624 to 1.456, NURSE in the range from 0.924 to 1.555, OCARE in the range from 0.465 to 1.496 and OTHER in the range from 0.436 to 2.180. The IUR values for OCARE and OTHER have the largest variation.

However, I will examine in the following a hypothesis to explain the variation in the IUR values for DOCTOR. In the Danish hospital sector it has been claimed that there are difficulties with attracting doctors to hospitals placed in nonurban areas especially in Jutland. I will analyse whether this is reflected in differences in the input utilisation rates for urban and non-urban areas such that DOCTOR tends to be overutilised in urban hospitals and underutilised in non-urban hospitals. Computing the average input utilisation rate for

[3]The occupancy rate is defined as actual patient days to potential patient days with potential patient days defined as $BEDS * 365$, when patient days are measured per year.

DOCTOR shows a large difference with the average input utilisation rate for urban hospitals being 1.161 and for non-urban hospitals being 0.947. This indicates that urban hospitals tend to overutilise DOCTOR while non-urban hospitals tend to underutilise DOCTOR.

I will use the non-parametric Mann-Whitney test to analyse whether this difference is significant. This tests whether the two groups have been drawn from the same distribution. Hospitals in non-urban areas are: the hospitals in Jutland except for Aarhus hospital (AARH) and Aarhus County Hospital (AHUS) and Aalborg Hospital (LBOR), hospitals in Fuen except for Odense hospital (ODEN), and Rønne hospital (RONN). Hospitals in urban areas are the hospitals in Zealand plus AARH, AHUS, LBOR and ODEN. The null hypothesis is that the input utilisation rates are distributed equally between non-urban hospitals and urban hospitals. The alternative hypothesis will be that the input utilisation rates for DOCTORS in urban hospitals are larger than the input utilisation rates for DOCTORS in non-urban hospitals. The Mann-Whitney test statistic z (approximating a normal distributed for large samples) is equal to 4.065 which is much larger than the 1 per cent critical level of 2.576. The conclusion is that the null hypothesis of equal distribution of the input utilisation rates for urban and non-urban hospitals is rejected; the data suggests that the urban hospitals tends to have larger input utilisation rates than non-urban hospitals. A possible reason for this can indeed be the above-mentioned difficulties with attracting doctors to the non-urban hospitals.

	All inp.	x1,x2,x3,x5	x1,x5	x5		All inp.	x1,x2,x3,x5	x1,x5	x5
RIGS	0.832	0.832	0.832	0.832	VARD	0.749	0.749	0.682	0.546
SUND	0.698	0.698	0.644	0.614	GRIN	0.946	0.946	0.782	0.730
BISP	0.781	0.781	0.781	0.754	BROE	0.930	0.930	0.801	0.763
HVID	0.755	0.755	0.755	0.755	BRAA	0.887	0.887	0.753	0.604
FRED	0.785	0.773	0.773	0.742	FRCI	1.000	1.000	1.000	0.925
GENT	0.998	0.998	0.946	0.946	GIVE	0.992	0.992	0.856	0.537
GLOS	0.829	0.829	0.780	0.780	HOOR	0.858	0.858	0.741	0.695
ELIS	1.000	1.000	0.789	0.789	KOLD	0.848	0.832	0.722	0.680
HERL	1.000	1.000	1.000	1.000	VEJL	0.797	0.775	0.726	0.707
HILL	0.778	0.778	0.778	0.747	HOLS	0.820	0.820	0.819	0.815
HORS	0.999	0.999	0.999	0.964	HERN	0.692	0.692	0.691	0.687
FRSU	1.000	1.000	1.000	0.989	TARM	0.910	0.910	0.844	0.619
HELS	0.927	0.927	0.907	0.895	RIKO	1.000	1.000	1.000	0.689
ROSK	0.896	0.883	0.882	0.882	LEMV	0.976	0.974	0.963	0.640
KOGE	0.734	0.734	0.706	0.685	SILK	0.763	0.763	0.721	0.694
HOLB	0.784	0.784	0.776	0.776	AARH	0.949	0.949	0.949	0.949
SLAG	0.714	0.714	0.659	0.658	AHUS	0.961	0.961	0.961	0.895
KALU	0.820	0.820	0.710	0.704	RAND	0.785	0.785	0.784	0.766
RING	0.719	0.719	0.637	0.626	ODDE	0.937	0.937	0.907	0.901
NAST	0.967	0.967	0.966	0.966	GREN	0.942	0.942	0.877	0.864
NYFA	0.887	0.887	0.683	0.679	TRAN	1.000	1.000	1.000	0.499
FAKS	1.000	1.000	0.834	0.834	VIBO	0.829	0.829	0.823	0.802
NAKS	0.838	0.838	0.836	0.833	SKIV	0.771	0.771	0.767	0.762
RONN	0.674	0.674	0.674	0.668	THIS	0.869	0.869	0.824	0.815
SVEN	0.750	0.750	0.739	0.738	MORS	0.765	0.765	0.729	0.691
ODEN	0.821	0.821	0.821	0.820	KJEL	0.738	0.738	0.689	0.685
NYBO	0.856	0.856	0.784	0.694	LBOR	0.842	0.842	0.842	0.842
FAAB	0.842	0.840	0.734	0.703	JORR	0.703	0.703	0.703	0.700
SOND	0.771	0.771	0.757	0.756	HOBR	1.000	1.000	1.000	0.999
HADE	0.789	0.789	0.776	0.647	ARSO	0.905	0.905	0.796	0.792
TOND	0.899	0.899	0.852	0.763	DRON	0.998	0.998	0.772	0.618
AABR	0.897	0.897	0.856	0.774	HAVN	0.964	0.964	0.852	0.721
ESBJ	1.000	1.000	1.000	1.000	SKAG	1.000	1.000	0.768	0.682
RIBE	0.879	0.879	0.841	0.747					
Mean	0.866	0.865	0.816	0.762					
Stdev	0.100	0.101	0.103	0.118					

Table 14. Output efficiency measures when the number of variable inputs changes from 0 to $m-1$.

	x1,x2,x3,x5	x1,x2,x5	x1,x5	x5		x1,x2,x3,x5	x1,x2,x5	x1,x5	x5
RIGS	1.000	1.000	1.000	1.000	VARD	1.000	0.968	0.911	0.729
SUND	1.000	1.000	0.923	0.880	GRIN	1.000	0.840	0.827	0.772
BISP	1.000	1.000	1.000	0.965	BROE	1.000	0.935	0.861	0.820
HVID	1.000	1.000	1.000	1.000	BRAA	1.000	0.875	0.849	0.681
FRED	0.985	0.985	0.985	0.945	FRCI	1.000	1.000	1.000	0.925
GENT	1.000	1.000	0.948	0.948	GIVE	1.000	0.917	0.863	0.541
GLOS	1.000	1.000	0.941	0.941	HOOR	1.000	0.980	0.864	0.810
ELIS	1.000	1.000	0.789	0.789	KOLD	0.981	0.976	0.851	0.802
HERL	1.000	1.000	1.000	1.000	VEJL	0.972	0.972	0.911	0.887
HILL	1.000	1.000	1.000	0.960	HOLS	1.000	1.000	0.999	0.994
HORS	1.000	1.000	1.000	0.965	HERN	1.000	1.000	0.999	0.993
FRSU	1.000	1.000	1.000	0.989	TARM	1.000	0.977	0.927	0.680
HELS	1.000	1.000	0.978	0.965	RIKO	1.000	1.000	1.000	0.689
ROSK	0.985	0.984	0.984	0.984	LEMV	0.998	0.995	0.987	0.656
KOGE	1.000	0.974	0.962	0.933	SILK	1.000	0.992	0.945	0.910
HOLB	1.000	1.000	0.990	0.990	AARH	1.000	1.000	1.000	1.000
SLAG	1.000	1.000	0.923	0.922	AHUS	1.000	1.000	1.000	0.931
KALU	1.000	0.952	0.866	0.859	RAND	1.000	0.999	0.999	0.976
RING	1.000	0.979	0.886	0.871	ODDE	1.000	0.983	0.968	0.962
NAST	1.000	0.999	0.999	0.999	GREN	1.000	0.993	0.931	0.917
NYFA	1.000	1.000	0.770	0.766	TRAN	1.000	1.000	1.000	0.499
FAKS	1.000	0.915	0.834	0.834	VIBO	1.000	0.993	0.993	0.967
NAKS	1.000	0.998	0.998	0.994	SKIV	1.000	0.995	0.995	0.988
RONN	1.000	1.000	1.000	0.991	THIS	1.000	1.000	0.948	0.938
SVEN	1.000	0.985	0.985	0.984	MORS	1.000	1.000	0.953	0.903
ODEN	1.000	1.000	1.000	0.999	KJEL	1.000	0.982	0.934	0.928
NYBO	1.000	0.925	0.916	0.811	LBOR	1.000	1.000	1.000	1.000
FAAB	0.998	0.873	0.872	0.835	JORR	1.000	1.000	1.000	0.996
SOND	1.000	1.000	0.982	0.981	HOBR	1.000	1.000	1.000	0.999
HADE	1.000	0.990	0.984	0.820	ARSO	1.000	0.927	0.880	0.875
TOND	1.000	0.976	0.948	0.849	DRON	1.000	0.797	0.774	0.619
AABR	1.000	0.977	0.954	0.863	HAVN	1.000	0.924	0.884	0.748
ESBJ	1.000	1.000	1.000	1.000	SKAG	1.000	0.825	0.768	0.682
RIBE	1.000	1.000	0.957	0.850					
Mean	0.999	0.975	0.943	0.885					
Stdev	0.005	0.045	0.067	0.122					

Table 15. Plant capacity utilisation rates.

	IUR1	IUR2	IUR3	IUR4		IUR1	IUR2	IUR3	IUR4
RIGS	1.348	1.478	1.292	1.466	VARD	0.972	1.386	0.801	1.792
SUND	1.274	1.512	1.250	1.654	GRIN	0.920	1.246	0.597	1.225
BISP	1.165	1.393	1.330	1.779	BROE	0.878	1.124	0.603	0.961
HVID	1.428	1.533	1.496	1.957	BRAA	0.884	1.436	0.630	1.201
FRED	1.149	1.555	1.132	0.898	FRCI	0.705	1.155	0.564	0.436
GENT	1.291	0.992	1.188	1.132	GIVE	0.730	1.331	0.612	0.839
GLOS	1.274	1.175	1.167	1.291	HOOR	0.985	1.243	0.752	0.934
ELIS	1.201	0.924	0.599	0.837	KOLD	1.026	1.229	0.885	0.790
HERL	1.000	1.000	1.000	1.000	VEJL	1.232	1.288	1.164	1.030
HILL	1.151	1.461	1.150	1.161	HOLS	1.005	1.252	1.046	1.317
HORS	0.790	1.022	0.779	0.819	HERN	1.196	1.445	1.101	1.228
FRSU	0.711	0.983	0.614	0.625	TARM	0.762	1.300	0.727	1.261
HELS	0.837	1.065	0.755	0.998	RIKO	0.624	1.392	0.809	1.373
ROSK	1.239	1.252	1.062	0.907	LEMV	0.650	1.399	0.810	1.311
KOGE	1.232	1.415	0.970	1.241	SILK	1.137	1.321	0.917	1.246
HOLB	1.456	1.257	1.093	1.898	AARH	1.301	1.344	1.294	1.501
SLAG	1.374	1.354	1.170	1.610	AHUS	0.782	1.246	0.799	0.778
KALU	1.048	1.242	0.700	1.513	RAND	1.143	1.352	1.042	1.328
RING	1.116	1.388	0.829	1.168	ODDE	0.868	1.068	0.637	0.958
NAST	1.199	1.085	0.913	1.118	GREN	0.811	1.049	0.646	1.039
NYFA	1.233	1.056	0.977	1.263	TRAN	0.678	1.482	0.769	2.180
FAKS	1.212	1.051	0.540	0.831	VIBO	1.081	1.426	0.938	1.612
NAKS	1.043	1.247	0.782	1.275	SKIV	1.257	1.332	1.179	1.232
RONN	1.099	1.541	0.964	1.561	THIS	0.866	1.128	0.742	1.411
SVEN	1.319	1.435	1.117	1.332	MORS	0.965	1.283	0.826	1.490
ODEN	1.217	1.474	1.367	1.372	KJEL	1.126	1.340	0.806	2.170
NYBO	0.888	1.288	0.672	0.988	LBOR	1.280	1.410	1.357	1.207
FAAB	0.958	1.373	0.642	1.035	JORR	1.177	1.504	1.128	1.573
SOND	1.203	1.274	1.052	1.299	HOBR	0.859	1.123	0.503	1.057
HADE	0.869	1.460	0.854	1.444	ARSO	1.000	1.158	0.605	1.398
TOND	0.811	1.133	0.671	0.973	DRON	0.858	1.400	0.498	1.141
AABR	0.810	1.139	0.672	0.929	HAVN	0.794	1.122	0.592	1.108
ESBJ	1.000	1.000	1.000	1.000	SKAG	0.897	1.188	0.465	1.111
RIBE	0.816	1.116	0.747	1.099					
Mean	1.033	1.271	0.886	1.234					
Stdev	0.211	0.164	0.254	0.340					

Table 16. Input utilisation rates when BEDS is the only fixed inputs.

6.7 An analysis of the efficiency structure for a group of private firms - 18 British hospitals.

6.7.1 Introduction.

This section will examine the efficiency structure for a group of private British hospitals. In this sense the following analyses will be similar to the previous ones on Danish hospitals. However additional aspects will be included. Since the group of British hospitals are privately organised the analysis can be used to shed some light on the efficiency within privately- organised firms. Unfortunately the data on British hospitals includes only private hospitals; if data on public British hospitals were included then more insight could have been obtained with respect to the efficiency advantages of both private and public organisation of production. The only possibility for some insight with respect to this aspect is to compare the efficiency results from the public Danish hospitals with the corresponding efficiency results from the private British hospitals. Certainly such a comparison is problematic due to non-comparabilities between the 2 data sets arising from different environmental conditions. For example private hospitals are, in general, concerned with providing care for profitable cases, while leaving the unprofitable cases to the public hospitals. In the case of public Danish hospitals vs. private British hospitals, this problem is of less importance since the hospitals in Denmark are almost all public such that the case-mix structure is not biased due to the organisational conditions. However this is the case for private British hospitals, which provide care for the easy and profitable case, while leaving the difficult cases to the public British hospitals. Moreover differences appear with respect to the sets of inputs and outputs used in the efficiency analyses of the two hospital samples. It is impossible to combine the two data sets into one comprehensive data set as suggested in chapter 3.

Therefore the efficiency comparison of public vs. private hospitals will be restrained to very general comments based on the Farrell concept of structural efficiency (see Farrell (1957) and chapter 2). The structural efficiency degree measures the overall efficiency within the data set (industry) and gives an indication of the proximity between the industry and the best- practice firms. In Sengupta (1989) it is mentioned that the degree of structural efficiency can be used to compare two or more industries (e.g. industry A and industry B) in order to state that industry A is more structurally efficient than industry B (if A's structurally efficiency degree is larger than B's). This implies that the firms in industry A are closer to its best- practice frontier than for the firms in industry B. Of course it is important to notice that this proximity to best-practice does not necessarily imply that the firms in industry A are more efficient if compared directly with the firms in industry B.

The analysis of the efficiency structure within the sample of private British hospitals will also include a comparison of each hospital's choice of input-output weights (u_i, v_j) (chosen such that the degree of efficiency of a given observation is maximised) in order to analyse the variation in optimal weights. In particular it will be shown that the input and output weights are chosen such that large weights are chosen for outputs which have relatively high levels and for inputs which have relatively low levels (as described in chapter 3).

Moreover the data set will be divided into two subsets corresponding to the location of the hospital, being either in the North or South part of UK. This partitioning of the data set is carried out in order to examine the effects of geographical location on the efficiency results. The idea corresponds to the concepts of programme efficiency vs. managerial efficiency as introduced in Charnes et al. (1981), described in chapter 3, with an example of a possible efficiency comparison of hospitals from different countries (e.g. Danish hospitals vs. Swedish hospitals).

The plan for this section is as follows. In section 6.7.2 the British hospital data for the efficiency analysis is described. The results are presented in section 6.7.3.

6.7.2 The data.

This application of Data Envelopment Analysis is based on data from the General Healthcare Group PLC's Earnings Review on their Acute Care Hospitals for Sept. 1993. These data contain information concerning costs, revenue and activity statistics for 18 acute care hospitals for the period Jan. 1993-Sept. 1993. As it has been seen previously this kind of data can be analysed easily using DEA to detect possible hospital specific inefficiencies. I have chosen to base the main analysis on the following input and output variables:

- x_1: The number of beds (BEDS)
- x_2: Total operating costs (COSTS)
- y_1: Number of equivalent admissions (EQADM)
- y_2: Number of medium difficult operations (SURG2)

The analysis will be based on 2 inputs and 2 outputs, which seems to represent a reasonable number of production categories, compared with the number of observations (18) to maintain the discriminating power of DEA. BEDS is included such that the hospital's capacity is taken into account in the efficiency evaluation. Total operating costs is the sum of: labour, supplies, medical fees and other sources[1]. By using total operating costs I take into account the consumption of all goods including labour. EQADM measures the total number of equivalent admissions, which contains total number of admissions for at least one whole day of hospitalisation plus converted out-patient visits (an out-patient visit is converted to what

[1] Costs with respect to medical fees are from payments to physicians not employed by the hospital. Other costs are mainly overhead costs.

it corresponds to in time in terms of an admission). EQADM covers both a hospital's in-patient care and out-patient care. The other output measure (SURG2) is related to which operations are performed during the hospital stay and measures the number of medium difficult operations. The operations are divided in 5 categories according to the difficulty of the operation: Minor, intermediate, major, major+ and complex major. However the main variation between the hospitals seem to be with respect to this major operations.

In table 17 the data set with some summary statistics are given.

	EQADM	SURG2	BEDS	COSTS
ALEX	7286	1579	164	10504
HIGH	3236	764	54	3269
PARK	4023	784	93	4060
PRIOR	4453	945	101	7079
ROSS	3416	718	92	4929
THOR	1869	518	55	3421
CHAT	615	191	16	803
BLAC	3186	905	64	3729
CHAU	2739	505	58	3237
CHIL	3813	788	73	4233
CLEM	5529	937	114	7558
HARL	3706	633	105	10970
PORT	3573	758	106	7703
PRIG	5328	1014	115	9692
PRIM	3155	783	90	4027
SLOA	2691	574	57	3000
BISH	1120	187	47	1215
RUNN	868	177	42	1344
Max.	7286	1579	164	10970
Min.	615	177	16	803
Mean	3367	709	80	5043
St.dev.	1679	328	35	3144
Median	3326	761	82	4044

Table 17. The 18 private British hospitals.

6.7.3 The results.

Table 18 gives the efficiency results for the 18 British hospitals, with costs and beds as inputs and with equivalent admissions and surgeries of major difficulty as outputs.

	$E_{1,DEA-V}$	$E_{2,DEA-V}$	E_3	E_4	E_5	$E_{1,DEA-D}$	RTS
ALEX	1.000	1.000	0.741	0.741	0.741	1.000	DRS
HIGH	1.000	1.000	1.000	1.000	1.000	1.000	CRS
PARK	1.000	1.000	1.000	1.000	1.000	1.000	CRS
PRIOR	0.857	0.894	0.736	0.859	0.823	0.857	DRS
ROSS	0.700	0.789	0.700	1.000	0.887	0.700	CRS
THOR	0.685	0.666	0.666	0.972	1.000	0.666	IRS
CHAT	1.000	1.000	0.980	0.980	0.980	0.980	IRS
BLAC	1.000	1.000	1.000	1.000	1.000	1.000	CRS
CHAU	0.863	0.855	0.855	0.991	1.000	0.855	IRS
CHIL	0.993	0.995	0.910	0.916	0.915	0.993	DRS
CLEM	1.000	1.000	0.809	0.809	0.809	1.000	DRS
HARL	0.631	0.715	0.589	0.933	0.824	0.631	DRS
PORT	0.593	0.692	0.562	0.948	0.812	0.593	DRS
PRIG	0.952	0.963	0.773	0.812	0.803	0.952	DRS
PRIM	0.823	0.884	0.823	1.000	0.931	0.823	CRS
SLOA	0.913	0.909	0.906	0.992	0.997	0.906	IRS
BISH	1.000	1.000	0.930	0.930	0.930	0.930	IRS
RUNN	0.751	0.704	0.652	0.868	0.926	0.652	IRS
Max.	1.000	1.000	1.000	1.000	1.000	1.000	
Min.	0.593	0.666	0.562	0.741	0.741	0.593	
Mean	0.876	0.893	0.813	0.931	0.910	0.863	
St.dev.	0.145	0.125	0.144	0.080	0.087	0.149	
Median	0.933	0.936	0.816	0.960	0.928	0.919	

Table 18. Efficiency results and returns to scale characterisation for 18 British hospitals.

According to the chosen specification of inputs and outputs the results show that there exists considerable inefficiency within the group of private hospitals, contradicting the general opinion that private organisation per se is efficient while public organisation is inefficient. Even for private firms there are opportunities to change production procedures in order to achieve higher efficiency levels. On the other hand, the measured inefficiency is related to the chosen variable specification and, therefore, the way the production is organised can indeed be optimal for each of the hospitals according to their implicit and explicit goals.

Looking more closely at the results table 18 shows that, on average, the gross scale efficiency measure is equal to 0.812 with a minimum of 0.562. The main part of this inefficiency is caused by pure technical inefficiency (the average pure technical input efficiency measure is equal to 0.876 with a minimum of 0.593) rather than by scale inefficiency (the average scale input efficiency measure is equal to 0.931 with a minimum of 0.809. This suggests that the hospitals are close to the optimal scale on average, but opportunities for reducing the inputs for given output levels (or increasing the outputs for given input levels) are present. The number of gross scale efficient hospitals is 3 out of 18 hospitals, while the number of pure technically efficient hospitals is 7 and that of scale input efficient hospitals is 5 (the number of scale output efficient hospitals is also 5). However as was the case with the Danish hospitals not all hospitals are scale efficient in terms of both inputs and outputs. Only 3 hospitals are scale efficient in terms of both inputs and outputs, BLAC, HIGH and PARK. These hospitals are the ones which are gross scale efficient.

However some of these hospitals can be efficient by default e.g. due to a very special production profile rather than being efficient by dominating other hospitals in terms of using less inputs and/or achieve higher output levels. Simply counting of the number of times each of the efficient hospitals are in the solution base can give an indication of which hospitals are efficient by default and which are really efficient; the maximum corresponds to the number of observations in the sample, in this case 18 times. HIGH achieves the highest number 15, while PARK and BLAC achieve 7 and 3 respectively. In particular the efficiency of BLAC can be claimed to be to self-evaluation, whereas HIGH is efficient to a large degree because it dominates other hospitals.

Moreover, it should further be noticed that the variation in the gross scale efficiency measures and the pure technical efficiency measures are larger than for the sample of Danish hospitals. The scale efficiency measures are more or less at the same level. The larger variation in the private British hospitals efficiency measures compared with the ones for the Danish hospitals is also one explanation behind the result. If the arithmetic average of the inputs and outputs is computed from both the British and Danish sample and included as an additional observation in the Danish and British hospital samples then the average Danish unit obtains higher E_3 and E_1 measures (0.776 and 0.78) than the average British unit (0.699 and 0.715). The only conclusion to be drawn from this finding is that the Danish observations vary less and are closer to the Danish best-practice frontier than is the case for the British hospitals. Later we will examine the possible explanations for this larger variation.

In table 18 the returns to scale characterisation for each hospital is also tabulated. It appears that 5 hospitals (HIGH, PARK, ROSS, BLAC, PRIM) produce at constant returns to scale when the returns to scale characterisation is based on the output vector, that is on the input scale efficiency measure E_4. If the characterisation is based on the output scale efficiency measure E_5, then HIGH, PARK and BLAC will still produce at constant returns to scale. ROSS and PRIM will not, since their output scale efficiency measure is less than 1. The input vectors for these two hospitals are at a level corresponding to decreasing returns to scale. However THOR and CHAU have output scale efficiency measures equal to 1, that is their input vectors are employed at a level according to constant returns scale. The other observations have the same returns to scale characterisation independent of whether the input scale efficiency or the output scale efficiency measure is used. Making the complete scale characterisation for the 18 hospitals with respect to the input scale efficiency measure E_4, 5 hospitals appear (as mentioned) to produce at constant returns to scale; 7 hospitals produce at decreasing returns to scale and 6 hospitals produce at increasing returns to scale. If the returns to scale characterisation is related to the number of beds, the following indication of the variation of the returns to scale with respect to the number of beds can be given. Hospitals with CRS have beds in the range from 54 to 93 with an average level equal to 78.6 beds, hospitals with DRS have beds in the range from 73 to 164 with an average of 111.1, hospitals with IRS have beds in the range from 16 to 58 with 45.8 as an average.

Compared with the analysis of Danish hospitals' returns to scale characterisation it seems that the optimal scale is a much lower number of beds for the private British hospitals than for the public Danish hospitals. The average number of beds for Danish hospitals with CRS was ca. 190, which is more than the double of the average level for the private British hospitals. The results from the returns to scale characterisation are confirmed in figure 13, where the $\sum_k \delta_k$ is plotted against the number of beds.

The figure indicates that the optimal bed size is in the region of 50-90 beds. The reason for the much smaller optimal bed size for the private British hospitals is probably due to a different treatment profile with more weight on treating diseases requiring short periods of hospitalisation, that is the private hospitals focus on cases with relatively less resource demands. In addition these hospitals concentrate on providing a narrow range of treatment, whereas most hospitals in Denmark provide a complete range.

As mentioned in chapter 3 the solution of the DEA problem provides, in addition to a efficiency measure, a set of input- output weights which indicates for each hospital how the inputs and outputs should be weighted in order to maximise efficiency for the hospital. There will be a tendency to put large weights on outputs which the hospital produces at relatively high levels and to put high weights on inputs which the hospital is using in relatively small amounts. This is confirmed in table 19 where the input and output weights are shown together with the ratios equivalent of admissions to surgeries and costs to beds.

As examples, consider the optimal weights for BISH, CHAT and HARL. BISH applies a positive output weight to equivalent admissions (0.00083) and a positive input weight to costs (0.00082), while the other weights are equal to zero. CHAT applies a positive output weight to surgeries of medium difficulty (0.00513) and a positive weight to costs (0.00125). HARL applies a positive output weight to EQADM (0.00016) and a positive input weight to beds (0.00952). These choices can be related to the output ratio $EQADM/SURG2$ and

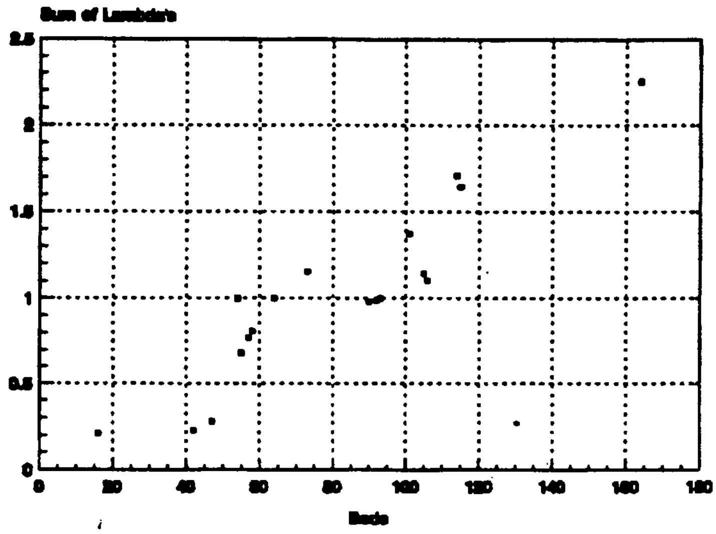

Figure 13. $\sum_k \delta_k$ plotted against number of beds.

the input ratio *COSTS/BEDS*. Thus BISH has the largest ratio of EQADM to SURG2 and the smallest ratio of COSTS to BEDS, which explains that BISH puts a large weight on EQADM and a large weight on COSTS. CHAT has the smallest ratio of EQADM to SURG2, such that it is optimal for CHAT to put a large weight on SURG2. HARL has the largest ratio of COSTS to BEDS and therefore it puts weight to BEDS. In addition its ratio of EQADM to SURG2 is quite large so the optimal for HARL is to put a large weight to EQADM. This relationship between the input ratio and the input weights and between the output ratio and the output weights is further confirmed by computing the correlation coefficients The output weight to SURG2 is negatively correlated with $EQADM/SURG2$ (-0.666) and the output weight to EQADM is positively correlated with $EQADM/SURG2$ (0.538). The same pattern is found with respect to the correlation between the input ratio and the input weights: the input weight to COSTS is negatively correlated with the input ratio $COSTS/BEDS$ (-0.637), while the input weight to BEDS is positively correlated to that ratio (0.522).

In relation to table 18 it was noticed that the variation in the gross scale efficiency measures and the pure technical efficiency had a rather high level. Below I will examine one possible hypothesis for this variation. The hypothesis concerns the effect of geografical location on the efficiency measures. The hospitals are grouped according to region, where the southern

	$EQADM/SURG2$	$COSTS/BEDS$	v_{BEDS}	v_{COSTS}	u_{SURG2}	u_{EQADM}
ALEX	4.614	64.049	0.00610	0.00000	0.00000	0.00010
HIGH	4.236	60.537	0.01852	0.00000	0.00000	0.00031
PARK	5.131	43.656	0.00004	0.00025	0.00000	0.00025
PRIOR	4.712	70.089	0.00990	0.00000	0.00000	0.00017
ROSS	4.758	53.576	0.00003	0.00020	0.00000	0.00020
THOR	3.608	62.200	0.01818	0.00000	0.00129	0.00000
CHAT	3.220	50.188	0.00000	0.00125	0.00513	0.00000
BLAC	3.520	58.266	0.01541	0.00000	0.00110	0.00000
CHAU	5.424	55.810	0.00005	0.00031	0.00000	0.00031
CHIL	4.839	57.986	0.00004	0.00024	0.00000	0.00024
CLEM	5.901	66.298	0.00877	0.00000	0.00000	0.00015
HARL	5.855	104.476	0.00952	0.00000	0.00000	0.00016
PORT	4.714	72.670	0.00943	0.00000	0.00000	0.00016
PRIG	5.254	84.278	0.00870	0.00000	0.00000	0.00015
PRIM	4.029	44.744	0.00000	0.00025	0.00083	0.00005
SLOA	4.688	52.632	0.00005	0.00033	0.00000	0.00034
BISH	5.989	25.851	0.00000	0.00082	0.00000	0.00083
RUNN	4.904	32.000	0.00000	0.00074	0.00002	0.00075
Max.	5.989	104.476	0.01852	0.00125	0.00513	0.00083
Min.	3.220	25.851	0.00000	0.00000	0.00000	0.00000
Mean	4.744	58.850	0.00582	0.00024	0.00047	0.00023

Table 19. The variation in input and output weights with respect to input ratios and output ratios.

region consists mainly of London and the surroundings, while the north region covers the rest. The northern region contains 7 hospitals, while the southern contains 9. Two hospitals could not be placed in either group (BISH and RUNN). The method for comparing two groups of observations was described in chapter 3, but will be shortly summarized here. Three efficiency measures for each hospital are computed ((1) separate efficiency measure, (2) pooled efficiency measure (3) and group efficiency measure), where (1) and (2) are directly computed while (3) is derived from the other two. First the separate efficiency measure is computed for each of the two groups by only including observations from the same group. Then the two groups of observations are pooled to one sample and the pooled efficiency measure is computed. Finally the group efficiency measure is defined as the pooled efficiency measure divided by the separate efficiency measure. Thus the pooled efficiency can be decomposed into separate efficiency and group efficiency. The separate efficiency concerns the technical (managerial) efficiency of an observation irrespective of the group, while the group efficiency is indicating the efficiency caused by belonging to the group.

This partitioning of the overall data sample made it necessary to include a reduced number of variables in the calculation of the efficiency measures in order to obtain an appropriate relation between the number of variables and number of observations. The reduced model included costs and beds as inputs and equivalent admissions as output. Table 20 shows the pooled, separate and group efficiency measures for each hospital for the purely technical input efficiency measure F_1 (DEA-V).

	$E_{1,DEA-V}$	E_s	E_g
ALEX	1.000	1.000	1.000
HIGH	1.000	1.000	1.000
PARK	1.000	1.000	1.000
PRIOR	0.850	0.862	0.986
ROSS	0.700	0.700	1.000
THOR	0.621	0.621	1.000
CHAT	1.000	1.000	1.000
Avg. northern	0.882	0.883	0.998
BLAC	0.864	1.000	0.864
CHAU	0.865	0.994	0.870
CHIL	0.993	1.000	0.993
CLEM	1.000	1.000	1.000
HARL	0.631	0.681	0.927
PORT	0.593	0.656	0.904
PRIG	0.946	0.950	0.996
PRIM	0.793	0.872	0.909
SLOA	0.919	1.000	0.919
Avg. southern	0.845	0.906	0.931

Table 20. The effect of geographical location on efficiency.

The averages (northern and southern) of the pooled efficiency measure are different, the northern hospitals have an average of 0.882 while the southern hospitals have an average of 0.845. This difference is not found for the average separate efficiency measures, where the southern hospitals have an average larger than the nouthern hospitals. The higher average pooled E_1 measure for the northern hospitals is not because of higher average technical efficiency. In fact the southern hospitals are closer to their separate frontier than the northern are. However the average group efficiency measure is larger for northern hospitals than for the southern ones. Therefore, geographical location does have an effect on the efficiency measure. I tested whether these differences were significant using the non-parametric Mann-Whitney test, that is I tested whether the efficiency measures for northern and southern hospitals could be from the same distribution (null hypothesis), see table 21.

	Northern		Southern
E_1	0.882		0.845
Mann-Whitney		0.2039	
E_s	0.883		0.906
Mann-Whitney		0.500	
E_g	0.998		0.931
Mann-whitney		0.0026	

Table 21. Technical efficiency decomposed into pooled, separate and group efficiency, average values.

For both the pooled efficiency measure, E_1, and the separate efficiency measure, E_s, the Mann-Whitney test suggest that the null- hypothesis can be accepted, that is these two efficiency measures for northern hospitals and southern hospitals are drawn from the same distribution. However for the group efficiency measure, E_g, the Mann-Whitney test rejects the null-hypothesis in favour of the alternative hypothesis, that the northern hospitals have larger group efficiency measures than the southern hospitals. The larger variation can be related to this finding that efficiency measures for the northern and southern hospitals are from different distributions.

Chapter 7

Conclusion.

On many occasions clear judgments on the performance of sectors or single organisations are formed without a firm methodological basis. But without the appropriate analysis such a judgment is not valid, it is more of a reflection of given opinions, which then are confirmed. This has often been the case regarding public organised institutions, which are characterised per se as inefficient in comparison with the efficient private sector. In this dissertation I have examined methods that can approach the problem of characterising the performance of organizations and through such methods it can be possible to make value-free judgments. Below the main conclusions from the thesis will be summarised.

In chapter 3 different aspects of non-parametric efficiency measurement methods were discussed. Methods such as DEA and FDH have a clear relation to production theory and therefore these methods are based on well-defined economic concepts. Provided that inputs and outputs can be defined and measured for a set of units these methods can yield measures which indicate the relative efficiency of a given unit compared with the other units. The comparison is based on construction of a benchmark, a production frontier, using the best-practices among the units in terms of maximum output levels for given inputs. The efficiency concept is concerned with to what extent it is possible to increase outputs for given input levels or to decrease inputs for given output levels. However, the methods are however also restricted to consider performance evaluations of similar units, units which can be described with the same inputs and outputs, e.g. comparison of a group of schools, banks or bus companies. Therefore, unfortunately, it is possible to make inter-sector comparisons such as comparing the performance of a school with the performance of a bank. Moreover, the performance evaluation of a group of similar production units is relative and not absolute, even production units which are characterised as efficient are only efficient with respect to the other included production units but are not necessarily efficient if other units are included.

DEA and FDH construct the frontier under different assumptions where FDH is the least restrictive, only free disposability of inputs and outputs is assumed. The choice between

FDH and DEA is far from evident. At least two levels are involved when determining the technology. Whether choosing DEA or FDH depends on the attitude towards the assumption of convexity. Moreover, having accepted convexity which follows from the DEA-models, one has to determine the proper returns to scale. There seems to be some arguments in favour of the unrestrictive formulation of free disposability. However, it is recommendable to consider all various kinds of technologies in order to get a more complete picture of the activities as such.

Another aspect which was discussed in chapter 3 was the problems regarding radial efficiency measures as these do not always provide a true indication of efficiency. However a radial measure is independent of the units of measurement. One solution is to introduce a slack-augmented DEA-model, but the problem with that procedure is that the efficiency measure in such a model is influenced by the units of measurement. Another solution was to use a non-radial efficiency measure such as the one proposed by Fāre and Lovell (1978) which provides correct efficiency indications and is independent of the units of measurement. Moreover, as the whole point of FDH and DEA models is to handle disaggregated activity data it is natural to operate with partial efficiency indices too rather than an over-all radial efficiency measure. However, strictly speaking the Fāre-Lovell efficiency measure is defined as the mean of the partial efficiency scores which makes it difficult to interpret as opposed to Farrell's radial efficiency measure. An alternative could be simply to use a radial efficiency measure, but to report the slacks in inputs and outputs along with it. Efficient units could then be restricted to be those with all slacks equal to zero and the radial efficiency measure equal to 1.

Other aspects of these models which could, at first sight, be seen as problematic were shown not to be serious. Although the standard models assume that inputs and outputs are continuous and adjustable it was shown that DEA (and FDH) could be extended to cover categorical and/or exogeneous variables. Moreover despite that most efficiency studies only included quantitative variables it was discussed how qualitatively oriented variables can be taken into account. One of the main results from an efficiency analysis should be that a ranking of the units is obtained. However as a large proportion of the units might be defined as efficient this is not immediately possible for all units. However a ranking of the efficient units can be obtained in DEA-models with constant returns to scale. Unfortunately, the problem is that this cannot be extended to other DEA models or FDH. It should be noted that even a result which indicate that an observation is efficient regardless of whether it can be stated that the observation is more efficient than other efficient observations is an important result and contains, as such, important information. The chapter also included a section on dynamic aspects of efficiency measurement thereby showing that the standard models with a single period can be extended to several periods. A problem with that extension is that it becomes necessary to distinguish between techhnology changes and efficiency changes. This can be carried out using the Malmquist index approach. However, this approach is only valid for a DEA model with constant returns to scale but not for other DEA models or FDH.

The chapter discussed the possible applications of the results from an efficiency analysis which was concerned with the actual results and the applications of these results. A DEA analysis provides a range of information about the observations included. It is possible to rank the observations and best-performances among these can be defined. Moreover, a complete scale characterisation can be obtained and if the observations included operate with different programs it is possible to analyse whether an observation's inefficiency is due to program or managerial inefficiency. This type of information is relevant for the manager/controller of a group of similar production units as knowledge about the relative performance is provided. The results from an efficiency analysis can also be used as a control instrument such that units with a high efficiency score receive a reward. However, it is necessary to be careful in how the reward system is constructed otherwise the reward system might induce incentives among the units to be "efficient" at the lowest possible work intensity level. The main role an efficiency analysis can play is to increase the information level about the production activity for a group of similar production units.

It is important to remember that technical efficiency indices are only technical of nature i.e. only related to the transformation of inputs into outputs. Hence, when it comes to explaining the technical efficiency result obtained it is necessary also to include institutional (environmental) factors. This can be done through a regression approach as illustrated but obviously there are problems related to the definition of relevant environmental variables. However, these problems are quite general whenever regression approaches are involved.

At present it seems that non-parametric efficiency measurement models have many advantages and few problems. They do not require information about input and output weights and in this way it is possible to measure efficiency when data only is available on quantities of inputs and outputs. This is particularly relevant for public sector institutions where prices often do not exist since prices could have been used as input and output weights. In addition and related to this advantage is the possibility of including multiple inputs and outputs. Moreover, the methods do not require assumption about the functional form of the relation between inputs and outputs. Except for the problem that the evaluation is relative and not absolute the main problem in non-parametric efficiency analysis is the risk of non-comparabilities in the data set e.g. due to too many inputs and outputs included or outliers among the observations. However, this represents more an empirical problem than a methodological one and illustrates the need for careful data selection.

Chapter 4 contained an analysis of the Danish health care sector in order to provide information about the sector chosen for empirical application of efficiency measurement models. It was described that this sector had a more or less constant share compared to the rest of the economy for the years 1970-90. The private-public shares have shown a similar stability with a clear public dominance financed through general county based income taxes. The health care producers are paid differently. Hospitals are budget financed while practioners are in general fee-for-service financed. Moreover, the hospitals are normally public organised and practioners are privately organised. The absence of consumer charges is another characteristic of the Danish health care sector. Looking closer to the hospital structure a main aspect is the county based hospital structure which has the form of several multi-hospital systems. This imply advantages in the organisation of hospital production within a multi- hospital system but difficulties when cooperation between multi-hospital systems is needed. However the fact that the county is the single buyer of hospital services provides strong support for cost control of the hospitals.

The problems of defining and measuring hospital production were analysed in chapter 5 with respect to the problems of defining and measuring hospital outputs. An important problem is that the available output measures are intermediate output measures and not final output measures. Thus the effect of the hospital treatment on the patients' health status is excluded in most hospital performance analyses. Additional studies are called for in order to obtain more valid outputs for the hospital sector. These studies will require interdisciplinary work between medicine and economics.

However even with the kind of hospital data available today it is possible to apply the efficiency measurement models to these data and obtain information about the efficiency structure of hospitals, as illustrated by the range of empirical applications in chapter 6. In addition to the traditional procedures for computing efficiency measures the chapter also included applications of some of the newly-proposed methods, which can be expected to become more important in the future evaluation of the non-parametric approach to efficiency measurement. Efficiency measures for each input are reported along with the radial measure in order to utilise the detailed information that can be extracted from an efficiency analysis if it is not restricted to focus on a single radial efficiency score. Moreover one of the applications focused on establishing interfaces between non-parametric and parametric efficiency measurement methods. This seems a promising way to improve the applicability of the efficiency measurement methods, since such mixed approaches combine the attractive properties of both. The first sections in that chapter included an efficiency analysis of a sample of Danish hospitals. In section 6.2 DEA and FDH were applied on this sample without taking into account possible case-mix differences between the hospitals. A standard specification of hospital production was defined with inputs the number of personnel in 4 different labour categories: DOCTOR, NURSE, OCARE, OTHER. Outputs were specified as number of discharges and number of outpatient visits. This specification showed a large difference between DEA and FDH results. The sensitivity of the results obtained from this standard specification was examined by changing the inputs or outputs included and which hospitals to include. This examination showed that the efficiency results were dependent on the chosen input-output specification as well as which hospitals were included. The information obtained from the different specifications was used to choose which variables were relevant in the explanation of the variation of the variation in the efficiency scores from the standard specification. These variables were then used in a regression model as explanatory variables and the efficiency scores from the standard specification as dependent variables. In this way an interface between non-parametric efficiency analysis and parametric analysis was obtained. The constructed regression model could explain a significant part of the variation in the efficiency scores. In the following section efficiency was analysed with allowance for case-mix. The chosen procedure for adjusting case-mix differences implied only insignificantly differences between a model with case-mix adjustment and one without. This can however be due to the chosen procedure for case-mix adjustments. The next section (sec. 6.4) was mainly concerned with a returns to scale characterisation of the group of Danish hospitals. The largest number of hospitals produce with decreasing returns to scale followed by hospitals with increasing returns to scale and then hospitals with constant returns to scale. The analysis seemed to indicate that the optimal scale is reached for quite small hospitals, measured in number of beds at a level around 200 beds. In section 6.5 a parametric efficiency measurement method was applied to the Danish hospital sample in order to compare these efficiency results to efficiency results from the non-parametric DEA-V model. This comparison indicated a weak association between the two sets of results and suggested the use of a non-parametric approach since the choice of the parametric function is likely to be wrongly specified. The following section (sec. 6.6) examined which inputs (DOCTOR, NURSE, OCARE or OTHER) restricted the outputs and the results seemed to indicate that the outputs mainly were restricted due to DOCTOR. In addition it was analysed whether the inputs were under- or overutilised and the results suggested that NURSE and OTHER were overutilised while OCARE was underutilised, DOCTOR was close to employed in optimal levels. Section 6.7 included an efficiency analysis of a group of private hospitals in the UK. The results showed that there exist considerable inefficiency within this group of hospitals. The returns to scale characterisation indicated that the optimal scale is lower for these hospitals than for the Danish hospitals, in the region of 50-90 beds. The analysis seemed to suggest that efficiency varied due to geographical region with the hospitals in North England being more efficient than the hospitals in South England per se.

Overall it is important to discuss the relevance of analysing hospitals in this way. as the efficiency measurement methods are very much quantitatively oriented in comparison to hospital activity which is more directed towards qualitative areas and indeed difficult to view from a purely quantitative point of view. The risk is that a quantitative evaluation could be unfair to some hospitals since the approach cannot properly capture the qualitative elements involved in the hospital production activities. Moreover one can question whether it is reasonable to consider hospitals as production units which on the basis of a range of inputs produce a range of outputs. The activity within a hospital is perhaps so complex that it is misleading to analyse it using methods based on defining a production process, as the concept of a production process is more directed towards manufacturing industries rather than service industries. Indeed the activity in hospitals provides measurable intermediate outputs such as the number of patient days or the number of operations, but the final outputs such as the extent of improved health status for patients are not easy to measure. The final outputs are only partly dependent on the hospital activity as they are also related to external conditions such as the patient's lifestyle, social class, health status before hospitalisation. However even for hospitals a set of resources can be defined which are transformed to inputs, combined to obtain intermediate outputs which in turn interact with external variables to provide final outputs. Therefore it is possible to define a production process for hospitals. It seems that the difficulties with applying efficiency analyses are not so much conceptual rather they are related to the provision of data.

In general efficiency analyses of hospitals have been restricted to consider efficiency in the production of intermediate outputs, because of the difficulties with measuring final outputs. Therefore, the analyses therefore only evaluate hospital performance based on one part of the production activity. The case with a hospital being inefficient in the production of intermediate outputs but efficient in the production of final outputs cannot be excluded. This is clearly a limitation with many hospital efficiency studies. However it is only a limitation with respect to the concrete studies not with respect to the methods as such, since if data for final outputs were available then the methods could be applied.

Moreover an efficiency analysis that shows that some hospitals are inefficient in the production of intermediate outputs can be used for further studies in order to reveal the reasons for the measured intermediate output inefficiency, e.g. that part of this inefficiency could be due to efficient provision of final outputs where the excessive resources in the production of intermediate outputs are efficiently utilised for the production of the final outputs. This is the case in the hospital study by Sherman (1984), where one hospital was inefficient because of larger staff levels due to a decision at the hospital to provide more individualised patient care (an element that could be expected to have a positive influence on the patients' well-being). Thus efficiency analyses can inform managers about the efficiency consequences of decisions and to make unknown decisions explicit, such as discussions of the type: do we, in fact, want to use resources in a particular way.

The results from hospital efficiency analyses imply in this way an increased information level for a hospital manager or the funding system. In particular concerning the funding system as a decision-maker regarding resource allocation to the hospitals this is relevant as this decision-maker often possess less information about the hospital production production processes than the hospital. Therefore the information from an efficiency analysis can establish a more equal bargaining situation between the hospital and the funding system when deciding on resource allocations. However the efficiency results cannot be used as an automatic device for the magnitude of the budget because the results are subject to uncertainty and should be compared to other information sources. Rather the funding system can use the efficiency results as an instrument for discussions with the hospitals about how the hospital production should be organised.

Therefore, the efficiency measurement methods seem to be promising instrument for hospital planning since the methods can give relevant information about hospitals' activity.

BIBLIOGRAPHY.

Afriat S. N. (1972): "Efficiency estimation of production functions"; International Economic Review, vol. 13, no. 3, p. 568-598.

Aguilar R. (1988): "Efficiency in production: Theory and an application on Kenyan smallholders"; Gothenburg.

Aigner D. J. & S. F. Chu (1968): "On estimating the industry production function"; American Economic Review, vol. 58, no. 4, p. 826-839.

Aigner D., C.A.K. Lovell & P. Schmidt (1977): "Formulation and estimation of stochastic frontier production function models"; Journal of Econometrics, vol. 6, p. 21-37.

Ali A. I. & L. M. Seiford (1993): "The mathematical programming approach to efficiency analysis", in Fried H. O., C.A.K. Lovell & S. S. Schmidt (eds.) "The measurement of productive efficiency: Techniques and applications"; Oxford University Press, Oxford.

Andersen B. (1989): "Kan der foretages produktivitetssammenligninger i sygehusvæsenet - hvad er muligt med de eksisterende data?" (title in Danish) in A. Ankjær-Jensen (ed.) "Produktivitets- og effektivitetsmål i sygehusvæsenet - En status"; DSI-Rapport 89.06.

Andersen P. & N. C. Petersen (1989): "A procedure for ranking efficient units in data envelopment analysis"; Publication from Department of Management, Odense University, No. 11/1989.

Ankjær-Jensen A. (1990): "Skal vi have minimum datasæt for opgørelse af sygehusvæsenets ressourceforbrug" (title in Danish), Tidsskrift for danske sygehuse, 3/90, p. 82-85.

Arrow K. (1963): "Uncertainty and the welfare economics of medical care": American Economic Review, vol. 53, p. 941-973.

Bak P. (1989): "Kvalitetssikring, DRG og klinisk-økonomisk analyse"; (title in Danish) in A. Ankjær-Jensen (ed.) "Produktivitets- og effektivitetsmål i sygehusvæsenet - En status"; DSI-Rapport 89.06.

Banker R. D. (1980): "A game theoretic approach to measuring efficiency"; European Journal of Operational Research. vol. 5, p. 262-266.

Banker R. D. (1984): "Estimating most productive scale size using data envelopment analysis"; European Journal of Operational Research, vol. 17, p. 35- 44.

Banker R. D., A. Charnes & W. W. Cooper (1984): "Models for estimation of technical and scale inefficiencies in Data Envelopment Analysis"; Management Science, vol. 30, p. 1078-1092.

Banker R. D., R. F. Conrad & R. P. Strauss (1986): "A comparative application of data envelopment analysis and translog methods: An illustrative study of hospital production"; Management Science, vol. 32, no. 1., p. 30-44.

Banker R. D. & Maindiratta (1988): "Non-parametric analysis of technical and allocative efficiency in production"; Econometrica, vol. 56, p. 1315-1332.

Banker R. D. & Morey (1986a): "Efficiency analysis for exogeneously fixed inputs and outputs"; Operations Research, vol. 34, no. 4, p. 513-521.

Banker R. D. & Morey (1986b): "The use of categorical variables in Data Envelopment Analysis"; Management Science, vol. 32, no. 12, p. 1613-1627.

Banker R. D. & R. M. Thrall (1992): "Estimation of returns to scale using data envelopment analysis"; European Journal of Operational Research, vol. 62, p. 74- 84.

Barer M. L. (1982): "Case mix adjusment in hospital cost analysis: Information theory revisited"; Journal of Health Economics, vol. 1, p. 53-80.

Barla P. & S. Perelman (1989): "Technical efficiency in airlines under regulated and deregulated environments"; Annals of Public and Cooperative Economics, vol. 60, p. 103-124.

Bauer P. W. (1990): "Recent developments in the econometric estimation of frontiers"; Journal of Econometrics, vol. 46, p. 39-56.

Bay-Nielsen H. & O. Olesen (1993): "Diagnose baseret normalomkostninger for danske sygehuse" (title in Danish), mimeo.

Berg S., F. Førsund, L. Hjalmarsson & M. Suominen (1993): "Banking efficiency in the Nordic countries"; Journal of Banking and Finance, vol. 17, p. 371-388.

Berg S., F. Førsund & E. Jansen (1992): "Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980–89", Scandinavian Journal of Economics, vol. 94, Supplement, p. 211-228.

Berry R. (1967): "Returns to scale in the production of hospital services"; Health Services Research, vol. 2, p. 123-139.

Bjurek H., F. R. Førsund & L. Hjalmarsson (1990): "Deterministic parametric and non-parametric estimation of efficiency in service production -A comparison"; Journal of Econometrics, vol. 46, p. 213-227.

Bjurek H., U. Kjulin & B. Gustafsson (1992): "Efficiency, productivity and determinants of inefficiency at public day care centers in Sweden"; Scand. Journal of Economics, vol. 94, Supplement, p. 173-187.

Blomquist Å. (1992): "Comment on B. A. Weisbrod, "Productivity and incentives in the medical care sector"; Scand. Journal of Economics, vol. 94, Supplement, p. 147-150.

Bogetoft P. (1990): "Strategic responses to DEA-control - a game theoretical analysis"; Working Paper, DASY, Copenhagen Business School.

Bogetoft P., O. Olesen & N. C. Petersen (1987): "Produktivitetsevaluering af 96 danske sygehuse - en præsentation af DEA-metoden og et eksempel pådens anvendelse"; Ledelse og Erhvervsøkonomi, p. 67-81.

Borcherding T. (ed.) (1977): "Budgets and bureaucrats"; Durham, N.C.: University of North Carolina Press.

Boussofiane A., R. G. Dyson & E. Thanassoulis (1991): "Applied Data Envelopment Analysis"; European Journal of Operational Research, vol. 52, p. 1-15.

Bowlin W. F., A. Charnes, W. W. Cooper & H. D. Sherman (1985): "Data Envelopment Analysis and regression approaches to efficiency estimation and evaluation"; Annals of Operations Research, vol. 2, p. 113-138.

Bradley M. & D. Baron (1993): "Measuring performance in a multiproduct firm: An application to the US postal service"; Operations Research, vol. 41, no. 3, p. 450-458.

Breyer F. (1987): "The specification of a hospital cost function"; Journal of Health Economics, vol. 6, p. 147-157.

Breyer F. & F.-J. Wodopia (1987): " The concept of "scale" and related topics in the specification of econometric cost functions: Theory and applications to hospitals" in Eichhorn W. (ed.) "Measurement in economics: Theory and applications of economic indices"; Physica-Verlag, Heidelberg.

Broek J.V.D., F. R. Forsund, L. Hjalmarsson & W. Meeusen (1980): "On the estimation of deterministic and stochastic frontier production functions"; Journal of Econometrics, vol. 13, p. 117-138.

Bruning E. R. & C.H. Register (1989): "Technical efficiency within hospitals do profit incentives matter?"; Applied Economics, 21, p. 1217-1233.

Byrnes P., R. Färe, S. Grosskopf & C.A.K. Lovell (1988):."The effect of unions on productivity: US surface mining of coal"; Management Science, vol. 34, no. 9, p. 1037-1053.

Bös D. (1987): "Public enterprise economics"; North-Holland.

Bös D. (1988): " Introduction: Recent theories on public enterprise economics"; European Economic Review, vol. 32, p. 409-414.

Chang K.-P. & Y. Y. Guh (1991): "Linear production functions and the data envelopment analysis"; European Journal of Operational Research, vol. 52, p. 215- 223.

Charnes A. & W. W. Cooper (1985): "Preface to topics in Data Envelopment Analysis"; Annals of Operations Research, vol. 2, p. 59-94.

Charnes A. & W. W. Cooper (1990): "Data Envelopment Analysis"; Operational Research, vol. 90, p. 641-646.

Charnes A., W. W. Cooper & C. Rhodes (1978): "Measuring the efficiency of decision making units"; European Journal of Operational Research, vol. 2, no. 6, p. 429-444.

Charnes A., W. W. Cooper & C. Rhodes (1979): "Short communication: Measuring the efficiency of decision making units"; European Journal of Operational Research vol. 3, no. 4, p. 339.

Charnes A, W. W. Cooper & C. Rhodes (1981): "Evaluating program and managerial efficiency: An application of Data Envelopment Analysis to program follow through"; Management Science, vol. 27, no. 6, p. 668-697.

Charnes A., W. W. Cooper, B. Golany, L. Seiford & J. Stutz (1985): "Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production function"; Journal of Econometrics, vol. 30, p. 91-107.

Charnes A., W. W. Cooper, Z. M. Huang & D. B. Sun (1990): "Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks"; Journal of Econometrics, vol. 46, p. 73-91.

Charnes A., W.W. Cooper, A.Y. Lewin, R.C. Morey & J.J. Rousseau (1985): "Sensitivity and stability analysis in DEA"; Annals of Operations Research, vol. 2, p. 139-156. *

Christensen F., P. Fristrup & J. Hougaard (1991): "Produktivitetsanalyser" (title in Danish); Jurist- og Økonomforbundets Forlag.

Christiansen T. (1991): "Reformer i England og Holland: Introduktion af interne markeder" (title in Danish) in P. Andersen & T. Christiansen "Styring og regulering påudbudssiden i sundhedssektoren"; Odense Universitetsforlag.

Cornwell C., P. Schmidt & R. C. Sickles (1990): "Production frontiers with cross- sectional and time-series variation in efficiency levels"; Journal of Econometrics, vol. 46, p. 185-200.

Cowing T. G. & A. G. Holtmann (1983): "Multiproduct short-run hospital cost functions: Empirical evidence and policy implications from cross-section data"; Southern Economic Journal, vol. 49, no. 1, p. 637-653.

Culyer A. J. (1989): "The normative economics of health care finance and provision"; Oxford Review of Economic Policy, vol. 5, no. 1, p. 34-58.

Danish Health Institute, DSI (1989): "Små sygehuse" (title in Danish).

Danish Home Office (1986): "Standardomkostninger og produktivitet for 96 somatiske sygehuse" (title in Danish).

Danish Statistical Department: "Statistisk Tiårsoversigt" (title in Danish); several years.

Debreu G. (1951): "The coefficient of resource utilization"; Econometrica, vol. 19, p. 273-292.

de Bruin G. P. (1987): "Economic theory of bureaucracy and public good allocation" in in J.-E. Lane (ed.) Bureaucracy and public choice"; Bristol: Sage Publications.

Deprins D., L. Simar & H. Tulkens (1984): "Measuring labour efficiency in post offices"; in M. Marchand, P. Pestieau and H. Tulkens (ed.), The performance of public enterprises, Amsterdam North-Holland.

Deprins D. & L. Simar (1989): "Estimating technical inefficiency with correction for environmental conditions with an application to railway companies"; Annals of Public and Cooperative Economics, vol. 60, p. 81-102.

Diewert W. E. & C. Parkan (1972): "Linear programming tests of regularity conditions for production functions" in Eichhorn W., K. Hern, K. Neumann & R. Shephard (eds.) "Quantitative studies on production and prices"; Physica-Verlag Wien.

Drummond M. F. (1989): "Output measurement for resource allocation decisions in health care"; Oxford Review of Economic Policy, vol. 5, no. 1, p. 59-74.

Dyson R. G. & E. Thanassoulis (1988): "Reducing weight flexibility in Data Envelopment Analysis"; Journal of Operational Research Society, vol. 39, no. 6, p. 563-576.

Dmitruk A. V. & G. A. Koshevoy (1991): "On the existence of a technical efficiency criterion"; Journal of Economic Theory, vol. 55, p. 121-144.

Enemark U. (1992): "Frit sygehusvalg" (title in Danish); Nationaløkonomisk Tidsskrift, vol. 130, p. 235-244.

Epstein M. K. & J. C. Henderson (1989): " Data Envelopment Analysis for managerial control and diagnosis"; Decision Sciences, vol. 20, p. 90-119.

Evans R. G. (1971): "Behavioural cost functions for hospitals"; Canadian Journal of Economics, vol. 4, p. 198-215.

Farrell M. L. (1957): "The measurement of productive efficiency"; Journal of Royal Statistical Society, vol. 120, A, p. 253-290.

Feldstein M. (1967): "Economic analysis for health service efficiency"; Amsterdam: North-Holland Publishing Company.

Ferrier G. D. & C.A.K. Lovell (1990): "Measuring cost efficiency in banking: Econometric and linear programming evidence"; Journal of Econometrics, vol. 46, p. 229-245.

Fetter R. & J. Freeman (1985): "Diagnosis Related Groups: product line management within hospitals"; The Academy of Management Review, Oct. 1985.

Field K. (1990): "Production efficiency of British building societies"; Applied Economics, vol. 22, p. 415-426.

Frank R.G. (1988): "On making the illustration illustrative: A comment on Banker, Conrad and Strauss"; Management Science, vol. 34, no. 8, p. 1026-1029.

Fried H. O., C.A.K. Lovell & S. S. Schmidt (1993): "The measurement of productive efficiency: Techniques and applications"; Oxford University Press, Oxford.

Fried H. O., C.A.K. Lovell & P. Vanden Eeckaut (1993): "Evaluating the performance of US credit unions"; Journal of Banking and Finance, vol. 17, p. 251-265.

Friedman B. & M. Pauly (1981): "Cost functions for a service firm with variable quality and stochastic demand: The case of hospitals"; Review of Economics and Statistics, vol. 63, p. 610-624.

Färe R. & S. Grosskopf (1983): "Measuring congestion in production"; Zeitschrift für Nationalökonomie, vol. 43, no. 3, p. 257-271.

Färe R. & S. Grosskopf (1983): A non-parametric cost approach to scale efficiency"; Scand. Journal of Economics, vol. 87, no. 4, p. 594-604.

Färe R., S. Grosskopf, B. Lindgren & P. Ross (1989): "Productivity developments in Swedish hospitals: a Malmquist output index approach"; Working Paper.

Färe R., S. Grosskopf & J. Logan (1985): "The relative efficiency of Illinois electric utilities"; Resources and Energy, vol. 5, no. 4, p. 349-367.

Färe R., S. Grosskopf & C.A.K. Lovell (1983): "The structure of technical efficiency"; Scandinavian Journal of Economics, vol. 85, no. 2, p. 181-190.

Färe R., S. Grosskopf & C.A.K. Lovell (1985): "The measurement of efficiency in production"; Kluwer-Nijhoff, Boston-Dordrecht.

Färe R., S. Grosskopf & C.A.K. Lovell (1987a): "Some observations on the new DEA"; Working Paper, Chapel Hill, University of North Carolina.

Färe R., S. Grosskopf & C.A.K. Lovell (1987b): "Scale elasticity and scale efficiency"; Working Paper, Chapel Hill, University of North Carolina.

Färe R., S. Grosskopf & C.A.K. Lovell (1987c): "An indirect efficiency approach to the evaluation of producer performance"; Working Paper, Chapel Hill, University of North Carolina.

Färe R., S. Grosskopf & C.A.K. Lovell (1987d): Non-parametric disposability tests"; Zeitschrift fr Nationalökonomie, vol. 47, no. 1, p. 77-85.

Färe R., S. Grosskopf, C.A.K. Lovell & C. Pasurka (1989): "Multilateral productivity comparisons when some outputs are undesirable: A non-parametric approach"; The Review of Economics and Statistics, vol. LXXI, p. 90-98.

Färe R., S. Grosskopf & D. Njinkeu (1988): "On piecewise reference technologies"; Management Science, vol. 34, no. 12, p. 1507-1510.

Färe R., S. Grosskopf, P. Roos & K. Ödegaard (1992): "Productivity and quality changes in some Swedish pharmacies"; Paper presented at the WEA Conference in San Francisco, 9-13 July 1992.

Färe R., S. Grosskopf & V. Valdmanis (1989): "Capacity, competition and efficiency in hospitals: A non-parametric approach"; The journal of Productivity Analysis, vol. 1, p. 123-138.

Färe R. & C.A.K. Lovell (1978): "Measuring the technical efficiency of production"; Journal of Economic Theory, vol. 19, p. 150-162.

Färe R., C.A.K. Lovell & K. D. Zieschang (1983): "Measuring the technical efficiency of multiple output production technologies" in Eichhorn W., K. Hern, K. Neumann & R. Shephard "Quantitative studies on production and prices", Physica-Verlag: Wien.

Färe R. & D. Primont (1987): "Efficiency measures for multiplant firms with limited data" in Eichhorn W. (ed.) "Measurement in economics: Theory and applications of economic indices"; Physica-Verlag, Heidelberg

Färe R. & L. Svensson (1980): "Congestion of production factors"; Econometrica, vol. 48, no. 7, p. 1745-1753.

Färe R. & G. Whittaker: "An intermediate input model of dairy production using complex survey data"; forthcoming.

Førsund F. R. (ed.) (1987): "Collected works of Leif Johansen"; Amsterdam: North- Holland.

Førsund F. R. (1990): "A comparison of parametric and non-parametric efficiency measures: The case of Norwegian ferries"; The Journal of Productivity Analysis, vol. 3, p. 25-43.

Førsund F. R. & L. Hjalmarsson (1979): "Generalised Farrell measures of efficiency: An application to milk processing in Swedish dairy plants"; The Economic Journal, vol. 89, p. 294-315.

Førsund F. R. & L. Hjalmarsson (1987): "Analyses of industrial structure: A putty-clay approach"; Stockholm: Almquist and Wiksell International.

Førsund F. R., C.A.K. Lovell & P. Schmidt (1980): "A survey of frontier production functions and their relationship to efficiency measurement"; Journal of Econometrics, vol. 13, p. 5-25.

Gathon H.-J. (1989): "Indicators of partial productivity and technical efficiency in the European urban transit sector"; Annals of Public and Cooperative Economics, vol. 60, p. 43-60.

Gaynor M. & G. F. Anderson (1991): "Hospital costs and the cost of empty hospital beds"; NBER Working Paper no. 3872.

Goudriaan R., H. de Groot & F. van Tulder (1986): "Public sector productivity: Recent empirical findings and policy applications" in "The relevance of public finance for policy-making"; Proceedings of 41'st congress of the International Institute of Public Finance, Madrid.

Gramlich E. M. (1981): "Benefit-cost analysis of government programs"; Prentice- Hall.

Green A., C. Harris & D. Mayes (1991): "Estimation of technical inefficiency in manufacturing industry"; Applied Economics, vol. 23, p. 1637-1647.

Greene W. H. (1980): "Maximum Likelihood estimation of econometric frontier functions"; Journal of Econometrics, vol. 13, p. 27-56.

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

Grosskopf S. (1986): "The role of the reference technology in measuring productive efficiency"; The Economic Journal, vol. 96, p. 499-513.

Grosskopf S. & V. Valdmanis (1987): "Measuring hospital performance: a non-parametric approach"; Journal of Health Economics, vol. 6, p. 89-107.

Grossman S. J & O. D. Hart (1983): "An analysis of the principal-agent problem"; vol. 51, no. 1, p. 7-45.

Guesnerie R. & J.-J. Laffont (1984): "The government control of public firms and the economics of incomplete information: An introduction" in M. Marchand, P. Pestieau & H. Tulkens (eds.) " The performance of public enterprises"; North- Holland.

Hanoch O. & M. Rothschild (1972): "Testing the assumptions of production theory: A non-parametric approach"; Journal of Political Economy, vol. 80, p. 256-275.

Hansen P. E. (1993): "Sundhedssektorens økonomi"; Samfundsøkonomen, Oct. 1993:6.

Harris J. (1977): "The internal organisation of hospitals: Some economic implications"; Bell Journal of Economics, p. 467-482.

Hatting A. (1989): "DRG-systemer - beskrivelse og anvendelsesmuligheder" (title in Danish) in A. Ankjær-Jensen (ed.) "Produktivitets- og effektivitetsmål i sygehusvæsenet - En status"; DSI-Rapport 89.06.

Health Systems International (1989): "DRG's: Diagnosis Related Groups: Definition Manual 6"; Revised Edition, New Haven, Connecticut.

Hildenbrand K. (1983): "Numerical computation of short-run-production functions" in Eichhorn W., K. Hern, K. Neumann & R. Shephard (eds.) "Quantitative studies on production and prices"; Physica-Verlag Wien.

Hildenbrand W. (1981): "Short-run production functions based on microdata"; Econometrica, vol. 49, p. 1095-1126.

Hjalmarsson L. (1991): "Metoder i forskning om produktivitet och effektivitet med tillämpninger påoffentlig sektor" (title in Swedish); ESO, Ds 1991:20.

Hougaard J. & M. Tvede (1993): "Intertemporal dominance analysis"; Working Paper 5-93, Institute of Economics - Copenhagen Business School.

Härdle W. (1990): "Applied non-parametric regression"; Cambridge University Press.

Immergut E. M. (1990): "Health care: The politics of collective choice"; Estudio/ Working Paper 1990/5.

Jackson P. M. (1983): " The political economy of bureaucracy"; Totowa: Barnes & Noble.

Jenkins A. W. (1980): "Multiproduct cost analysis: service and case-type cost equations for Ontario hospitals"; Applied Economics, vol. 12, p. 103-113.

Johansen L. (1972): "Production functions"; Amsterdam.

Johnston J. (1984): "Econometric methods"; New York: McGraw-Hill.

Jondrow J., C.A.K. Lovell, I. S. Materov & P. Schmidt (1982): "On the estimation of technical inefficiency in the stochastic frontier production function model"; Journal of Econometrics, vol. 19, p. 233-238.

Joskow P. L. (1980): "The effects of competition and regulation on hospital bed supply and the reservation quality of the hospital"; Bell Journal of Economics, vol. 11, no. 2, p. 421-447.

Kamakura W. A. (1988): "A note on the use of categorical variables in Data Envelopment Analysis"; Management Science, vol. 34, no. 10, p. 1273-1276.

Kittelsen S. (1992): "Stepwise DEA: Choosing variables for measuring technical efficiency in the Norwegian electricity distribution"; Paper presented at the Nordic workshop on Productivity and Growth in Gothenburg 25-27 November 1992.

Kittelsen S. & F. R. Førsund (1992): "Efficiency analysis of Norwegian District courts"; Journal of Productivity Analysis, vol. 3, p. 277-306.

Koopmans T. C. (1951): "Analysis of production as an efficient combination of activities", in Koopmans T. C. (ed.), Activity analysis of production and allocation, New York, Wiley, p. 33-97.

Koopmans T. C. (1957): "3 essays on the state of economic science"; Mcgraw-Hill.

Koopmans T. C. (1977): "Examples of production relations based on micro data" in G. C. Harcourt (ed.) "The microeconomic foundation of macroeconomics"; London.

Kopp R. (1981): "The measurement of productive efficiency: A reconsideration"; Quarterly Journal of Economics, vol. 96, no. 3, p. 477-503.

Kopp R. J. & W. E. Diewert (1982): "The decomposition of frontier cost function deviations into measures of technical and allocative efficiency": Journal of Econometrics, vol. 19, p. 319-331.

Lancaster K. & R. G. Lipsey (1957): "The general theory of second best"; Review of Economic Studies, p. 11-32.

Land K. C., C.A.K. Lovell & S. Thore (1988): "Chance-constrained efficiency analysis"; Working Paper, Chapel Hill, University of North Carolina.

Lane J.-E. (1987): "Introduction: The concept of bureaucracy" in J.-E. Lane (ed.) Bureaucracy and public choice"; Bristol: Sage Publications.

Larsen T. (1990): "Udviklingsfinansiering gennem klinisk produktivitets- udvikling" (title in Danish) in "Struktur og finansiering af det danske sundhedsvæsen i 1990'erne"; Lundbechfondens Prisopgave 1990, Jurist- og Økonomforbundets Forlag.

Lave J. R. & L. B. Lave (1970): "Hospital cost functions"; American Economic Review, vol. LX, p. 379-395.

Lewin A. Y. & R. C. Morey (1981): "Measuring the relative efficiency and output potential of public sector organisations: An application of Data Envelopment Analysis"; International Journal of Policy Analysis and Information Systems, vol. 5, p. 276-285.

Lidegaard ø. & J. P. Steensen (1990): "Sundhedsvæsenet mod år 2000 - kvalitet, differentiering og incitamenter" (title in Danish) in "Struktur og finansiering af det danske sundhedsvæsen i 1990'erne"; Lundbechfondens Prisopgave 1990, Jurist- og Økonomforbundets Forlag.

Lotz J. (1991): "Samspillet mellem stat, amter og kommuner" (title in Danish) in P. Andersen & T. Christiansen "Styring og regulering påudbudssiden i sundhedssektoren"; Odense Universitetsforlag.

Lotz J. (1993): "Rammestyring påsundhedsområdet"; Samfundsøkonomen, Oct. 1993:6.

Lovell C.A.K. (1993): "Production frontiers and productive efficiency" in Fried H. O., C.A.K. Lovell & S. S. Schmidt (eds.) "The measurement of productive efficiency: Techniques and applications"; Oxford University Press, Oxford.

Lovell C.A.K., R. Färe and S. Grosskopf (1990): "Indirect productivity measurement "; Working Paper, Chapel Hill, University of North Carolina.

Lovell C.A.K., A. Sarkar and R. C. Sickles (1987): "Output aggregation and the measurement of productive efficiency"; Working Paper, Chapel Hill, University of North Carolina.

Lovell C.A.K. & P. Schmidt (1979): "Estimating technical and allocative inefficiency relative to stochastic production and costs frontiers; Journal of Econometrics, vol. 9, p. 343-366.

Lovell C.A.K. & P. Schmidt (1988): "A comparison of alternative approaches to the measurement of productive efficiency" in Dogramaci A. & R. Färe "Applications of modern production theory: Efficiency and productivity"; Boston, Kluwer Academic Publishers.

Lovell C.A.K., L. C. Walters and L. L. Wood (1990): "Stratified models of education production using DEA and regression models"; Working Paper, Chapel Hill, University of North Carolina.

Lovell C.A.K. and L. L. Wood (1988): "Monitoring the performance of Soviet cotton refining enterprises: Sensitivity and findings to estimation techniques"; Working Paper, Chapel Hill, University of North Carolina.

Magnussen J. (1992): "Efficiency variations between different types of hospitals"; Paper presented at the Nordic workshop on Productivity and Growth in Gothenburg 25-27 November 1992.

Maindiratta A. (1990): "Largest size-efficient scale and size efficiencies of decision-making units in data envelopment analysis"; Journal of Econometrics, vol. 46, p. 57-72.

Martin S. (1993): "Advanced Industrial Economics"; Basil Blackwell.

Mcguire A., J. Henderson & G. Mooney (1988): "The economics of health care"; London: Routledge & K. Paul.

Mcpherson K. (1990): "Intertemporal differences in medical care practices" in OECD "Health care systems in transition: The search for efficiency"; Paris.

Meeusen W. and J. Broeck (1977): "Efficiency estimation from Cobb-Douglas production functions with composed error"; International Economic Review, vol. 18, no. 2, p. 435-444.

Moene K. O. (1986): "Types of bureaucratic interaction"; Journal of Public Economics, vol. 24, p. 333-345.

Morey R. C., D. J. Fine & S. W. Loree (1990): "Comparing the allocative efficiencies of hospitals"; OMEGA, vol. 18, no. 1, p. 71-83.

Mueller D. C. (1979): "Public choice"; Cambridge: Cambridge University Press.

National Board of Health (1976): "Personale- og økonomistatistik for sygehusvæsenet 1976".

National Board of Health (1991): "Virksomheden ved sygehuse 1989" (title in Danish).

Newhouse J. P. (1970): "Toward a theory of non-profit institutions: An economic model of hospitals"; American Economic Review, vol. 6, no. 1, p. 64-74.

Newhouse J. P. (1992): "Medical care costs: How much welfare loss?"; Journal of Economic Perspectives, vol. 6, no. 3, p. 3-21.

Nielsen V. (1993): "Styring af sygesikringsudgifterne" (title in Danish), Samfundsøkonomen, Oct. 1993:6.

Niskanen W. A. (1971): "Bureaucracy and representative government"; Chicago: Aldine Publishing Company.

Nunamaker T. R. (1985): "Using Data Envelopment Analysis to measure the efficiency of nonprofit organisations: A critical evaluation"; Managerial and Decision Economics, vol. 6, no. 1, p. 50-58.

OECD (1987): "Financing and delivering health care. A comparative analysis of OECD countries"; Paris.

OECD (1993): "OECD Health systems"; Health Policy Studies, no. 3, vol. I and II, Paris.

Olesen O. & N. C. Petesen (1993): "Incorporating quality into Data Envelopment Analysis: A stochastic dominance approach"; Publications from Department of Management, Odense University, no. 2/1993.

Olesen O. & N. C. Petersen (1993): "Indicators of ill-conditioned data sets and model misspecification in Data Envelopment Analysis: Aggregation as a potential remedy"; Publications from Department of Management, Odense University, no. 3/1993.

Pedersen K. Møller (1988): "Diagnoserelaterede grupper DRG" (title in Danish); Odense University (Forskningsrapport no. 7).

Pedersen K. Møller (1990): "Nye økonomiske mekanismer og ændret politisk- administrativ struktur" (title in Danish) in "Struktur og finansiering af det danske sundhedsvæsen i 1990'erne"; Lundbechfondens Prisopgave 1990, Jurist- og Økonomforbundets Forlag.

Pedersen K. Møller (1993): "Sundhedsreformernes fremtid - 4 reformforslag" (title in Danish), Samfundsøkonomen, Oct. 1993:6.

Perelman S. & P. Pestieau (1988): "Technical performance in public enterprises: A comparative study of railways and postal services"; European Economic Review, vol. 32, p. 432-441.

Perelman S. and B. Thiry (1989): "Measuring the performance of public transport companies"; Annals of Public and Cooperative Economics, vol. 60, p. 3-8.

Pestieau P. & H. Tulkens (1993): "Assessing and explaining the performance of public enterprises: Some recent evidence from the productive efficiency viewpoint", Finanzarchiv Neue Folge, vol. 50, no. 3, p. 293-323.

Petersen N. C. (1990): "Data Envelopment Analysis on a relaxed set of assumptions"; Management Science, vol. 36, no. 3, p. 305-314.

Petersen N. C. and O. Olesen (1989): "Chance constraints efficiency evaluation"; Publication from Department of Management, Odense University, No. 9/1989.

Phlips L. (1988): "The economics of imperfect information"; Cambridge, Cambridge University Press.

Premchand A. (1993): "Public expenditure management"; Washington, IMF.

Richmond J. (1974): "Estimating the efficiency of production"; International Economic Review, vol. 15, p. 515-521.

Roos P. (1993): "Three essays on the measurement of productivity changes"; Licentiate Dissertation in Economics at Lund University.

Rousseau J. J. & J. H. Semple (1993): "Notes: Categorical outputs in Data Envelopment Analysis"; Management Science, vol. 39, no. 3, p. 384-386.

Rousseeuw P. J. & A. M. Leroy (1987): "Robust regression and outlier detection"; New York: Wiley.

Russell R. R. (1985): "Measures of technical efficiency"; Journal of Economic Theory, vol. 35, p. 109-126.

Russell R. R. (1987): "On the axiomatic approach to the measurement of economic efficiency" in Eichhorn W. (ed.) "Measurement in economics: Theory and applications of economic indices"; Physica-Verlag, Heidelberg.

SAS Institute (1985): "SAS user guide on statistics".

Schieber G. J. & J.-P. Poullier (1990): "Overview of international comparisons of health care expenditures" in OECD "Health care systems in transition: The search for efficiency"; Paris.

Schmidt P. (1976): "On the statistical estimation of parametric frontier production functions"; Review of Economics and Statistics, vol. 58, no. 2, p. 238-239.

Schmidt P. (1985-86): "Frontier production functions"; Econometric Reviews, vol. 4, p. 289-328.

Schmidt P. and T. Lin (1984): "Simple tests of alternative specifications in stochastic frontier models"; Journal of Econometrics, 13, p. 349-361.

Schmidt P. and C.A.K. Lovell (1979): "Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers"; Journal of Econometrics, vol. 9, p. 343-366.

Schmidt P. and C.A.K Lovell (1980): "Estimating stochastic production and cost frontiers when technical and allocative inefficiency are correlated"; Journal of Econometrics, vol. 13, p. 83-100.

Schmidt P. and R. Sickles (1984): "Production frontiers and panel data"; Journal of Business and Economic Statistics, vol. 2, p. 57-66.

Seiford L. M. (1990): "A bibliography of Data Envelopment Analysis (1978-1990)"; Department of Industrial Engineering and Operations Research, The University of Massachusetts, Amherst, MA, USA.

Seiford L. M. and Thrall (1990): "Recent developments in DEA"; Journal of Econometrics, vol. 46, p. 7-38.

Sengupta J. K. (1986): "Efficiency measurement in nonmarket systems"; in Stochastic optimization and economic models, Dordrecht: D. Reidel Publishing.

Sengupta J. K. (1987): "Production frontier estimation to measure efficiency: A critical evaluation in light of Data Envelopment Analysis"; Managerial and Decision Economics, vol. 8, p. 93-99.

Sengupta J. K. (1988): "A robust approach to the measurement of Farrell efficiency"; Applied Economics, vol. 20, p. 273-283.

Sengupta J. K. (1989): "Efficiency analysis by production frontiers - the non- parametric approach"; Kluwer Academic Publishers, London.

Sengupta J. K. (1990): "Transformations in stochastic DEA-models"; Journal of Econometrics, vol. 46, p. 109-123.

Sengupta J. K. (1990a): "Testing Farrell efficiency by stochastic dominance"; Economic Notes, vol. 3, p. 429-439.

Sengupta J. K. and R. E. Sfeir (1988): "Efficiency measurement by data envelopment analysis with econometric applications"; Applied Economics, vol. 20, p. 285-293.

Sherman H. D. (1984): "Hospital efficiency measurement and evaluation"; Medical Care, vol. 22, no. 10.

Siegel S. & M.J. Castellan (1988): "Non-parametric statistics"; McGraw-Hill, NY.

Silverman B. W. (1990): "Density estimation for statistics and data analysis"; London, Chapman & Hall.

Simar L. (1991): "Estimating efficiencies from frontier models with panel data: A comparison of parametric, non-parametric and semiparametric methods with bootstrapping, CORE Discussion Paper No. 9126.

Simon H. A. (1965): "Administrative behaviour"; FP.

Sundhedsstyrelsen: "Personale- og økonomistatistik"; several years.

Sundhedsstyrelsen: "Virksomheden ved sygehuse"; several years.

Søgaard J. (1991): "Internationale sammenligninger af sundhedsudgifter og deres finansiering" (title in Danish) in P. Andersen & T. Christiansen "Styring og regulering påudbudssiden i sundhedssektoren"; Odense Universitetsforlag.

Sørensen R. J. (1987): "Bureaucratic decision-making and the growth of public expenditure" in J.-E. Lane (ed.) Bureaucracy and public choice"; Bristol: Sage Publications.

Tatchell M. (1983): "Measuring hospital output: A review of the service mix and case mix approaches"; Social Science Medicine, vol. 17, no. 13, p. 871-883.

Thanassoulis E, A. Boussfiane & R. G. Dyson (1991): "DEA models for the assessment of the provision of perinatal care in England on the basis of quantity and quality of outputs"; Warwick Business School Research Papers no. 26.

Thiry B. and H. Tulkens (1988): "Allowing for technical inefficiency in parametric estimates of production functions"; CORE Discussion Paper 8841.

Thiry B. and H. Tulkens (1989): "Productivity, efficiency and technical progress. Concepts and measurement."; Annals of Public and Cooperative Economics, vol. 60, p. 9-42.

Timmer C. P. (1971): "Using a probabilistic frontier production function to measure technical efficiency"; Journal of Political Economy, vol. 79, p. 776-794.

Torup J. (1991): "Udbudsregulering ved hjælp af DRG" (title in Danish) in P. Andersen & T. Christiansen "Styring og regulering påudbudssiden i sundhedssektoren"; Odense Universitetsforlag.

Tulkens H. (1990): "Non-parametric efficiency analyses in four service activities: Retail banking, municipalities, courts and urban transit"; CORE Discussion Paper 9050.

Tulkens H. (1993): "On FDH efficiency analysis: Some methodological issues and applications to retail banking courts and urban transit"; Journal of Productivity Analysis, vol. 4 (1/2), p. 183-210.

Tulkens H. (1993a): "Efficiency dominance analysis: A frontier-free efficiency evaluation method"; Paper presented at the Third European Workshop on efficiency and productivity measurement CORE 21-23 Oct. 1993.

Tulkens H. & P. Vanden Eeckaut (1991): "Non-frontier measures of efficiency, progress and regress"; CORE Discussion Paper 9155.

Vallgårda S. (1992): "Sygehuse og sygehuspolitik i Danmark. Et bidrag til det specialiserede sygehusvæsens historie 1930-1987"; København. DJØF's Forlag.

Valdmanis V. (1992): "Sensitivity analysis for DEA models. An empirical example using public vs. NFP hospitals"; Journal of Public Economics, vol. 48, p. 185-205.

Vanden Eeckaut P., H. Tulkens & M-J. Jamar (1993): "A study of cost-efficiency and returns to scale for 235 municipalities in Belgium"; chapter 12 (pp. 300-334) in Fried H. O., C.A.K. Lovell & S. S. Schmidt (eds.) "The measurement of productive efficiency: Techniques and applications"; Oxford University Press, Oxford.

Varian H. (1984): "Microeconomic analysis"; New York, Norton and Co.

Varian H. (1984a): "The non-parametric approach to production analysis"; Econometrica vol. 52, p. 579-597.

Westerberg, I. (1989): "Technical efficiency in public dental care in Sweden"; Paper presented at the workshop "Efficiency and productivity measurement in the service industries" at CORE, Oct. 20-21, 1989.

Weisberg S. (1980): "Applied linear regression"; New York, Wiley.

Weisbrod B. A. (1991): "The health care Quadrilemma: An essay on technological change, insurance, quality of care and cost containment"; Journal of Economic Literature, vol. 29, p. 523-552.

Weisbrod B. A. (1992): "Productivity and incentives in the medical care sector"; Scand. Journal of Economics, vol. 94, Supplement, p. 131-145.

Zieschang K. D. (1984): "An extended Farrell technical efficiency measure"; Journal of Economic Theory, vol. 33, p. 387-396.