Ole. B. Olesen and Niels. C. Petersen

The Use of Data Envelopment Analysis with Probabilistic Assurance Regions for Measuring Hospital Efficiency

First Draft May 1997
Revised July 1998
Abstract.

This paper uses Data Envelopment Analysis (DEA) for an estimation of the cost efficiency of 70 Danish hospitals. The analysis relates to a cost function based on 483 outputs in combination with a set of probabilistic assurance regions defined by the cost distributions for each output. It is demonstrated that the probabilistic assurance region approach allows for i) a frontier estimation in the full output space, i.e. no fixed aggregation is required, and ii) a controlling of the variation in heterogeneity of the output clusters, in casu Diagnosis Related Groups. The likelihood of the estimated efficiency score for a given hospital can be measured based on the sensitivity of the score w.r.t. the probability levels used in the specification of confidence intervals for the probabilistic assurance regions.

1. Introduction.

Inefficiency is widely conjectured to be an important contributor to the cost of the hospital sector. Unfortunately, efficiency measurement within this sector is not straightforward because it is difficult to measure the outputs and because hospitals may not be homogeneous with respect to the types of outputs actually produced.

Hospitals produce to an extreme degree multiple output-services using multiple inputs. The ultimate product is ambiguous but has of course to do with the specific goods and services given to patients. These goods and services comprise x-ray, medication, lab-tests, nursing care, operating room facilities, services from the various physicians, hotel services, and social services. Many such goods and services are produced at ancillary departments producing services for the bed departments. The individual patient sent to the hospital in a given condition is often used as the entity for definition of the multidimensional product. The services provided to this patient are seen as intermediary products used as inputs to provide the final output: the treatment and discharge of the patient with a certain combination of illnesses. The diagnoses and some of the treatment characteristics are in combination with other information often used for a distinction between different types or clusters of treated patients.

The hospital sector is characterized by a complicated structure. Hospitals have different mixes of medical specialties, they are of different sizes, and the product as defined by the number of discharges within different clusters of patients sharing some clinical and socioeconomic characteristics is often of varying homogeneity regarding expected consumption of resources per patient. The sector often is characterized by a size hierarchy, where

i) larger hospitals have many very specialized high tech and expensive treatments,
ii) hospitals of medium size have many difficult treatments of more common diagnoses, and
iii) small hospitals are primarily dealing with rather trivial procedures and may
function as convalescence centers for local patients initially sent to larger hospitals located at some distance from their home area.

Observe that the larger Danish hospitals also serve as local hospitals besides being characterized by high tech treatments of more uncommon specialties. Hence, a substantial amount of activities in larger hospitals relate to treatments requiring rather trivial procedures.

Most methods for efficiency measurement or frontier estimation require the existence of many similar Decision Making Units (DMUs), i.e. many organizations or productive units with similar structure, each of which using the same types of inputs to produce the same types of well defined homogeneous outputs. It is not straightforward to justify the fulfillment of these requirements within the hospital sector. The characteristics of this sector and the problems hereby created for an efficiency analysis are the primal concern of Newhouse (1994) in his comments on the usefulness of frontier estimation. It is argued that the efficiency of a hospital sector cannot be measured with sufficient accuracy to warrant its use in decision making because

i) it is difficult to measure the outputs of the sector,
ii) several inputs including capital cannot be measured and must for this reason be omitted,
iii) measures for case mix are questionable since one cannot expect the variation within a cluster to be random by hospital,
iv) strong and non-testable assumptions concerning noise in data or the distribution of inefficiency must be made in an implementation of the available methods for frontier estimation, and
v) a parametric approach to frontier estimation based upon econometrics is almost impossible without a drastic aggregation because of lack of degrees of freedom.

Omitted outputs are considered almost certain since the increase in hospital cost during the past 30 years is of a size which cannot be justified by an increased need for hospital treatments or a drastically increase in case mix. The omission of quality dimensions is also considered a problem. Newhouse (1994) concludes: "I am doubtful that regulators can recover 'true' or efficient cost or production parameters from observed data with any degree of precision".

We agree that the combination of the state of the art of efficiency analysis and the complexity of the performance of hospitals implies that the validity of frontier analyses within the hospital sector is questionable and subject to discussion. However, we are also of the opinion that an ongoing development of the methodological framework for efficiency measurement along with the incorporation of more information related to the activities of hospitals is the more appropriate approach for improving the validity when applying frontier analyses. An extension of the DEA framework combined with the incorporation of the distributional characteristics of the (empirical) cost distributions for

---

1 Quality may be even more difficult to measure than output itself, but the need for quality control can be argued to be conditional on the purpose of the analysis. An introduction of quality measures in an efficiency analysis tends to reorient focus from efficiency towards effectiveness.
each cluster of patients is suggested in this paper. This construction improves upon the measurement of outputs and avoids a fixed aggregation of clusters of patients.

Inefficiency within a sector of an economy is measured by frontier estimation techniques. Let the sector be made up of a number of individual firms or DMUs termed DMU\(_j\), \(j = 1, \ldots, n\). For each DMU\(_j\) its consumption of inputs \(X_j\) used for its production of outputs \(Y_j\), \(j = 1, \ldots, n\), has been observed. The geometric locus of efficient production can be represented by a production frontier to be estimated based upon the set of observed input-output configurations \((X_j, Y_j)\), \(j = 1, \ldots, n\). Efficiency analysis is concerned with the identification of the production frontier as well as the question whether a given DMU is to be termed efficient or not.

The available set of techniques for efficiency analysis comprises two approaches, the parametric and the non-parametric approach. The parametric approach has focused on the development of stochastic frontier models which allow for random noise around the estimated frontier. Output is in the case of one output and multiple inputs predicted from inputs by a functional relationship \(Y_j = f(X_j; \theta) + \epsilon_j\), \(j = 1, \ldots, n\). The residual \(\epsilon_j\) is a so-called composed error term, i.e. \(\epsilon_j = v_j + u_j\), where \(v_j\) is a symmetric noise residual and \(u_j\) is a non-positive inefficiency residual. The expression of the frontier output \(Y_j = f(X_j, \theta) + v_j\), \(j = 1, \ldots, n\), reveals that the frontier itself is stochastic. The parametric model can be estimated by a maximum-likelihood technique when a parametric functional form for \(f(\cdot; \theta)\) is specified and specific distributional assumptions on the residuals are invoked. Major limitations of the parametric approach are implied by i) the restrictiveness and ultimate untestability of the parametric forms chosen for the production function and ii) the employed distributional assumptions on the inefficiency residual \(u_j\). In addition, a practical implementation of the stochastic frontier model in the case of multiple inputs as well as multiple outputs is not easy as witnessed by Kumbhakar (1996).

The non-parametric approach has focused on the development of multiple inputs and multiple outputs models based upon the concept of DEA developed in the seminal paper by Charnes et al. (1978 & 1979). The DEA approach relies on a specification of nonparametric functional forms and is for this reason termed non-parametric. The production frontier is determined by a piecewise linear envelopment of observed input-output configurations for all DMUs in combination with assumptions concerning returns to scale and disposability of inputs and outputs. The envelopment surface is referred to as the empirical production frontier and used as a benchmark for efficiency evaluations of the DMUs in the sample. Following Schmidt (1985), the production frontier estimated by DEA belongs to the class of deterministic frontier functions, because the frontier is bounded by a deterministic function. Most models within this family of frontier functions do not allow for random noise in data, with Banker and Maindiratta (1992) as an important exception. The implicit assumption of no random noise in data is an important limitation of this type of model.

Both approaches for frontier estimation have been applied within the area of health care as witnessed by a number of recent papers, see for instance Chilingarian & Sherman (1996), Clement et al. (1996), Morey & Dittman (1996), Ozcan et al. (1996), and Salinas-Jiménez & Smith (1996) in the special issue of Annals of Operations Research vol. 67
(1996) on operations research and health care as well as Kooreman (1994), Vitaliano & Toren (1994), and Zuckerman et al. (1994) in the Journal of Health Economics vol.13 (1994). These authors advocate the usefulness of frontier estimation or efficiency measurement within the area of health economics. On the other hand, Dor (1994), Newhouse (1994), and Skinner (1994) argue in a series of comments on the above papers in the Journal of Health Economics vol.13 (1994), that efficiency cannot be measured with sufficient accuracy to warrant its use in decision making. It is not the intention to comment further on that debate in this paper. Suffice it to say that we are of the opinion that frontier estimation is a potentially useful tool, also within the health care sector, in spite of the inherent difficulties, including measurement of inputs and outputs within that specific area. And that the development of new theory is the most constructive approach for addressing shortcomings in available models.

This paper addresses the issue of aggregation of outputs in an estimation of a cost frontier for the hospital sector and the issue of homogeneity of each output. As observed by Newhouse (1994), hospitals are multi-output firms. Case-mix controls usually by Diagnosis Related Groups (DRGs) involve around 500 such groups which are themselves aggregated measures of outputs. It is in view of this relatively large number of groups not surprising that existing studies have aggregated (both inputs and) outputs drastically, see for instance Granneman et al. (1986), Grosskopf & Valdmanis (1987), Vita (1990), Vitaliano (1987), and Wagstaff (1989). Newhouse (1994) observes that the resulting aggregation emphasizes how the estimated inefficiency may represent measurement errors from inappropriately aggregating (inputs and/or) outputs. One of the main contributions to be made in the context of the present paper is that DEA allows for an estimation of a production frontier in an output space which involves all DRGs explicitly along with a number of additional outputs. The DEA approach is for this reason used to derive hospital-specific measures of inefficiency for the Danish hospital sector.

The empirical production frontier to be used as a benchmark for efficiency evaluations of the DMUs in a sample is in DEA determined by a piecewise linear envelopment of observed input-output configurations for all DMUs in combination with some additional assumptions. The piecewise linear envelopment of data is reflected by the fact that the inefficiency of any given DMU can be estimated by the solution of a linear optimization problem. Duality theory for linear programming implies in turn that the linear optimization problem either relates to the input-output space or equivalently to a corresponding price space. The formulation of the problem in the price space allows for an easy incorporation of the relative (random) cost information offered by the weighting scheme of the DRG system. A model with cost information incorporated as a number of side constraints allows for an efficiency analysis in an output space of dimension say 500 or larger. Another of the main contributions to be made in the context of the present paper is that DEA allows for the incorporation of information on random cost characterizing each output cluster. Incorporation of random cost allows for the controlling of variation in heterogeneity of the output-clusters, i.e. the DRGs. Hence, the likelihood of the efficiency score from each hospital can be measured based on its sensitivity with respect to the probability levels used in the specification of confidence regions for the probabilistic assurance regions.

DEA has also been criticized by Newhouse (1994) for not allowing for measurement errors or random fluctuations in data. Four observations are appropriate concerning this
viewpoint. First, Banker and Maindiratta (1992) combines a "non-parametric functional form" enveloping observed data points with a composite error term and a maximum likelihood estimation based on specific distributional assumptions on the error terms. Second, only measurement errors or errors in variables regarding the dependent variable are allowed for in the standard stochastic frontier analysis.\footnote{Kmenta (1971) chapter 9 distinguishes between error terms representing errors in equation and errors in variables.} Consider a cost efficiency analysis with cost explained by services from the hospitals in a number of clusters. There is no easy way to incorporate measurement errors in the many output clusters neither in stochastic frontier analyses nor in DEA-like analyses. Third, a substantial body of recent work has opened ways to exploring the statistical significance of DEA estimators of production frontiers, estimation of bias, estimates of the bias corrected confidence intervals for the efficiency scores, estimation of sample distributions and rates of convergence, see Simar and Wilson (1995), Wilson and Simar (1995), Kneip and Simar (1995), Simar (1995), Kneip et al. (1996), Banker (1995), Banker and Chang (1995), Kittelsen (1995) and Grosskopf (1995). Most of this work is based on an assumption of no error in data and no error in equations with only one random one-sided residual representing inefficiency included. However, there is a general consensus that it is important to extend these models to allow for errors in equations or variables and an important first step is made by the development of a statistical foundation. Fourth, DEA models with stochastic inputs and outputs based upon the theory of chance constraints developed by Charnes et al. (1963) are available, see Land et al. (1993), Olesen & Petersen (1995), and Cooper et al. (1997).

Inputs and outputs are in principle considered non-stochastic in the model to be suggested below. However, the cost information offered by the DRG system is considered random. The random cost information is incorporated into a formulation of the DEA model in the price space and subsequently handled by a specification of a set of side constraints as probabilistic constraints or chance constraints.

The paper unfolds as follows. The model for efficiency evaluation of hospitals is developed in Section 2 with all technical details relegated to an appendix. The use of the DRG system on Danish data as well as the design of the empirical analysis is discussed in Section 3. Numerical results are presented in Section 4. Finally, Section 5 concludes.

2. Development of the DEA model for hospital efficiency evaluation

The DEA models and the necessary model extensions to allow for probabilistic constraints on the relative prices are presented in Appendix A1. These model extensions are presented more thoroughly in Olesen and Petersen (1998).

A decision has to be made on which inputs and outputs to include in the efficiency analysis and how to measure these inputs and outputs. Basically, inputs can be measured by observed cost for each hospital and outputs by the number of discharges in each group of discharges to be defined by some classification system. The measurement of outputs is of particular importance in this respect since a control for case mix is required. The DRG-system is one of a number of classification systems available and appropriate for this purpose. An explicit connection between the choice of measurement of outputs and the
structure of the probabilistic constraints is a new feature of the models used in this study compared to previous DEA-analyses. The probabilistic constraints incorporate the relative heterogeneity of the chosen measurement of outputs into the efficiency analysis.

The DRG-system involves an aggregation of individual discharge-related clinical data into about 500 aggregated groups. The main requirements concerning the aggregation of clinical data into aggregated groups are that the resulting groups should be

i) clinical meaningful,
ii) homogeneous with respect to use of resources, and
iii) of a reasonable number.

The resulting grouping can be seen as an operational compromise between conflicting requirements and is subject to ongoing refinements and tuning. It is not the intention to discuss the comparative advantages of the DRG-system vs. other classification systems in the current context. Suffice it to say that the DRG-system offers a possible platform for an efficiency analysis of the hospital sector. Clearly, the validity of the results to be reported is for obvious reasons conditional upon the validity of the underlying classification system.

Existing studies have been criticized for aggregation of outputs beyond the DRG-system, and large variations in treatment cost within aggregated groups of patients have typically been ignored. A DEA model based on so-called probabilistic assurance regions\textsuperscript{3} offers an approach which avoids this aggregation and at the same time incorporates the relative variation in treatment cost within different DRGs. It is a flexible aggregation that is proposed. Probabilistic constraints can be used to incorporate the relative heterogeneity of the DRGs into the efficiency analysis. Hence, the approach resolves some of the criticized characteristics of existing efficiency analyses.

The DRG-system allows for an easy identification of upper and lower bounds on the virtual output multipliers needed in the AR-approach for a hospital efficiency evaluation since relative average cost estimates for the treatment of the average patient in all DRGs are available. Deterministic assurance regions in the output space can for instance be specified as follows:

\[
\alpha_{r_1,r_2}^{u} \equiv .9 \times \frac{\overline{C}_{DRG_{r_1}}}{\overline{C}_{DRG_{r_2}}} \leq \frac{u_{r_1}}{u_{r_2}} \leq 1.1 \times \frac{\overline{C}_{DRG_{r_1}}}{\overline{C}_{DRG_{r_2}}} \equiv \beta_{r_1,r_2},
\]

\[r_1, r_2 = 1, \ldots, s, r_1 < r_2\]  \hspace{1cm} (1)

where $\overline{C}_{DRG_r}$ is the $r$'th DRG-weight (the average cost of a discharge in DRG$_r$). This formulation implies that the relative DEA-prices are not allowed to deviate more than 10\% in either one direction from the corresponding relative average cost estimates provided by the DRG-weights.

\textsuperscript{3} The notion of assurance regions was originally proposed by Thompson et al. (1986) as constraints in the form of upper and lower bounds on relative input and/or output multipliers in DEA. An introduction to assurance regions is in Appendix.
Assume that we have a stochastic benchmark vector \( \mathbf{e} \equiv [e_1, \ldots, e_s]^\top \) available, where each component \( e_i, \, i = 1, \ldots, s \) is a random variable representing random cost in DRG_1. These \( s \) random variables are distributed according to some multivariate distribution. Consider some set of realizations of the random vector \( \mathbf{e} \) that is regarded as having a high likelihood (it could for example be a confidence region). We want to restrict all relative output prices to belong to the set of ratios of all realized components in this random vector \( \mathbf{e} \) constrained to this set of realizations of high likelihood.

The vector of all median cost estimates or all average cost estimates for all DRGs are point estimates with neighborhoods of high likelihood. Hence, we may choose to restrict the output multipliers to be equal to either the median or the mean cost estimates. Moreover, available data allows for an estimation of the variation of cost in each of the DRGs by estimating upper and lower endpoints for confidence intervals concerning cost. Thus, the DRG-system allows for the following specification of probabilistic assurance regions in the output space:

\[
\begin{align*}
  u_r &\leq k \chi_r^+(\alpha) & r = 1, \ldots, s \\
  u_r &\geq k \chi_r^-(\alpha) & r = 1, \ldots, s
\end{align*}
\]  

(2)

where \( k \) is a non-negative scalar and the \((100 - 2\alpha)\) percent confidence interval for cost for DRG_\(r\) is \([\chi_r^-(\alpha), \chi_r^+(\alpha)]\).\(^4\) Observe that the formulation is fairly compact and reflects the stochastic nature of the cost estimates provided by the DRG-system. The model to be used can now be formulated as follows:

---

\(^4\) In the main text we specify confidence intervals at probability level \((100-2\alpha)\) percent, with equal probability mass \( \alpha \) cutoff two sided. The approach in appendix is more general with lower and upper confidence interval endpoint \( \kappa^-(\alpha^-), \kappa^+(\alpha^+) \) characterized by the following:

- the probability of \( u_r \) being greater than or equal to \( \kappa^-(\alpha^-) \) is equal to \( \alpha^- \).
- the probability of \( u_r \) being less than or equal to \( \kappa^+(\alpha^+) \) is equal to \( \alpha^+ \).

Clearly, (2) can be obtained by rewriting (A4) in Appendix A1 with \( \chi_r^-(\alpha) = \kappa^-(\alpha^-), \chi_r^+(\alpha) = \kappa^+(\alpha^+) \) and \((100-2\alpha) = (\alpha^* - \alpha^-), r = 1, \ldots, s.\)
\[
\max \sum_{r=1}^{s} u_r y_{rj_0} \tag{3.1}
\]
\[
\text{s.t.} \quad \sum_{r=1}^{s} u_r y_{ij} - v x_j \leq 0 \quad j = 1, \ldots, n \tag{3.2}
\]
\[
v x_{j_0} = 1 \tag{3.3}
\]
\[
u_r - k \chi^+_r (\alpha) \leq 0 \quad r = 1, \ldots, s \tag{3.4}
\]
\[
u_r - k \chi^-_r (\alpha) \geq 0 \quad r = 1, \ldots, s \tag{3.5}
\]
\[
k \geq 0, v \geq 0, u_r \geq 0 \quad r = 1, \ldots, s \tag{3.6}
\]

The working of the model can be demonstrated by the following simple example with 4 hospitals each of which produces two outputs, discharges in DRG_L and DRG_H using identical cost. The number of discharges produced by each hospital in each DRG is listed in Table 1:

<table>
<thead>
<tr>
<th>Hospital</th>
<th>#discharges, DRG_L</th>
<th>#discharges, DRG_H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>B</td>
<td>40</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>50</td>
<td>5</td>
</tr>
<tr>
<td>D</td>
<td>60</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1. Data for illustrative example.

To make this illustrative example realistic we assume that the cost in DRG_L (DRG_H) \( C_L \) (\( C_H \)) are random variables distributed according to some skew distributions. For illustrative purposes we assume two log-normal distributions. Let \( E(\log(C_L)) = \log(1) \) and \( E(\log(C_H)) = \log(10) \), i.e. median value of cost of a discharge in DRG_H is ten times the median cost in DRG_L. Furthermore, assume that \( \text{Var}(\log(C_L)) = \text{Var}(\log(C_H)) = 0.5 \). Then \( \text{Var}(C_L) \approx 1.07 \) and \( \text{Var}(C_H) \approx 107 \), i.e. the variance of the cost of discharges in DRG_H is hundred times as large as the variance of the cost of discharges in DRG_L.\(^5\) The cost of discharges in DRG_H is relatively high and much more uncertain compared the cost of discharges in DRG_L.

\(^5\) \( E(\log(C_L)) = \log(1) \) and \( E(\log(C_H)) = \log(10) \) implies that the median values are

\[
\text{Median}(C_L) = 1 \text{ and } \text{Median}(C_H) = 10.
\]

The mean values are \( \bar{C}_L = \text{Med}(C_L)e^{\text{Var}(\log(C_L))} = 1e^{10.5} \approx 1.28 \), and \( \bar{C}_H = \text{Med}(C_H)e^{\text{Var}(\log(C_H))} = 10e^{10.5} \approx 12.84 \). Notice, that the mean value is always greater than then median value but the difference is asymptotically equal to zero for \( \text{Var}(\log(C)) \to 0 \).
The shape of the empirical length of stay distributions in the different DRGs resembles in many cases the log-normal distributions. Hence, a log-normal is a realistic illustration of the cost distributions. However, in the actual efficiency analysis we have used a more flexible parametric family of distributions, namely the truncated normal distribution with endogenous truncation point. Figure 1 depicts these two illustrations of log-normal cost distributions with marks indicating the median cost, the mean cost and the $(100 - 2\alpha)$ percent confidence interval for $\alpha = 30$ percent.

Figure 1. The two log-normal cost distributions with A, B, C and D indicating the median cost, the mean cost and the endpoints of the $(100 - 2\alpha)$-confidence interval for $\alpha = 30$ percent. logn1 and logn2 are cost distributions for DRG_L and DRG_H. A, B, C, and D are located close to each other in logn1, because the variance of cost of discharges in DRG_L is small compared to the variance of cost of discharges in DRG_H; however, the relative positions of A, B, C, and D are the same in the two distributions which is the reason why dots are not labeled in logn1. Observe that the mean cost (indicated by B) in both cases
exceeds the median cost (indicated by A). It is a well known phenomenon in asymmetrical distributions that the mean tends to be located somewhat away from the center of the distribution. Thus, the mean cost in $logn1$ ($logn2$) is only exceeded in about 36% of the cases. The median of the distribution has a more central position in the sense that it divides the distribution into two equal parts. Thus, the median cost is in both $logn1$ and $logn2$ exceeded in exactly 50% of the cases.

Consider first the situation with no variation in prices allowed, i.e. the ratio between a pair of virtual DEA-multipliers must equal the ratio between the corresponding pair of mean or median costs.\footnote{In Figure 2 we have chosen the median values 1 and 10 to restrict the virtual DEA multipliers. Alternatively, one could have chosen the mean values 1.28 and 12.84, i.e. a ray with a slope equal to 0.0997.} This situation is illustrated in Figure 2.

![Diagram](image)

Figure 2. Probabilistic productivity indices with relative prices required to be identical to the relative median costs.

The cone of feasible multipliers is in this case the single ray with slope .10 since no deviation from the relative median costs is allowed; this cone is shown in the orthant in the upper right segment of the figure. Hence, the marginal rate of substitution between DRG$_L$ and DRG$_H$ is 10:1. Let $y_{rj}$ denote the number of discharges in DRG$_r$, $r = \{L, H\}$ from hospital $j$, $j = \{A, B, C, D\}$. The output possibility set is defined by points in the positive orthant below the hyperplane

$$\mathcal{H} = \{(y_L, y_H) : y_L = 100 - 10.0 \times y_H\}$$

since the marginal rate of substitution between DRG$_L$ and DRG$_H$ as defined by the AR is 10:1 and since B and C have been observed to produce (40, 6) and (50, 5) discharges in...
DRG<sub>L</sub> and DRG<sub>H</sub>, respectively. Hospitals B & C are termed efficient and A and D inefficient with indices $\frac{OA}{OA} \approx 0.9$ and $\frac{OD}{OD} \approx 0.9$.

Consider next the situation with $\alpha = 30\%$, i.e. 30% of the probability mass is cut off in both sides for $C_L$ and $C_H$. We get<sup>7</sup>

$$\chi_L^i(30\%) = \exp\left(\log(1) + \Phi^{-1}(\beta)\sqrt{\text{Var}(\log(C_L))}\right), \beta = \begin{cases} 30\% & \text{if } i = - \\ 70\% & \text{if } i = + \end{cases}$$

$$\chi_H^i(30\%) = \exp\left(\log(10) + \Phi^{-1}(\beta)\sqrt{\text{Var}(\log(C_H))}\right), \beta = \begin{cases} 30\% & \text{if } i = - \\ 70\% & \text{if } i = + \end{cases}$$

This case is illustrated in Figure 3.

![Figure 3. Probabilistic productivity indices with 40% confidence intervals for $C_L$ and $C_H$, i.e. 30% of the probability mass is cut off in both sides.](image)

The cone of feasible multipliers is in this case the one shown in the orthant in the upper right segment of the figure. This cone is generated from the Cartesian product of the confidence intervals (the intervals from C to D in Figure 1), i.e. approximately the area $[0.7,1.45] \times [7,14.5]$. Deviations from the relative benchmark prices corresponding to marginal rates of substitution between approximately $\frac{1.45}{7} : 1$ and $\frac{1.45}{0.7} : 1$ are now allowed.

<sup>7</sup> This follows from

$$\chi_L^i(30\%) = z, \text{ where } \Phi\left(\frac{\log(z) - \log(C_L)}{\sqrt{\text{Var}(\log(C_L))}}\right) = \Phi\left(\frac{\log(0.25) - \log(1)}{0.25}\right) = 30\%$$

$$\chi_H^i(30\%) = z, \text{ where } \Phi\left(\frac{\log(z) - \log(C_H)}{\sqrt{\text{Var}(\log(C_H))}}\right) = \Phi\left(\frac{\log(0.25) - \log(1)}{0.25}\right) = 70\%$$
The output possibility set is now defined by points in the positive orthant below the three hyperplanes drawn with bold lines since B and C have been observed to produce (40, 6) and (50, 5) discharges in DRG_L and DRG_H, respectively, and since the marginal rates of substitution are as defined by the imposed ARs. Observe that the output possibility set has become smaller compared to Figure 2 because the restriction imposed by the ARs has been relaxed. Hospitals B & C are still termed efficient and A and D inefficient but now assigned higher indices OA_A > 0.95 and OD_D > 0.95. A further relaxation of the imposed ARs implies a further reduction of the possibility set. At some point all hospitals are termed efficient since a larger variation in relative prices is allowed.

3. Using the DRG-system on Danish data

Implementation of model (3) requires the availability of data on the confidence intervals 
\[ \chi_r^-(\alpha), \chi_r^+\alpha), r = 1, \ldots, s. \] The fulfillment of these data requirements is the issue to be addressed next.

We consider the DRG-system appropriate in the context of the current study because its basic validity has been demonstrated in a number of studies and because the system provides the information needed for an implementation of model (3). Clinical data for each patient discharged from a Danish hospital in a given calendar year are available in the so-called LPR-database. These data are designed to be used for clinical purposes and are much too disaggregated to be used directly in an efficiency evaluation since they describe diagnoses and procedures related to the treatment of individual inpatients. An aggregation of individual discharges by means of a classification system is therefore required. The DRG-system is available for an efficiency evaluation of Danish hospitals, because an assignment to a DRG is possible for each discharge in the LPR-database.

The estimation of the relative average cost for each DRG, i.e. the components in the vector of Danish DRG-weights, is performed as follows. The computer code for assigning discharges in the LPR-database to one of the DRGs relates to a version of the DRG-system which includes a total of 471 groups. The cost component, \( \bar{W}_r \), corresponding to a given DRG_r, \( r = 1, \ldots, 471 \), measures the ratio between the average cost, \( \bar{C}_r \), for a discharge for patients in this group, and the average cost, \( \bar{C} \), for discharges in all DRGs. Hence,

\[ \bar{W}_r = \frac{\bar{C}_r}{\bar{C}} \]  

(4)

Let \( j = 1, \ldots, n \) denote an index set for the set of hospitals, and let \( y_{rj} \) denote the number of discharges in DRG_r from hospital j. Then \( \sum_{r=1}^{s} y_{rj} \bar{C}_r \) measures the expected cost for the j'th hospital and by construction

\[ \sum_{r=1}^{s} \sum_{j=1}^{n} y_{rj} \bar{C}_r = \bar{C} \times \sum_{r=1}^{s} \sum_{j=1}^{n} y_{rj} \]  

(5)

which in turn implies
\[
\frac{\sum_{i=1}^{s} \sum_{j=1}^{s} W_{r}}{\sum_{i=1}^{s} \sum_{j=1}^{s} W_{i}} = 1
\]  
\[\text{(6)}\]

i.e. average cost measured in terms of DRG-weights must equal one by construction.

Estimations of \( W_{r}, r = 1, \ldots, 471 \), have been carried out in a number of countries including Norway, see Slåtetebrak (1990). The Norwegian cost estimates have been used as a starting point in the current study since Danish cost estimates are not available. It is not without problems of its own to use cost estimates from Norway on Danish data, because the Norwegian cost estimates reflect clinical practice in Norway which may differ from clinical practice in Denmark. These differences imply among other things that the average length of stay in each DRG differs between the two countries. The Norwegian analysis includes a decomposition of total average cost for each DRG into 10 components, some of which are affected by length of stay and others not. The decomposition makes it possible to adjust the Norwegian estimates to Danish clinical practice.

Let \( FC_{r}^{s} \) denote average fixed cost per discharge in DRG \( r \), i.e. cost unaffected by length of stay and let \( VC_{r}^{s} \) denote average variable cost per day for patients in DRG \( r \), i.e. cost affected by length of stay. For each DRG we assume that fixed cost per discharge and variable cost per bed day are non random, i.e. these entities can be determined with very high precision, and no noise is present.\(^8\) \( LS_{r}^{s} \) denotes a random variable representing number of bed days of discharges in DRG \( r \). The superscript \( s \) refers to country, i.e. \( s = \{ N, D \} \) where \( N \) indicates Norway and \( D \) Denmark. The random cost \( C_{r}^{s} \) of a treatment in each DRG \( r \) is now determined as

\[
C_{r}^{s} = FC_{r}^{s} + VC_{r}^{s} \times LS_{r}^{s}, \ r = 1, \ldots, 471, \ s = N, D
\]
\[\text{(7)}\]

where we assume that \( FC_{r}^{s} \) and \( VC_{r}^{s} \) are non random variables. The mean cost \( \overline{C}_{r}^{s} \) of a treatment in the \( r \)th DRG is

\[
\overline{C}_{r}^{s} = FC_{r}^{s} + VC_{r}^{s} \times \overline{LS}_{r}^{s}, \ r = 1, \ldots, 471, \ s = N, D
\]
\[\text{(8)}\]

where \( \overline{LS}_{r}^{s} \) is the mean value of \( LS_{r}^{s} \). The random normalized cost in each DRG is

\[
W_{r}^{s} = \frac{\overline{C}_{r}^{s}}{C} = FC_{r}^{s} + VC_{r}^{s} \times \overline{LS}_{r}^{s}, \ r = 1, \ldots, 471, \ s = N, D
\]
\[\text{(9)}\]

and the Norwegian mean cost values or DRG-weights \( \overline{W}_{r}^{N} \) can now be written

\[
\overline{W}_{r}^{N} = \frac{\overline{C}_{r}^{N}}{C} = FC_{r}^{N} + VC_{r}^{N} \times \overline{LS}_{r}^{N}, \ r = 1, \ldots, 471
\]
\[\text{(10)}\]

\(^8\) These assumptions are necessary because we do not have access to the empirical distributions of these cost components for each DRG. However, it is relatively easy to modify the model, if such data on the distributions are available and the analyst wants to incorporate this variation in data into the model.
Assume that the ratios between average fixed and variable cost do not differ between Norway and Denmark, i.e.

\[
\frac{FC_r^N}{C_r^D} = \frac{FC_r^D}{C_r^D} \quad \text{and} \quad \frac{VC_r^N}{C_r^D} = \frac{VC_r^D}{C_r^D}, \quad r = 1, \ldots, 471
\]  

(11)

It is now possible to adjust for differences in length of stay between Norway and Denmark by defining Danish DRG-weights as follows:

\[
W_r^D = \frac{C_r^D}{C_r^D} = \frac{FC_r^D}{C_r^D} + \frac{VC_r^D}{C_r^D} \times LS_{r}^{D} = \\
(1 - \tau_r)W_r^N + \tau_r\bar{W}_r^N \times \frac{LS_r^D}{LS_r^N}, \quad r = 1, \ldots, 471
\]  

(12)

where \((1 - \tau_r)W_r^N \equiv \frac{FC_r^N}{C_r^N}\) and \(\tau_r\bar{W}_r^N \equiv \left[\frac{VC_r^N}{C_r^N} \times \frac{LS_r^N}{LS_r^N}\right]\), i.e. \(\tau_r\) is the relative share of the length of stay affected component of \(W_r^N\), \(r = 1, \ldots, 471\)

The empirical length of stay distribution for each DRG is in general highly skewed. It is possible to work directly with these empirical skew distributions. However, we have assumed that the length of stay distributions are truncated normal with endogenous lower truncation points for all DRGs in order to allow for an estimation of confidence intervals which is not dependent on a single or a few observations. A normal distribution truncated endogenously from below is a flexible distribution, see Cohen (1950) and Cohen and Woodward (1953). Using the trim points implies that abnormally long lengths of stay (exceeding the trim points) within a given DRG are considered outliers and are removed\(^9\)

The support or range of these distributions is the interval \([1, \infty)\). Hence, the probability assigned to a length of stay less than one day is zero for all inpatients.

Maximum likelihood estimates for the mean, the variance and the endogenous lower trim point of the normal distribution are identified, see Cohen (1950) and Cohen and Woodward (1953) for details, and the resulting truncated distribution used for estimation of \((100 - 2\alpha)\)-confidence intervals for length of stay for various probability levels \(\alpha\). The lower and upper end points for the confidence intervals, \([LS_r^-, LS_r^+]) \subset [1, \infty)\), can now be determined by

\[
Prob(LS_r^D \leq LS_r^-) = \frac{Prob(LS_r^D \geq LS_r^+)}{\alpha} = \frac{\alpha}{(1 - \alpha)}
\]  

(13)

with the probabilities evaluated in the estimated truncated normal distribution for DRG\(_r\). Using (12) we have the corresponding confidence interval for \(W_r^D\) from

\[
Prob\left(W_r^D \leq \frac{FC_r^D}{C_r^D} + \frac{VC_r^D}{C_r^D} \times LS_r^-\right) = Prob\left(W_r^D \geq \frac{FC_r^D}{C_r^D} + \frac{VC_r^D}{C_r^D} \times LS_r^+\right) = \alpha
\]  

(14)

---

\(^9\) Removing observations above the trim point also implies that some of the skewness is removed. Less skewness will move the mean value closer to the median value and thereby make the analysis less sensitive to whether or not the mean cost or the median cost are chosen as the DRG-weight.
or the interval

$$\left[ \left( \frac{V_{CD}}{C_r} \times L_{S_r^-} \right), \left( \frac{V_{CD}}{C_r} \times L_{S_r^+} \right) \right] + \frac{F_{CD}}{C_r}, r = 1, \ldots, 471.$$  

The DRG-system is designed for a grouping of inpatient somatic discharges and does not include long-term medical treatments, psychiatric treatments and ambulatory visits. Long-term medical and psychiatric treatments are carried out at a set of specialized hospitals but also in specialized departments at larger hospitals. Hospitals which specialize in long-term medical and psychiatric treatments are considered atypical and for this reason excluded from the sample. However, long-term medical and psychiatric treatments cannot be isolated from the somatic inpatient treatments in a number of hospitals, and ambulatory activity is an integral activity on most hospitals. Average cost estimates and their variation comparable with the DRG-weights are required for these activities. The set of activities not covered by the DRG-system relates to:

- i) hospitalized AIDS-patients,
- ii) one day hospitalizations in medical wards,
- iii) acute and non-acute ambulatory visits,
- iv) psychiatric hospitalizations, and
- v) outlier bed days.

Somatic inpatient activities related to AIDS-patients are registered in the LPR-database. Psychiatric inpatient activities are registered in a special psychiatric database, to which we do not have access. However, we know the number of bed days and discharges of psychiatric patients at each hospital. Average cost estimates $\overline{C}_{AIDS}$ and $\overline{C}_{PSYC}$ per bed day for activities i) and iv) are available. Thus, the ratios $\frac{C_{AIDS}}{C}$ and $\frac{C_{PSYC}}{C}$ define weights comparable with the DRG-weights. A similar construction is available for activities ii) and v). The Danish data allows for a distinction between acute and non-acute ambulatory visits. Moreover, the Danish health authorities have suggested that acute ambulatory visits can be divided into three groups and non-acute ambulatory visits into five groups and suggested a set of relative prices. Average cost for the least expensive ambulatory visit has been set to D.Kr. 600, which in turn allows for a calculation of DRG-comparable weights for ambulatory visits as for activities i), ii), iv), and v).\(^{10}\) Let $\overline{W}_r^D$, $r = 472, \ldots, 483$, denote the resulting weight estimates.

Information on the variation in the random DRG-comparable weights $W_r^D$, $r = 472, \ldots, 483$ is not available. It is assumed that these weights are distributed normally with mean values $\overline{W}_r^D$ and variances $\left( \frac{W_r^D}{4} \right)^2$, $r = 472, \ldots, 483$.

In summary, input is measured by observed cost (including physician cost) for each hospital in the final sample made up by 70 Danish hospitals. Outputs for a hospital are defined by the number of discharges within 471 DRGs and 12 additional groups related to

\(^{10}\) The price of D.Kr. 600 for an ambulatory visit is calculated as 1/4 of the average cost of a bed day.
psychiatric bed days and ambulatory visits along with the treatment of AIDS-patients in a
given calendar year.

An estimation of a stochastic frontier cost function is one way to proceed. Clearly, this
approach requires a drastical aggregation of the 483 outputs within a sample of size 70.
The aggregation might be based on the available DRG-weights in the sense that
discharges with DRG-weights of similar magnitudes might be pooled into aggregated
groups. But it is hard to argue that the resulting say about 10 output dimensions can be
characterized as homogeneous. Equally important, the procedure does not take the noise
in the DRG-weights into account.

We are of the opinion that a DEA-based efficiency analysis of the (Danish) hospital
sector which allows for

i) a specification of the model in the total output space of dimension 483,
ii) the incorporation of relative bounds on the parameters to be estimated, i.e. the
    virtual output multipliers, and
iii) a control for the stochastic noise in the DRG-weights

is superior compared to a stochastic frontier approach.11 Model (3) developed in Section
2 allows for an efficiency evaluation which does not require additional aggregation
beyond the one implied by the DRG-system and which facilitates the incorporation of
uncertainty on the average cost estimates provided as an integral part of the DRG-system.

4. Presentation of numerical results

Model (3) has been solved for each of the 70 hospitals in the sample with probability
levels set equal to 20%, 10%, and 5%, i.e. so that 20%, 10%, and 5% of the probability
mass in the estimated truncated normal distribution for each DRG-weight is cut off two-
sided.12 Thus, the stability of the estimated DEA-indices with respect to level of
probability is demonstrated. The model has been solved both for an upper truncation of
the length of stay distribution (trimmed data) and no upper truncation (untrimmed data).
Model (3) has also been solved for each of the 70 hospitals in the sample with probability
levels set equal to 50%, i.e. all outputs are aggregated into one output using the DRG
weights. It is common practice to use the (normalized) mean cost and not the
(normalized) median cost as DRG-weights. However, probability equal to 50% implies
an aggregation based on the median cost. Hence, we have included both efficiency scores
based on an aggregation from the mean cost and from the median cost. The results are
reported in Table 2 and graphically illustrated in Figures 4 and 5.

11 Thus, an aggregation beyond the DRG-system is not required and the stochastic noise in the
    DRG-weights is incorporated explicitly.
12 Observe that data actually allows for 2-sided cutoffs on some probability level in the empirical
distributions for the DRG-weights as well as in the estimated truncated normal approximations.
Table 2. Estimated DEA-indices for 70 Danish hospital on 3 levels of probability and on trimmed and untrimmed data.

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Cutoff at Trim points</th>
<th>No cutoff at Trim points</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>640,24</td>
<td>701,24</td>
<td>745,25</td>
</tr>
<tr>
<td>2</td>
<td>789,05</td>
<td>841,12</td>
<td>870,26</td>
</tr>
<tr>
<td>3</td>
<td>807,94</td>
<td>867,52</td>
<td>907,32</td>
</tr>
<tr>
<td>4</td>
<td>669,22</td>
<td>718,30</td>
<td>750,54</td>
</tr>
<tr>
<td>5</td>
<td>888,06</td>
<td>945,60</td>
<td>981,24</td>
</tr>
<tr>
<td>6</td>
<td>830,31</td>
<td>893,36</td>
<td>938,44</td>
</tr>
<tr>
<td>7</td>
<td>832,32</td>
<td>887,72</td>
<td>926,32</td>
</tr>
<tr>
<td>8</td>
<td>1000,00</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>9</td>
<td>743,77</td>
<td>804,12</td>
<td>848,42</td>
</tr>
<tr>
<td>10</td>
<td>900,12</td>
<td>992,46</td>
<td>1000,00</td>
</tr>
<tr>
<td>11</td>
<td>1000,00</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>12</td>
<td>1000,00</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>13</td>
<td>991,85</td>
<td>987,52</td>
<td>1000,00</td>
</tr>
<tr>
<td>14</td>
<td>754,17</td>
<td>794,18</td>
<td>823,27</td>
</tr>
<tr>
<td>15</td>
<td>701,11</td>
<td>746,72</td>
<td>780,29</td>
</tr>
<tr>
<td>16</td>
<td>783,36</td>
<td>832,95</td>
<td>869,34</td>
</tr>
<tr>
<td>17</td>
<td>831,80</td>
<td>888,62</td>
<td>930,20</td>
</tr>
<tr>
<td>18</td>
<td>905,12</td>
<td>949,98</td>
<td>979,11</td>
</tr>
<tr>
<td>19</td>
<td>935,22</td>
<td>987,52</td>
<td>1000,00</td>
</tr>
<tr>
<td>20</td>
<td>1000,00</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>21</td>
<td>961,20</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>22</td>
<td>892,39</td>
<td>940,72</td>
<td>976,31</td>
</tr>
<tr>
<td>23</td>
<td>964,11</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>24</td>
<td>937,07</td>
<td>994,20</td>
<td>1000,00</td>
</tr>
<tr>
<td>25</td>
<td>788,40</td>
<td>837,19</td>
<td>874,96</td>
</tr>
<tr>
<td>26</td>
<td>831,65</td>
<td>882,34</td>
<td>920,86</td>
</tr>
<tr>
<td>27</td>
<td>765,92</td>
<td>836,36</td>
<td>881,16</td>
</tr>
<tr>
<td>28</td>
<td>964,47</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>29</td>
<td>898,47</td>
<td>941,35</td>
<td>968,94</td>
</tr>
<tr>
<td>30</td>
<td>816,30</td>
<td>912,77</td>
<td>990,06</td>
</tr>
<tr>
<td>31</td>
<td>856,79</td>
<td>923,99</td>
<td>972,64</td>
</tr>
<tr>
<td>32</td>
<td>881,64</td>
<td>926,87</td>
<td>957,11</td>
</tr>
<tr>
<td>33</td>
<td>997,39</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>34</td>
<td>931,53</td>
<td>978,52</td>
<td>1000,00</td>
</tr>
<tr>
<td>35</td>
<td>964,21</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
</tbody>
</table>
Table 2. (continued) Estimated DEA-indices for 70 Danish hospital on 3 levels of probability and on trimmed and untrimmed data.

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Cutoff at Trim points</th>
<th>No cutoff at Trim points</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40 pct. \ 20 pct. \ 10 pct.</td>
<td>40 pct. \ 20 pct. \ 10 pct.</td>
<td>100 pct.</td>
</tr>
<tr>
<td>36</td>
<td>972,39</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>37</td>
<td>1000,00</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>38</td>
<td>895,72</td>
<td>956,46</td>
<td>999,44</td>
</tr>
<tr>
<td>39</td>
<td>988,92</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>40</td>
<td>896,75</td>
<td>956,95</td>
<td>999,54</td>
</tr>
<tr>
<td>41</td>
<td>935,32</td>
<td>982,30</td>
<td>1000,00</td>
</tr>
<tr>
<td>42</td>
<td>934,23</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>43</td>
<td>950,62</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>44</td>
<td>956,32</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>45</td>
<td>875,06</td>
<td>966,62</td>
<td>1000,00</td>
</tr>
<tr>
<td>46</td>
<td>944,39</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>47</td>
<td>963,46</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>48</td>
<td>921,31</td>
<td>966,67</td>
<td>996,89</td>
</tr>
<tr>
<td>49</td>
<td>874,41</td>
<td>919,66</td>
<td>951,13</td>
</tr>
<tr>
<td>50</td>
<td>909,19</td>
<td>964,46</td>
<td>999,07</td>
</tr>
<tr>
<td>51</td>
<td>874,39</td>
<td>918,31</td>
<td>950,84</td>
</tr>
<tr>
<td>52</td>
<td>744,94</td>
<td>821,59</td>
<td>879,84</td>
</tr>
<tr>
<td>53</td>
<td>861,95</td>
<td>927,94</td>
<td>977,90</td>
</tr>
<tr>
<td>54</td>
<td>792,80</td>
<td>836,28</td>
<td>868,48</td>
</tr>
<tr>
<td>55</td>
<td>877,05</td>
<td>921,11</td>
<td>953,85</td>
</tr>
<tr>
<td>56</td>
<td>954,01</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>57</td>
<td>895,14</td>
<td>968,22</td>
<td>1000,00</td>
</tr>
<tr>
<td>58</td>
<td>838,46</td>
<td>924,66</td>
<td>991,15</td>
</tr>
<tr>
<td>59</td>
<td>823,17</td>
<td>868,53</td>
<td>900,22</td>
</tr>
<tr>
<td>60</td>
<td>888,64</td>
<td>972,08</td>
<td>1000,00</td>
</tr>
<tr>
<td>61</td>
<td>830,73</td>
<td>889,71</td>
<td>931,21</td>
</tr>
<tr>
<td>62</td>
<td>847,03</td>
<td>906,55</td>
<td>948,74</td>
</tr>
<tr>
<td>63</td>
<td>754,38</td>
<td>803,27</td>
<td>833,40</td>
</tr>
<tr>
<td>64</td>
<td>752,05</td>
<td>801,92</td>
<td>839,68</td>
</tr>
<tr>
<td>65</td>
<td>921,44</td>
<td>973,50</td>
<td>1000,00</td>
</tr>
<tr>
<td>66</td>
<td>863,80</td>
<td>914,62</td>
<td>952,85</td>
</tr>
<tr>
<td>67</td>
<td>943,59</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>68</td>
<td>973,21</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>69</td>
<td>960,67</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
<tr>
<td>70</td>
<td>961,43</td>
<td>1000,00</td>
<td>1000,00</td>
</tr>
</tbody>
</table>
It has been decided not to identify hospitals by name, since the reported results are considered illustrative for the suggested method; additional fine tuning and validation of data is needed for a more conclusive analysis. A number for the hospital at hand and its Danish classification code are reported in columns 1 and 11. Hospitals classified by '1' are the largest ones with national and/or regional specialties, those classified by '2' are large but with no such specialties, those classified by '3' ('4') have at least 3 (exactly 2) clinical wards, and hospitals classified by '5' define the remaining set. Efficiency indices for probability levels 20%, 10%, and 5% and trimmed data are reported in columns 2, 3, and 4 and the corresponding indices for untrimmed data are reported in columns 5, 6, and 7. Efficiency indices for probability levels 50% based on the mean cost and the median cost are reported in columns 8, 9 and 10. Indices for lower probability levels do not fall below those for higher levels because confidence regions become larger as the level of probability decreases.\textsuperscript{13} It is well known that the reported DEA-indices can be given an interpretation in terms of the minimal proportional reduction in cost required in order to be characterized as efficient.\textsuperscript{14}

32 hospitals are characterized as efficient on the 5% probability level, 23 are characterized as efficient on the 10% level, and 5 are termed efficient on the 20% level for trimmed data. 42 hospitals are efficient on the 5% level, 27 on the 10% level, and 8 are termed efficient on the 20% level for untrimmed data. These observations illustrate the fact that efficiency indices increase as probability levels decrease because the production possibility set becomes smaller.

Indices that are relatively stable with respect to changes in the level of probability are the more likely in a statistical sense. Conversely, the exact values of indices that increase relatively much as the level of probability is reduced are the less likely. Compare hospitals \#40, \#41, \#45, and \#57 with hospitals \#1, \#2, \#3, \#4, \#5, \#6, and \#7 in Figures 4 and 5. Indices for the first group are more likely in a statistical sense than those for the second group. Compare next hospitals \#31 and \#32. The ranking of these hospitals according to efficiency index is affected by level of probability. \#31 is the less efficient on the 20% level, the indices are very close on the 10% level, and \#32 is the less efficient on the 5% level.

The results indicate that a number of Danish hospitals are termed inefficient no matter the level of probability and the trimming or not trimming of data. DEA identifies for an inefficient hospital a reference set, namely all efficient hospitals with a strictly positive dual multiplier in (3.2). A combination of these efficient hospitals is capable of producing at least equally many discharges in each dimension at a lower cost compared to the inefficient hospital given the bounds imposed on the rates of substitution by (3.4) and (3.5) (see appendix). Hence, the DEA-approach also provides information on which hospitals' routines to be copied in order to be characterized as efficient.

\textsuperscript{13} More stringent, the models are nested in the sense that the production possibility set for a lower probability level is a subset of the production possibility set for a higher probability level. The nesting property is caused by the fact that a larger variation in prices is allowed with lower as compared to higher probability levels.

\textsuperscript{14} Reported indices have been scaled by a factor 1000.
Figure 4. Efficiency scores for the 70 Danish hospitals. Trimmed Data.

Figure 5. Efficiency scores for the 70 Danish hospitals. Untrimmed Data.

A DEA cost function based on 483 outputs and so-called probabilistic assurance regions has been used for an estimation of the cost efficiency of 70 Danish hospitals. It has been argued that the imposed probabilistic assurance regions allow for a frontier estimation in the full 483 dimensional output space, i.e. the approach does not involve a fixed aggregation. Furthermore, the incorporation of the cost distribution for each patient cluster allows for a control of variation in heterogeneity of output clusters. Finally, the measured efficiency for each hospital can be given different likelihood based upon the sensitivity of the score w.r.t. the probability levels used in the specification of confidence regions for the probabilistic assurance regions.

Efficiency measurement within the hospital sector is challenging. The state of the art of efficiency analysis in combination with the complexity of the performance of hospitals makes it no easy task, and neither Data Envelopment Analysis nor Stochastic Frontier Analysis is without inherent problems. We agree with Newhouse (1994) that one should be careful when using results of an efficiency analysis as part of the foundation for decision making. However, the results can be useful for asking questions to the manager of a hospital, allowing each hospital the opportunity to come up with explanations as to why it has been assigned a low or a high efficiency index. An explanation of a low index may reveal omitted outputs. On the other hand, a low score may reflect true inefficiency if no explanation is available, and the DMUs termed efficient may serve as benchmarks for currently best practice.

Although we find it advisable to use results from such analyses with caution (that is partly the reason for having omitted the names of the hospitals in the application) we are of the opinion that frontier estimation is a potentially useful tool also within the health care sector. The development of new theory is the most constructive approach for addressing shortcomings in available models. The estimation of an efficient cost frontier for the Danish hospital sector has called for extensions of the existing methodologies combined with incorporation of more information related to the activities of the hospitals. An extension of the DEA-framework based upon the incorporation of the distributional characteristics of the (empirical) cost distributions for each patient cluster has been suggested. This construction has improved the measurement of hospital outputs, it avoids any fixed aggregation of patient clusters, and it allows for an estimation of the frontier in a full 483-dimensional cost output space.

We have addressed the question of aggregation of multiple outputs in an estimation of a cost frontier, and argued that it is possible to introduce probabilistic assurance regions in DEA based on observable information of the cost distribution for each output category. Hence, it is possible to incorporate variation in heterogeneity across the output categories and allow the efficiency measurement to be based on a flexible aggregation defined by the distributions of cost, provided cost data is available. This provision is often fulfilled in the hospital sector and most certainly so in Denmark.

We have applied the model to measure hospital efficiency in Denmark. Danish data on individual discharges allows for an estimation of the cost distribution in a number of
clusters. We have used the DRG-system to define such clusters. It has been demonstrated that it is possible to avoid aggregating discharges in the 483 different DRGs, and that it is possible to incorporate the differences in homogeneity in the different DRGs in the analysis. The DEA framework allows each hospital to select output prices in a cone spanned by confidence intervals derived from the cost distributions. This implies that the relative prices corresponding to relatively heterogeneous DRGs are allowed to vary more than those corresponding to relatively homogeneous DRGs.

We have argued that estimated efficiency scores which are insensitive to variation in probability level have a high likelihood. Stability of an efficiency score reflects that it is not affected by the varying degree of homogeneity of the DRGs. On the other hand, efficiency scores which are highly sensitive to the level of probability have a relatively low likelihood. However, the difficulties encountered when measuring the efficiency of hospitals remain. Scores may reflect deficiencies in the DRG system in the form of aggregation bias in the sense that discharges with very different levels of resource consumption are lumped together in a DRG.

References.


Apprendix.

A1. Development of the DEA model for hospital efficiency evaluation

The classical DEA model suggested by Charnes et al. (1978 & 79) and termed the CCR-model is formulated as follows:

\[
\min \quad \theta_{j_0} - \left( \sum_{i=1}^{m} s_{i}^+ - \sum_{r=1}^{s} s_{r}^- \right) \epsilon \tag{A1.1}
\]

\[
\text{s.t.} \quad \sum_{j=1}^{n} \lambda_{j} x_{ij} + s_{i}^+ = \theta_{j_0} x_{ij_0}, \quad i = 1, ..., m \tag{A1.2}
\]

\[
\sum_{j=1}^{n} \lambda_{j} y_{rj} - s_{r}^- = y_{rj_0}, \quad r = 1, ..., s \tag{A1.3}
\]

\[
\lambda_{j} \geq 0, \quad j = 1, ..., n \tag{A1.4}
\]

\[
s_{i}^+, s_{r}^- \geq 0, \quad i = 1, ..., m, \quad r = 1, ..., s \tag{A1.5}
\]

\(x_{ij}\) measures the use of input \(i, i = 1, ..., m\), and \(y_{rj}\) the production of output \(r, r = 1, ..., s\), at DMU\(_j\), \(j = 1, ..., n\). \(\lambda_{j}\), \(j = 1, ..., n\), is a non-negative scaling factor and \(\theta_{j_0}\) is the CCR-efficiency index to be estimated for DMU\(_{j_0}\) currently subjected to evaluation. Finally, \(\epsilon\) is a non-Archimedian. The interpretation of the program is well known. A linear combination of DMUs capable of producing at least as much as the DMU under evaluation is required by (A1.3). Moreover, the particular linear combination satisfying (A1.3), which uses the smallest possible amount of inputs in the same mix as the DMU under evaluation is identified by (A1.1 & 2). This combination is usually termed the reference unit for DMU\(_{j_0}\). In effect, the optimal value of \(\theta_{j_0}\) measures the maximal possible proportional reduction of observed inputs for the unit under evaluation allowed by observed data.

An efficiency index is obtained for all DMUs in the sample by the solution of the above linear programming model for \(j_0 = 1, 2, ..., n\). DMU\(_{j_0}\) is termed efficient if and only if the optimal \(\theta_{j_0}\) is equal to 1, see Charnes et al. (1978 & 79) for details.

The dual counterpart for model (A1) is as follows:

\[
\max \quad \sum_{r=1}^{s} u_{r} y_{rj_0} \tag{A2.1}
\]

\[
\text{s.t.} \quad \sum_{i=1}^{m} v_{i} x_{ij_0} = 1 \tag{A2.2}
\]

\[
\sum_{r=1}^{s} u_{r} y_{ij} - \sum_{i=1}^{m} v_{i} x_{ij} \leq 0, \quad j = 1, ..., n \tag{A2.3}
\]

\[
u_{r} \geq \epsilon, \quad r = 1, ..., s \tag{A2.4}
\]

\[
u_{i} \geq \epsilon, \quad i = 1, ..., m
\]
This problem can be given an interpretation in terms of a pricing problem. Strictly positive output prices \( u_r, r = 1, \ldots, s \), and input prices \( v_i, i = 1, \ldots, m \), must be determined so that the resulting value of outputs does not exceed the value of inputs for any DMU by (A2.3 & 4). Moreover, a pricing scheme which maximizes the value of output for the DMU under evaluation subject to the constraint that the value of inputs for that DMU must equal 'l' is required by (A2.1 & 2). It is well known, that the ratios between each pair of multipliers can be given an economic interpretation in terms of a marginal rate of substitution, see Banker et al. (1984).

The optimal criterion in (A2) equals by construction the corresponding optimal criterion in (A1), since the two problems are each others duals and both are bounded and feasible. Hence, the set of CCR-indices can be estimated by the solution of either one of the two models.

The indices obtained provide measures of technical efficiency only. Hence, DMU \( j_o \) is termed efficient by (A1) (and consequently also by (A2)) if no reference unit capable of producing the same or a larger amount of outputs using a smaller amount of inputs compared to DMU \( j_o \), can be constructed as a non-negative linear combination of observations in the sample. However, a corresponding pricing scheme as obtained by (A2) may be judged unreasonable for a number of reasons. For instance, suppose that market prices on inputs are available, and that one input is assigned a much larger price than another on the market and a much smaller price than the other according to the DEA model. Or suppose that the firms under evaluation are public, that market prices for outputs are not available, and that the production of a unit of a given output is known to require a much larger amount of inputs than the production of one unit of another output. DEA prices which do not reflect this relationship between the two outputs may be judged unreasonable. Finally, suppose that technological information indicates that the marginal rate of substitution between a given pair of outputs does not fall below e.g. .5 and does not exceed 2.0. The inverse ratio between the corresponding pair of acceptable DEA prices should in this case belong to the interval [.5, 2.0].

The DEA framework allows for an easy incorporation of a priori information concerning the relative prices of inputs and outputs. This information can either be incorporated into (A1) or equivalently into (A2). (A2) is the more appropriate formulation when additional price information is to be incorporated, since the model is formulated in the price space. A number of different formulations are available depending on the price information at hand. The so-called assurance region or AR-formulation suggested by Thompson et al. (1986 & 90) imposes upper and lower bounds on the relative prices and can be written as follows:

\[
\alpha^{u}_{r_1 r_2} \leq \frac{u_{r_1}}{u_{r_2}} \leq \beta^{u}_{r_1 r_2}, \quad r_1, r_2 = 1, \ldots, s, \quad r_1 < r_2 \\
\alpha^{v}_{i_1 i_2} \leq \frac{v_{i_1}}{v_{i_2}} \leq \beta^{v}_{i_1 i_2}, \quad i_1, i_2 = 1, \ldots, m, \quad i_1 < i_2
\]
Hence, the ratio between each pair of output multipliers and between each pair of input multipliers are required to belong to some intervals. \( (A2.4') \) can be given the interpretation of a tightening of \( (A2.4) \) since \( (A2.4) \) can be obtained by setting all \( \alpha \)-values equal to \( \epsilon \) and all \( \beta \)-values equal to \( \infty \). Observe that a linking of output and input prices is not imposed.

The upper and lower bounds in the AR-specification, i.e. the \( \alpha \)- and \( \beta \)-values, are parameters to be determined a priori e.g. by expert opinion or by historical data on prices/rates of substitutions. These bounds are in a number of applications subject to uncertainty. For instance, the average cost estimates in the DRG-system are uncertain, and some estimates are more uncertain than others. Clearly, this uncertainty should be incorporated into the efficiency analysis. Olesen & Petersen (1998) suggest the following approach\(^\text{15} \) for that purpose.

Assume that we have a stochastic "norm-vector" \( \mathbf{e} = [e_1, \ldots, e_s]^T \), where each component \( e_i \), \( i = 1, \ldots, s \) is a random variable and these \( s \) random variables are distributed according to some multivariate distribution. Consider some set of realizations of the random vector \( \mathbf{e} \) that is regarded as having a high likelihood (it could for example be a confidence region). We want to restrict all relative output prices to belong to the set of ratios of all realized components in this random vector \( \mathbf{e} \) constrained to this set of realizations of high likelihood.

To simplify the presentation of the probabilistic ARs we employ the following assumptions\(^\text{16} \): Assume that \( \mathbf{e} \) is distributed according to a symmetric multivariate distribution, where all marginal distributions are identical and only parametrized by a location and a scale parameter. Let \( \mathbf{e} = [\mathbf{e}_1, \ldots, \mathbf{e}_s]^T \) be the vector of median or mean values\(^\text{17} \) of the vector \( \mathbf{e} \), and let \( \kappa_\alpha^- (\alpha^-_r) \) and \( \kappa_\alpha^+ (\alpha^+_r) \) be lower and upper endpoints of a confidence interval for \( e_r \) at probability levels \( (\alpha^+_r - \alpha^-_r) \), \( r = 1, \ldots, s \). Hence, \( \text{Prob}(e_r \leq \kappa_\alpha^- (\alpha^-_r)) = \alpha^-_r \) and \( \text{Prob}(e_r \geq \kappa_\alpha^+ (\alpha^+_r)) = \alpha^+_r \), \( r = 1, \ldots, s \). We will now restrict all relative output prices by only allowing \( u_i \) and \( u_j \) to belong to the cone \( C \) generated by the confidence intervals for \( e_i \) and \( e_j \), \( i, j = 1, \ldots, s \), \( i < j \):

\[
C \equiv \left\{ u \in \mathbb{R}_+^s : \frac{\kappa_\alpha^- (\alpha^-_i)}{\kappa_\alpha^+ (\alpha^+_j)} \leq \frac{u_i}{u_j} \leq \frac{\kappa_\alpha^+ (\alpha^+_i)}{\kappa_\alpha^- (\alpha^-_j)}, i, j = 1, \ldots, s, i < j \right\} \quad (A3)
\]

\( ^{15} \) The approach can be implemented in the input as well as the output space. Focus is for obvious reasons on the output space in the context of the current hospital efficiency evaluation.

\( ^{16} \) To present the general idea behind this approach we focus in this illustration on a symmetric distribution. However, length of stay distributions, even after trimming the distribution by cutoff points or trim points, are highly skew. Hence, no assumption of symmetric length of stay distributions is employed in the hospital application.

\( ^{17} \) We have in general to distinguish between the median cost and the mean cost since we are dealing with cost distributions that are highly skew. For \( \alpha^n_p \rightarrow 50\% \), \( i = 1, \ldots, s \), \( p \in \{+, -\} \) the confidence intervals are getting smaller and smaller and for \( \alpha^n_p = 50\% \) the confidence intervals equal to the median cost and not the mean cost. Hence, one could argue in favor of using the median cost in each DRG as the DRG-weight and not the average cost, which are common practice when estimating DRG weights.
Figure A1 illustrates the case of $\kappa_i^- (\alpha_i^-) = \bar{e}_i - \sigma_i$, $\kappa_i^+ (\alpha_i^+) = \bar{e}_i + \sigma_i$, for $\bar{e}_i = 60$, $\sigma_i = 20$ for $i = \{1, 2, 3\}$. The gray squares in each of the three projections are seen to correspond to the Cartesian products of the confidence intervals [40, 80].

Figure A1. Assurance regions in multiplier space. The leftmost figure illustrates the three dimensional cone obtained in the example in the text and the rightmost figure its three projections.

Observe that the formulation above involves $s(s - 1)$ constraints. It is demonstrated in Olesen and Petersen (1998) that the set of constraints can be rewritten as

\[
ku_r = \mu_r \kappa_i^- (\alpha_i^-) + (1 - \mu_r) \kappa_i^+ (\alpha_i^+), \quad r = 1, \ldots, s
\]

\[
\mu_r \in [0, 1], \quad k \geq 0 \quad r = 1, \ldots, s
\]  \hspace{1cm} (A4)

This formulation defines the set of feasible output multipliers by a scaling of points in the confidence region defined by the Cartesian product of all confidence intervals and is much to prefer in practical applications since only $2s$ constraints are needed.

Let us return to the assurance region $C$ defined by (A3). Formulation (A3) is suggested as a simple and intuitively appealing approach for imposing the requirement that the ratios between each pair of multipliers must belong to a set of all corresponding ratios from realizations of the stochastic benchmark vector considered of high likelihood. It will now be shown that (A3) can be given a theoretical foundation within the theory of chance constrained programming. This result is established by demonstrating that $C$ is a so-called certainty equivalent for a set of probabilistic constraints defined in relation to the stochastic benchmark vector $(e_1, \ldots, e_s)$. Consider for each $(i,j) \in IJ$ the probabilistic statement:

*The probability of each of the two events $\frac{c_{ij}}{u_j} \leq c_{ij}$ and $c_{ij} \leq \frac{c_{ij}}{u_j}$ must belong to the intervals $[\alpha_i^-, \alpha_i^+]$ and $[\alpha_j^-, \alpha_j^+]$ centered around 0.5 for some $c_{ij} \in \mathbb{R}_+$.***
Thus, we want for each \((i,j) \in IJ\) to constrain the probability of the event \(\frac{u_i}{u_j} \leq \frac{u_j}{u_i}\) to belong to some narrow interval around 0.5. We want these constraints represented as simple as possible, which in this case means that we focus on the marginal and not the joint distribution of \(e_i\) and \(e_j\) and therefore on the probabilistic statement above. The statement can be written

\[
\alpha_i^+ \geq \text{Prob}\left( \frac{e_i}{u_i} \leq c_{ij} \right) \geq \alpha_i^- \quad \text{and} \quad \alpha_j^+ \geq \text{Prob}\left( \frac{e_j}{u_j} \geq c_{ij} \right) \geq \alpha_j^-
\]

(A5)

for some \(c_{ij} \in \mathbb{R}_+, \forall (i,j) \in IJ\),

The certainty equivalents for the set of probabilistic constraints in (A5) are

\[
\frac{c_{ij}u_i - \bar{e}_i}{\sigma_i} \in [\eta_i^-, \eta_i^+], \quad \frac{c_{ij}u_j - \bar{e}_j}{\sigma_j} \in [\eta_j^-, \eta_j^+]
\]

(A6)

for some \(c_{ij} \in \mathbb{R}_+, \forall (i,j) \in IJ\)

where \(\eta_k^- \equiv \Phi^{-1}(\alpha_k^-), \eta_k^+ \equiv \Phi^{-1}(\alpha_k^+), k = i, j\), and \(\Phi()\) is the distribution function for the distribution with zero mean and variances equal to one. It is easily seen that (A7) is implied by (A6):

\[
\frac{\bar{e}_i + \eta_i^+ \sigma_i}{\bar{e}_j - \eta_j^+ \sigma_j} \geq \frac{u_i}{u_j} \geq \frac{\bar{e}_i + \eta_i^- \sigma_i}{\bar{e}_j - \eta_j^- \sigma_j}, \forall (i,j) \in IJ,
\]

(A7)

and it is shown in Olesen and Petersen (1998) that any \(u\) satisfying (A7) also satisfies (A6) with a particular choice of the \(c_{ij}\). Hence (A6) and (A7) are equivalent representations of the same cone.

The following two cases may occur in the definition of the set of cone constraints by (A7):

\[i) \quad 0.5 - \alpha_i^- = \alpha_i^+ - 0.5 \text{ for all } i\]

\[ii) \quad 0.5 - \alpha_i^- \neq \alpha_i^+ - 0.5 \text{ for some } i\]

\(i)\) is the standard case with two sided confidence intervals determined with equal probability mass on each side of the mean values. It is demonstrated below that (A7) in this case defines a cone spanned by the Cartesian products of the confidence intervals. \(ii)\) is the case of confidence intervals where probability masses of unequal size are cut off in

\[18\] Using this simplification implies that the constraints are based on an approximation of the true joint distribution. The simplification is made in order to establish an equivalence between (A3) and the set of chance constraints in (A5) and allows for the above simple presentation of the general idea behind a probabilistic constraint. However, the simplification is also useful if the joint distribution turns out to be highly complicated, as in the hospital application. It is in this application for practical purposes not possible to estimate a 483 dimensional joint distribution which fits data reasonably well.
the ends of a 2-sided confidence interval. Olesen and Petersen (1998) show that the cone
deﬁned by (A7) in this case is no longer spanned by the Cartesian product of the partial
conﬁdence intervals and it is affected by the order in which the s prices are represented in
the set IJ.

We consider only case i) in relation to the application, because this is the case related to
the standard deﬁnition of conﬁdence intervals. The assumption 0.5 − αi− = αi+ − 0.5 for
all i maintained in case i) implies along with the assumption that the distributions of the
components of e are symmetric that (A7) can be rewritten identical to formulation (A3):

\[
\frac{\kappa_i^+ (\alpha_i^+)}{\kappa_j^-(\alpha_j^-)} = \frac{\varepsilon_i + \eta_i^- \sigma_i}{\varepsilon_j + \eta_j^+ \sigma_j} \geq \frac{u_i}{u_j} \geq \frac{\varepsilon_i + \eta_i^- \sigma_i}{\varepsilon_j + \eta_j^+ \sigma_j} = \frac{\kappa_i^-(\alpha_i^-)}{\kappa_j^+(\alpha_j^+)}, \forall (i, j) \in IJ,
\]  

(A8)

(A3) is suggested as a simple and intuitively appealing approach for imposing the
requirement that the ratios between each pair of multipliers must belong to a set of all
corresponding ratios from realizations of the stochastic benchmark vector considered of
high likelihood. Hence, letting η_i^- = Φ^-1(α_i), η_i^+ = Φ^-1(α_i), for α_i ∈ (0,0.5),
i = 1, . . . , s, C in (A3) is a certainty equivalent for the cone C':

\[
C' \equiv \left\{ u \mid 1 - \alpha_i \geq \text{Prob} \left( \frac{e_i}{u_i} \leq c_{ij} \right) \geq \alpha_i, 1 - \alpha_j \geq \text{Prob} \left( \frac{e_j}{u_j} \geq c_{ij} \right) \geq \alpha_j \right\}, (A9)
\]

for some c_{ij} ∈ \mathbb{R}_+, (i, j) ∈ IJ

The equivalence between cones C and C' is established by assigning each pair of
multipliers (u_i, u_j), (i, j) ∈ IJ, a scaling factor c_{ij} of its own. Otherwise, we would
have (A7) implied by (A6) but not vice versa. The scaling factors c_{ij} are not necessarily
equivalent for different pairs of multipliers. However, it is shown in Olesen and Petersen
(1998) that it is possible to ﬁnd at least one common c ∈ \mathbb{R}_+ such that c_{ij} = c,
(i, j) ∈ IJ, if 0.5 − α_i^- = α_i^+ − 0.5 for all i, as maintained here.

Symmetric versus non-symmetric distributions.

The approach outlined above is presented in Olesen and Petersen (1998) with a focus on
symmetric distributions. However, a generalization to the case of non-symmetric
distributions is straightforward, when we restrict ourselves to the case of two-sided
conﬁdence intervals with equal probability mass cut off at both sides.

Assume now that the components in the stochastic "norm-vector" e = [e_1, . . . , e_s]^T are
distributed according to some non-symmetric multivariate distribution. Let again \kappa_r^-(\alpha_r^-)
and \kappa_r^+(\alpha_r^+) be lower and upper endpoints of a conﬁdence interval for e_r at probability
levels α_r^- and α_r^+, \ r = 1, . . . , s. We will now restrict all relative output prices by only
allowing u_i and u_j to belong to the cone generated by the conﬁdence intervals for e_i and
e_j, i, j = 1, . . . , s, i < j. As before we want for each (i, j) ∈ IJ to constrain the
probability of the event \frac{u_i}{e_j} ≤ \frac{u_j}{e_i} to belong to some narrow interval around 0.5 and with a
focus on the marginal and not the joint distribution of e_i and e_j. We consider as in (A5)
the probabilistic statement:
\[ \alpha_i^+ \geq \text{Prob}\left( \frac{e_i}{u_i} \leq c_{ij} \right) \geq \alpha_i^- \quad \text{and} \quad \alpha_j^+ \geq \text{Prob}\left( \frac{e_j}{u_j} \geq c_{ij} \right) \geq \alpha_j^- \quad (A10) \]

for some \( c_{ij} \in \mathbb{R}_+, \forall (i, j) \in IJ, \)

The certainty equivalents for the set of probabilistic constraints in (A10) can now be expressed as

\[ c_{ij}u_i \in \left[ \Phi_i^{-1}(\alpha_i^-), \Phi_i^{-1}(\alpha_i^+) \right], \quad c_{ij}u_j \in \left[ \Phi_j^{-1}(1 - \alpha_j^+), \Phi_j^{-1}(1 - \alpha_j^-) \right] \quad (A11) \]

for some \( c_{ij} \in \mathbb{R}_+, \forall (i, j) \in IJ \)

where the lower and upper endpoints \( \kappa_k^- (\alpha_k^-) \) and \( \kappa_k^+ (\alpha_k^+) \) for the confidence intervals mentioned above are \( \kappa_k^- = \Phi_k^{-1}(\alpha_k^-), \kappa_k^+ = \Phi_k^{-1}(\alpha_k^+), \) \( k = i, j, \) and \( \Phi_r() \) is the distribution function for the untransformed non-symmetric distribution of \( e_r, \) \( r = 1, \ldots, s. \) Assuming \( 0.5 - \alpha_i^- = \alpha_i^+ - 0.5 \) for all \( i, \) it is easily seen that (A11) can be rewritten as

\[ c_{ij}u_i \in \left[ \Phi_i^{-1}(\alpha_i^-), \Phi_i^{-1}(\alpha_i^+) \right], \quad c_{ij}u_j \in \left[ \Phi_j^{-1}(\alpha_j^-), \Phi_j^{-1}(\alpha_j^+) \right] \quad (A12) \]

for some \( c_{ij} \in \mathbb{R}_+, \forall (i, j) \in IJ \)

and (A13) is implied by (A12):

\[ \frac{\kappa_i^+(\alpha_i^+)}{\kappa_j^-(\alpha_j^-)} = \frac{\Phi_i^{-1}(\alpha_i^+)}{\Phi_j^{-1}(\alpha_j^-)} \geq \frac{u_i}{u_j} \geq \frac{\Phi_i^{-1}(\alpha_i^-)}{\Phi_j^{-1}(\alpha_j^-)} = \frac{\kappa_i^-(\alpha_i^-)}{\kappa_j^+(\alpha_j^+)} \quad \forall (i, j) \in IJ, \quad (A13) \]
A2. Use of the DRG-system in Denmark

The use of the DRG-system for an efficiency evaluation of Danish hospitals has been subject to discussion. Emphasis has been on the questions whether

i) all DRGs are sufficiently homogeneous and reflect clinical practice in Denmark,

ii) specialty should be used as an indicator for cost weights, and

iii) transfer of inpatients between local and specialized hospitals needs to be controlled for.

We are of the opinion, that the problems as indicated by the questions above are not of sufficient importance to reject the use of the DRG-system on Danish data. The viewpoint is supported by a number of analyses published in Danish (and for that reason not included in the list of references).

The DRG-based case mix index for the \( j \)'th hospital denoted \( \text{CMI}_j, j = 1, \ldots, 70 \), is easily calculated as

\[
\text{CMI}_j = \frac{\sum_{r=1}^{483} \nu_{rj} \cdot W_{pr}}{\sum_{r=1}^{483} \nu_{rj}}, j = 1, \ldots, 70
\]

A case mix index above (below) one indicates that the hospital has treated a mix of patients with a relatively large (small) expected consumption of resources per discharge, compared to the average consumption of resources per treatment. Clearly, observed cost may differ from expected cost for a number of reasons, of which stochastic noise and differences in efficiency are the most important. The number of discharges within each DRG produced at each hospital within a given calendar year can be measured fairly precisely. The main sources for uncertainty with respect to this issue are

i) classification errors due to the clinical data reported in the LPR-database and the assignment of each discharge to a DRG,

ii) errors caused by patients submitted to a hospital in one calendar year and discharged in another, and

iii) errors caused by transfers of patients between hospitals.

Observed outputs are for these (and other) reasons in principle stochastic variables. Equally and in our opinion even more important, stochastic noise also relates to the fact that patients in a given DRG are not homogeneous with respect to the required use of resources. However, the relationship between the DRG-based casemix index and observed average cost indicates whether \( \text{CMI}_j \) times the average cost per discharge for all DRGs and all hospitals can be seen as a reasonable estimate for \( \overline{C}_j \), i.e. the average cost per discharge for all DRGs at hospital \( j, j = 1, \ldots, 70 \). If not, the possible reasons are:

i) Danish hospitals are characterized by large differences in efficiency, or

ii) stochastic noise in data, or
iii) inconsistencies between clinical practice in Denmark and the DRG-system, or
iv) a combination of i)-iii)

The legitimacy for using the DRG-system on Danish data is supported by the following plot of DRG-based case mix indices vs. average cost per discharge for the set of hospitals in the sample.\textsuperscript{19}

Insert Figure A2

Figure A2. The relationship between average cost per discharge and case mix indices for 70 Danish hospitals.

The figure indicates a positive relation between $\text{CMI}_j$ and $\text{AC}_j$, $j = 1, \ldots, 70$, as expected and is considered a fairly strong indication of the validity of using the DRG-system on Danish data.

\textsuperscript{19} As observed above, the DRG-system does not include activities related to AIDS, psychiatric, and ambulatory visits all of which are integral activities at least at some Danish hospitals. We have for that reason adjusted observed cost at each hospital w.r.t. these activities. To be more specific, a bed day for an AIDS-patient is assigned a cost of D.Kr. 5,470-, a bed day for a psychiatric patient a cost of D.Kr. 1,880-, and an ambulant treatment (converted to weight 1) a cost of either D.Kr. 400-, 600-, or 800- with D.Kr. 600- used in the above plot.