Working Paper 07-11
Business Economics Series 05
March 2007

Departamento de Economía de la Empresa
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 6249607

A REASONABLE BENCHMARKING FRONTIER USING DEA: AN INCENTIVE SCHEME TO IMPROVE EFFICIENCY IN PUBLIC HOSPITALS*

Diego Prior [1] and Jordi Surroca [2]

Abstract

There exists research relating management concepts with productivity measurement methods that offers useful solutions for improving management control in the public sector. Within this sphere, we connect agency theory with efficiency analysis and describe how to define an incentives scheme that can be applied in the public sector to monitor the efficiency and productivity of managers. To fulfill the main objective of this research, we propose an iterative process for determining what we define as a 'reasonable frontier', a concept that provides the foundation required to establish the incentive scheme for the managers. Our 'reasonable frontier' has the following properties: i) it detects the presence of outliers, ii) it proposes a procedure to establish the influence introduced by extreme observations, and iii) it sorts out the problem of data masking. The proposed method is applied to a sample of hospitals taken from the public network of the Spanish health service. The results obtained confirm the applicability of the proposal made. Summing up, we define and apply a useful method, combining aspects of agency theory and efficiency analysis, which is of interest to those public authorities trying to design effective incentive schemes which influence the decision making of the public managers.

***Keywords**:* Incentives, Agency Theory, DEA, Outliers Detection, Data Masking
***JEL Classification**:* C14, C61, D23, D24

* Los autores quieren agradecer el respaldo financiero del Ministerio de Ciencia y Tecnología (royecto # SEC2003-04770), del Ministerio de Educación y Ciencia (proyecto # SEJ2006-01731); y de la Comunidad de Madrid y Universidad Carlos III de Madrid. (Proyecto # UC3M-ECO-05-042). Las opiniones de los autores son de su entera responsabilidad.

[1] Diego Prior Jiménez. Catedrático del Departamento de Economía de la Empresa, Universitat Autònoma de Barcelona 08193 Cerdañola (Barcelona), España. Tel.: +34-93 581 15 39 Fax: +34-93-581 25 55. diego.prior@uab.es
[2] Jordi Surroca Aguilar. Profesor Visitante del Departamento de Economía de la Empresa, Universidad Carlos III de Madrid 28903 Getafe (Madrid), España. Tel.: +34-91 624 86 40 Fax: +34-91 624 96 07. jsurroca@emp.uc3m.es

Introduction

In many national health systems, healthcare costs have increased dramatically during the last few decades. Increases in treatment prevalence, cost per treated case and population are the largest factors driving expenditures up. As a consequence, a survey conducted by The Economist in 2002 (July 15th 2004) shows, for example, that Germany spends 11% of its *GDP* on healthcare and France 10%. The *US* case is somewhat different because, although the expenditure is up to around 15%, more than half of this comes directly from its citizens' pockets rather than taxes.

As is well known, in European countries healthcare is financed by compulsory contributions from employers and employees. This social-insurance model has proved vulnerable to two main forces: rising healthcare costs and a shortfall in contributions resulting from unemployment. As a consequence of actual macroeconomic and demographical conditions (i.e. slow growth, drop in specific industrial and manufacturing activities, unemployment, ageing population, low birth rates, generous pension provision, lowering of taxes) these expenditures are putting pressure on already overloaded national budgets. In response to this problem, many European countries, especially Germany and France, have embarked on an ambitious package of reforms to raise efficiency and to get more value for money in medical care.

Hospital managers are viewed as crucial to the success of the above-mentioned reforms. The full participation of the managers in the reforms, however, is far from guaranteed. In this case, as Agency Theory points out, as in many other occasions, the decisions taken by managers are in conflict with the interests of the organizations they manage. In an agency relationship, the principal transfers decision-making responsibilities to the agent by means of a contract. Very often this relationship is characterized by bounded rationality and imperfect information. As such, the contract does not usually specify how the agent ought to proceed when faced with circumstances that are difficult to foresee. But such circumstances are very common in normal business life, so contracts tend to be incomplete. This situation gives the agents freedom to use their judgment in the decision making process and to take advantage of the asymmetric information situation (Holmström, 1982).

Tirole (2001) argues that the conflict can be resolved, or at least lessened, with the adoption of governance mechanisms such as supervision or incentive schemes. The adoption of those mechanisms is particularly complex in the case of stakeholder organisations, such as the public sector, because it is impossible to maximize more than one objective at the same time (i.e. a broad mission). As a consequence, the multiplicity of objectives will leave managers without a way to make a reasoned decision (in effect, it leaves the manager without an objective), it makes the process of decision-taking more difficult, it politicizes the corporation, and it leaves the managers with the power to exercise their own preferences in spending the firm's resources. In addition, the absence of competitive pressure on managers hinders the achievement of the maximum possible level of efficiency; Leibenstein (1966). Consequently, managers of public sector organisations often lack the incentives to: 1) limit possible waste of resources; 2) make optimal use of resources; or, 3) produce outputs in such a way that current needs are met; Tullock (1965), Niskanen (1971 and 1975), and Breton and Wintrobe (1975).

Examples of the above-mentioned difficulties in national healthcare systems are widespread. It must also be noted that, when public authorities have taken steps to introduce managerial incentive schemes, these have not always produced the desired motivational effect. As an example, Byrnes and Valdmanis (1994) discuss the *US* experience of paying hospitals according to the cost incurred, a practice that does not provide any incentive for cost reduction. Consequently, during the eighties, this system was replaced by another where hospitals' payments are determined prospectively (instead of retrospectively, as in the 80's) based on predetermined prices, which are determined by the average unit costs of the required products. The Dutch health authorities have adopted a similar solution: hospitals are financed on the basis of general indicators, such as "costs per case" and "bed-days per case" (Kuntz and Scholtes, 2000; Verheyen, 1998). Thus, if total expenditures exceed the budget a hospital has been assigned, expenditure cuts in subsequent years offset overdrafts in the previous financial year.

As a result of this and others reforms in the EU, the increase in health expenditures in European countries has been lower than in the US. The European experience with cost-control measures thus demonstrates that it is technically possible to control healthcare costs by government regulation of supply rather than demand; Verheyen (1998). By linking managerial incentives and hospital funding to efficiency, healthcare authorities aim to control expenditure and to stimulate competition between internal units and between hospitals. It's hardly surprising, thus, that a considerable body of literature in economics and management science is devoted to the problem of measuring efficiency in healthcare services. An extensive review can be found in Worthington (2004).

However, despite the importance of efficiency measurement in the healthcare sector, an important question still remains unanswered. Indeed, the very definition of Pareto-efficiency presumes knowledge of the technological production possibility set. Thus, a decision-making unit (*DMU*) can attain full efficiency if, and only if, none of its inputs or outputs can be improved without worsening some of its other inputs or outputs. As complete information regarding the transformation function that relates inputs to outputs is generally unknown, the preceding definition is therefore simplified by emphasizing its use only with the information that is empirically available. This is how Data Envelopment Analysis (*DEA*) operates. Consequently, emphasis is put on relative efficiency: a *DMU* is to be rated as fully efficient if and only if the performance of other *DMUs* in the sample does not show that some of its inputs or outputs can be improved without worsening some of its other inputs and outputs. In this sense, the existing observations are used for the construction of a non-parametric empirical frontier where it is not necessary to know the exact form of the function that relates inputs to outputs. Thus, an efficiency score can be obtained for each *DMU*, in comparison to the rest of the *DMUs*, which measures the distance that separates a specific *DMU* from the frontier.

The unanswered question of how the efficiency measurement of an empirically-based production technology is affected by the composition of the sampled efficient *DMUs* arises immediately. In order to introduce incentive schemes that can motivate hospitals to maximize the achievement of targets set by the health authority, it is desirable that the frontier contains hospitals whose performance can be matched. On the contrary, if the performance of the hospitals on the frontier is unattainable because they use a very different technology, operate in very different circumstances or it is simply that the data about them contains errors, efficiency targets will be distorted for the remaining hospitals, and this may create disincentives rather than incentives.

The identification of hospitals whose performance cannot be matched is the subject of the present article. Departing from this idea, we study how to create an incentive-based remuneration scheme when we cannot verify the efforts made by agents (i.e. hospital's managers), and when we do not know the productive technology employed by the organisations to which they belong. Our investigation has antecedents in a limited number of theoretically-oriented papers that have already examined how empirical production frontiers may contribute to the incentives of *DMUs*, namely Bogetoft (1994, 1995) and Agrell et al. (2002). In this paper we extend those works and we propose the identification of a "reasonable frontier" that should set the objectives to be met by agents. This frontier has two essential characteristics: first, it is based on the concept of technical efficiency; second, the frontier sets optimal targets, but we would like these targets to be feasible and indisputable. The identification of a reasonable frontier brings to the fore the need to identify outlying and influential observation, and the measurement of their impact on the efficiency of the remaining decision units in the sample.

The structure of the paper is as follows. The next Section introduces the budget system of the Spanish public hospitals and discusses the need for incentives. The economic problem to be studied is defined in Section 1, and its relationship with *DEA* is established. Section 2 is also concerned with some of the essential properties of our incentive system. This should motivate managers both in inefficient and in efficient organisations. It should also avoid setting objectives that cannot be achieved. Section 3 takes these premises as a starting point and describes our proposed incentive scheme and how this is related to the construction of a reasonable frontier. Section 4 describes the application of the methodology using a sample of Spanish public hospitals, and includes a discussion of the data, the variables used, the methodology, and the results obtained. The final section of the paper illustrates the managerial relevance of the methodology.

1 The need for incentives in Spanish public hospitals

In Spain, as in most Western economies, there is a concerted effort to make the public health system more efficient. There has been an attempt to overcome the traditional lack of financial discipline and the chronic deficit of Spanish public hospitals through the *INSALUD;* the relevant central administration body, which regulates, finances and administrates the provision of health services in those Spanish regions (so-called "Comunidades

Autónomas") without delegated responsibility in healthcare. To address the lack of interest by hospital managers in increasing efficiency, between 1992 and 1997, a "contract programme" was introduced to regulate the relationship between Spanish hospital and the *INSALUD*. Under this contract, year by year, hospitals individually negotiated production objectives and budget allocations to be received in the forthcoming period. This forward-looking system was handicapped because hospitals tended to hide the level of activity they could potentially achieve. The existence of a "ratchet effect" was confirmed by the observation that the hospitals that were better at meeting their production objectives were also the ones that had the most increase in the targets in later periods; Gonzalez and Ventura (2004).

Nowadays, the strategy followed by central government is based on the decentralization of responsibility and on the delimitation of roles and risks between public authorities and hospitals. The consensus is that these changes will stimulate competition between hospitals and increase internal control. The separation of the financing and decision-taking functions is viewed as a unique way to break with the bureaucracy of the past, which jeopardized efficiency improvement. The package of measures adopted includes: the substitution of the *INSALUD* by the "Sistema Nacional de Salud", *SNS* (National Health System); the actualization of the old contract-program; a new juridical status for hospitals, which are incorporated to public enterprises, public foundations or consortiums; and innovative formulas for the external contracting of healthcare services, through privately-owned hospitals. The last-mentioned measure serves to distinguish between direct management, when health services are produced by public entities, and indirect management, which is characterized by the production technologies of private enterprise.

The change in the characteristics of program-contracts is to be noted; gradually, the old system of retrospective payments has been replaced by a system in which payments are prospective. Nowadays, a hospital budget is calculated by applying to the production goals, measured in physical units, prices set to the estimated average costs for a sample of hospitals. However, the successful implementation and subsequent development of the above-mentioned measures has been limited. This is because incentive-programs do not provide enough incentives to induce improvement in efficiency and financial discipline in public hospitals (Gonzalez and Barber, 1996). In effect, the current system is handicapped by an old shortcoming; incentive pay to improve quality and/or decrease expenditures is not included in the package of measures; and implicit incentives such as professional promotion are not possible. Similarly, no major reform has been proposed to regulate the exclusive dedication of professionals. Signed commitments, to be in force until 2005, only contain a total wage rise of between 5 and 12 per cent, and a moderate staff increase; while substantial reforms have been postponed to the next agenda.

In the next section, we study how to create an incentive-based remuneration scheme for hospital managers, based on the empirical knowledge of the Spanish healthcare system and the recognized need for incentives. To discuss possible solutions, we ought to be concrete about the context: agents (hospital managers) make decentralised decisions on the amount of effort they apply, where information is non-symmetric and non-verifiable by third parties, and where the production functions of individual hospitals are unknown.

2 The relationship between Agency Theory and Data Envelopment Analysis.

2.1 The agency contract and the problem of incentives

The context of this paper is the agency relationship established between a principal –the health authority–, and several agents –hospital managers–. By means of the contract that regulates this relationship, the principal delegates to each agent *i*, $i = 1,\ldots,I$, the task to transform *n* production resources, $x^i = \left(x_1^i,\ldots,x_n^i\right) \in \Re_+^n$, into *m* hospital products, $y^i = \left(y_1^i,\ldots,y_m^i\right) \in \Re_+^m$. We assume that the amount finally produced depends on the productive resources used and on the effort $\left(e^1,\ldots,e^I\right)$ made by each one of the *I* agents, $y^i = F\left[x^i, e^i\right]$.

$$\left[\left(x_1^i,\ldots,x_n^i\right)_{i=1,\ldots,I}, \left(e^1,\ldots,e^I\right)\right] \rightarrow \left(y_1^i,\ldots,y_n^i\right)_{i=1,\ldots,I} \qquad [1]$$

For the contract between the principal and the agent to be complete, the relationship must not suffer from information problems; i.e., no situations of incomplete information and/or asymmetric information. Hence, the agents are required by contract to apply a level of effort $e^{i**} = \left(e^{1**},\ldots,e^{I**}\right)$ such that the wealth of the principal is maximised. Wealth is defined in this case as the difference between total product value and the sum of the opportunity costs incurred by the *I* agents, $W = F\left[x^i, e^i\right] - \sum_{i=1}^{I} v^i\left(e^i\right)$. In this way the principal will obtain a maximum wealth (*first best*), $W^{**} = F\left[x^i, e^{i**}\right] - \sum_{i=1}^{I} v^i\left(e^{i**}\right)$, and the agents will receive a compensation for their effort which equals their opportunity cost, $s^i = v^i\left(e^{i**}\right)$.

When information cannot be verified, the agency relationship cannot be regulated by complete contracts. The need to base the relationship between the principal and the agent on incomplete contracts confers a degree of discretion on the agent; and agents can take decisions that harm the interests of the principal. In this new information context, the principal could add a clause to the contract forcing the agent, *i*, to make an effort $e^i = e^{i**}$, but then the question arises of whether this contract is self-enforcing (i.e., whether $e^{i**} = \left(e^{1**},...,e^{I**}\right)$ is the Nash solution).

Since effort is not verifiable, it cannot be stated in a contract. Effort creates a private cost to the agent. Each agent will contribute a final effort such that his/her utility function, $U^i\left(s^i, e^i\right) = u^i\left(s^i\right) - v^i\left(e^i\right)$, is maximised. This function is defined by the difference between the utility that he/she obtains from his/her salary s^i (as defined by the principal), and the residual utility obtained from $v^i\left(e^i\right)$. Assuming that both functions satisfy appropriate conditions (u^i is concave and monotonously increasing, and v^i is also monotonously increasing), each agent solves his/her own maximisation problem. The effort level of the *I* hospital managers will be $e^{i*} = \left(e^{1*},...,e^{I*}\right)$, a Nash equilibrium; and, therefore, final product levels are $y^{i*} = F\left[x^i, e^{i*}\right]$. Holmström (1982) proves that the solution to the problem of maximising the agent's utility does not coincide with the first best optimum because the agents face a problem of moral hazard when deciding their level of effort.[1] This moral hazard situation can be quantified by combining the agent's individual solution, e^{i*}, and the social optimum solution, e^{i**}, as follows:

---

[1] More specifically, Holmström (1982) points out that there is an exclusive way to ensure the agent's effort is optimal. This situation is attained when there is a unique agent collecting the total amount of generated wealth. In this specific case, the roles of principal and agent are played by the same person.

$$MH^i = e^{i**} / e^{i*} \qquad [2]$$

The value of MH^i is greater than one, since $e^{i**} > e^{i*}$.

The degree to which the optimal levels –from the point of view of the principal– of the *m* products are achieved using the given levels of the *n* productive resources within an agency theory framework, will depend on the conflict of interests that emerges between the agent and the principal, and on the extent to which existing incentive and control mechanisms can stop agents from making decisions that favour their own interest. Holmström's (1982) solution is that the principal should set up a system of rewards and punishments based on the amount of product achieved, leaving agents the freedom to set their own effort levels (as these are not verifiable). Under such an incentive system, agents would receive their reserve utility, $v^i\left(e^{i**}\right)$, if the final output was the *first best* optimal solution, y^{i**}, and would receive nothing otherwise. It can be shown that in this case agents choose as optimal a level of effort, $e^{i**} = \left(e^{1**},..., e^{I**}\right)$, and that the wealth of the principal is maximised, $F\left[x^i, e^{i**}\right] - \sum_i^I v^i\left(e^{i**}\right)$.

Unfortunately, any attempt to introduce such an incentive scheme encounters an important problem: how to collect the necessary information to identify the function that maps productive resources into product levels. It is necessary to identify this function in order to build a production possibility frontier, and use it in order to set targets for each decision unit, in our case each hospital; see Bogetoft (1994).

2.2 Performance valuation using DEA

Given that it is not possible to ascertain the level of effort of the agents, and since we ignore the transformation function that maps production factors into product levels, we will instead study the decisions taken by each hospital manager in the context of the decisions taken by managers of the remaining hospitals. Such a study gives us an approximate way of assessing moral hazard. The actual effort of agent *i*, e^{i*}, can be approximated by means of the relationship between the level of *output actually observed* and the consumption of production factors, y^i / x^i; where $y^i = \left(y_1^i,..., y_m^i\right) \in \Re_+^m$ denotes the vector of observed outputs and

$x^i = \left(x_1^i,\dots,x_n^i\right) \in \Re_+^n$, the vector of inputs. The optimal effort, e^{i**}, could be approximated by means of the relationship between the potential level of output that could be achieved if the agent were to take maximum advantage of all the transformation possibilities offered by the existing technology, $y^{i**} = F\left[x^i, e^{i**}\right]$, and the resources used. This relationship can be quantified as $F\left[x^i, e^{i**}\right] / x^i$. It follows that, given an observed consumption of productive resources x^i, the problem of moral hazard that was defined in *[2]* can be approximated through the relationship between *potential output* and *observed output*:

$$MH^i = e^{i**} / e^{i*} \sim y^{i**} / y^{i*} \qquad [3]$$

But, how do we find out what the production function $F\left[x^i, e^i\right]$ is? A reasonable way forward is to use *DEA*.

DEA is particularly relevant in this case, as it does no attempt to estimate the form of the production function. *DEA* uses existing observations in order to create a non-parametric empirical frontier. To identify this frontier it is not necessary to know the exact form of the function that relates resources to production factors. *DEA* assumes the existence of a production possibility set that includes what is observed and what could have been observed. In our own particular case it would include all hospitals in the data set and those that could possibly exist, given existing technologies, but have not been observed. This set, *T*, has to satisfy a series of conditions, such as convexity and monotonicity; see Banker and Thrall (1992).

$$T = \left\{ (x, y) : x \geq \sum_{k=1}^{K} l^k x^k, y \leq \sum_{k=1}^{K} l^k y^k,\ l^k \geq 0, k = 1,\dots,K \right\} \qquad [4]$$

The set *T* is limited by an efficient frontier. This frontier can be used to measure the potential output that a hospital can deliver. To do so, it is necessary to solve, for each hospital, a Linear Programme (*LP*). In total *I* *LP's* need to be solved, one for each hospital; see Charnes et al. (1978):

$$\begin{aligned} &\underset{q^\circ, l^i}{\text{Max}}\ q^\circ \\ &\text{s.t.}\quad \sum_{i=1}^{I} l^i y_j^i \geq q^\circ y_j^\circ \quad ; \quad j = 1,\dots,m \\ &\qquad\ \sum_{i=1}^{I} l^i x_k^i \leq x_k^\circ \quad ; \quad k = 1,\dots,n \end{aligned} \qquad [5]$$

where the superscript "$\circ$" refers to the hospital that is being assessed. The solution to this problem, $q^{\circ**} \geq 1$, measures the proportion by which the *m* products of the hospital being assessed need to increase for the hospital to be located on the production possibility frontier; i.e., on the frontier that measures good practice or less greedy managers in the management of the principal's interests. Under this formulation, a manager is presumed to apply optimal effort, $e^{\circ**}$, and to run a hospital efficiently, only if $q^{\circ**} = 1$. The *DEA* algorithm identifies a reference set against which the performance of the hospital under observation $(\circ)$ is assessed. This reference set is associated with non-zero values of $l^{\circ**}$. In the particular case when $(\circ)$ forms part of the frontier we will find $l_{\circ}^{\circ} = 1$ y $l_{i}^{\circ} = 0$.

The potential output that could be achieved by a hospital is, under this DEA formulation, obtained by proportionally expanding the levels of all the outputs:

$$y^{\circ**} = q^{\circ**} \cdot y^{\circ*} \qquad [6]$$

where $y^{\circ**} = \left(y_j^{\circ**}\right)_{j=1,\dots,m}$. This makes it possible to quantify the moral risk problem, through the relationship between *potential output*, as suggested by the *DEA* algorithm, and *observed output* (the result of the actual decisions taken by the agent, $e^{\circ*}$):

$$MH^{\circ} = \frac{y^{\circ**}}{y^{\circ*}} = \frac{q^{\circ**} \cdot y^{\circ*}}{y^{\circ*}} = q^{\circ**}, \qquad [7]$$

Considering that this is greater than one, the principal could introduce an incentive scheme as follows:

$$s^{\circ}\left[q^{\circ}\right] = \begin{cases} v^{\circ}\left(e^{\circ**}\right) & \Leftrightarrow \quad q^{\circ**} = 1 \\ 0 & \Leftrightarrow \quad q^{\circ**} > 1 \end{cases}, \qquad [8]$$

Under this scheme, the agent would be paid a salary that would compensate him/her for the utility of his/her effort only if observed output equals potential output. He/she would not receive any compensation otherwise[2]. This incentive scheme motivates managers in inefficient hospitals to increase their effort until they reach *potential output*. It also creates new problems:

---

[2] Once could also think of an incentive scheme where the remuneration were decreasing on q.

Problem 1 (The measurement of performance in efficient hospitals) *The remuneration system defined by [8] does not discriminate between efficient hospitals: The managers of efficient hospitals will lack incentives to increase their effort once the potential output has been reached.*

Consider the decisions taken by the manager of an efficient hospital ($q^{\circ **}=1$). The monetary compensation that this agent receives for his/her effort is $v^{\circ}\left(e^{\circ **}\right)$. If the agent increases his/her effort, he/she will continue to receive the same reward, but he/she will still incur the whole of the private cost caused by the increase in effort. Consequently, the manager will lack incentives to go beyond the level of effort that places him/her on the production possibility frontier.

In order to overcome this problem, we follow Bogetoft's (1994) suggestion of estimating the potential output of a particular hospital using the concept of super-efficiency; Andersen and Petersen (1993). Under Andersen and Petersen's (1993) method the frontier is estimated without including in the reference set the hospital to be assessed. This approach makes it possible to order efficient hospitals. The ordering of inefficient hospitals is based on the results of *LP [5]*. The mathematical formulation of *LP [5]* that incorporates Bogetoft's (1994) suggestion is given by:

$$\begin{aligned} &\underset{\hat{q}^{\circ},\hat{l}^{i}}{\text{Max}} \quad \hat{q}^{\circ} \\ &\text{s.t.} \quad \sum_{\substack{i=1 \\ i\neq\circ}}^{I} \hat{l}^{i} y_j^i \geq \hat{q}^{\circ} y_j^{\circ} \quad ; \quad j=1,\ldots,m \\ &\qquad \sum_{\substack{i=1 \\ i\neq\circ}}^{I} \hat{l}^{i} x_k^i \leq x_k^{\circ} \qquad ; \quad k=1,\ldots,n \end{aligned} \qquad [9]$$

This modification to the algorithm has a consequence: $\hat{q}^{\circ **}$ is no longer constrained to values greater than or equal to 1. It can –and, in general, it will– take values smaller than 1. $\hat{q}^{\circ **}<1$ can be interpreted as the proportion by which hospital $(\circ)$ can reduce all its outputs and still be considered efficient. In the case of inefficient hospitals, *[5]* and *[6]* return the same value for $\hat{q}^{\circ **}$.

After this correction, the *potential output* that hospital $(\circ)$ must achieve in order to enter the production frontier, is:

$$\hat{y}^{\circ**} = \hat{q}^{\circ**} \cdot y^{\circ*}. \qquad [10]$$

and the problem of moral risk can be quantified as

$$\widehat{MH}^{\circ} = \frac{\hat{y}^{\circ**}}{y^{\circ*}} = \frac{\hat{q}^{\circ**} \cdot y^{\circ*}}{y^{\circ*}} = \hat{q}^{\circ**}, \qquad [11]$$

This is not constrained to being less than 1. It can take any positive value.

The system of incentives to be used by the principal in order to motivate agents would thus be modified to:

$$\hat{s}^{\circ}\left[\hat{q}^{\circ}\right] = \begin{cases} v^{\circ}\left(\hat{e}^{\circ**}\right) & \Leftrightarrow \quad \hat{q}^{\circ**} \leq 1 \\ 0 & \Leftrightarrow \quad \hat{q}^{\circ**} > 1 \end{cases}, \qquad [12]$$

Under this scheme, the manager of hospital $(\circ)$ will receive a reward that will compensate him/her for any increase in effort that raises the observed output beyond an efficient frontier.

The system of incentives thus defined rewards agents according to the distance that separates the amounts they produce from the amounts that would be produced under an efficient frontier that excludes their hospital (this has been rewritten because they are rewarded according to results, not according to decisions). This solves the problem of motivating managers in efficient hospitals, but a different problem may appear if the frontier is defined by hospitals whose performance is difficult to attain because they use a very different technology, operate in very different circumstances, or simply have error-strewn data. We will now expand and discuss this topic.

Problem 2 (The presence of hospitals whose performance cannot be matched) *If we include hospitals whose performance cannot be matched in the definition of the frontier, we run the risk of erroneously describing as sub-optimal a performance because we are judging it against a frontier that cannot be reached. This could have perverse consequences on the agent: if the agent considers that no amount of effort will bring the hospital to the frontier, the agent may keep effort to a minimum.*

In the context of frontier analysis, this problem is associated with the presence of extreme and influential observations. In practice, very small values of $\hat{q}^{i**}$ suggest that we are in the presence of an atypical observation that is far from the frontier. In other words, the relevant hospital could substantially reduce the amounts of products it generates, while keeping constant its consumption of resources, and still continue to be efficient- if judged against a frontier the construction of which excluded the hospital.

In order to identify hospitals whose performance cannot be matched, we need to start with a discussion of influential observations; see Timmer (1971), Wilson and Jadlow (1982), Dusansky and Wilson (1994 and 1995), Wilson (1995), Pastor *et al.* (1999) Fox et al., (2004). The presence of an influential observation has a disproportionate impact on the assessment of the efficiency of the rest of the observations (Dusasnky and Wilson, 1994), and its detection can be complicated by the existence of multiple outputs (Fox et al., 2004), as in the case of public hospitals. If an influential observation is removed from the data set, a large change will be seen in the efficiency of the remaining observations. This will not be the case if the observation removed is not influential.

In general, the *DEA* literature classifies an observation as a potential outlier only if it is included in the frontier, and ignores the possibility that observations not included in the frontier may also be influential.

Problem 3 (The masking of hospitals whose performance cannot be matched) *If the most influential hospitals are excluded from the data, other hospitals that did not form part of the frontier may join the efficient set, and some of these may now reveal a disproportionate impact on the efficiency of the rest of the sample.*

Data masking has already received attention in a parametric context- see, for example, Ezzamel and Mar-Molinero (1990)-, but has been ignored in *DEA*. In the following section we suggest an iterative way to identify hospitals whose performance cannot be matched. This procedure does not exclude the possibility of classifying as *influential* observations not located on the frontier. In the parametric context, the method introduced by Fox et al. (2004) does not distinguish between efficient and non-efficient observations, making possible outliers diagnosis for all the sample. Coming back to the non-parametric context, the exclusion of the most obviously influential observations may bring to the fore other observations not previously on the frontier. The literature does not offer any recipes in this case: the emphasis is on the identification of influential observations; judgment and detailed analysis of individual observations are then used to select the final sample.

This study discusses outlying hospitals within the framework of a negotiation between the health authority (principal) and hospital managers (agents). The removal of the most obviously influential observations will increase the degrees of freedom that the principal has when negotiating the system of incentives with the agents (or with its representatives, as would be the case with trade unions in a pay bargaining context).

3 Performance assessment by means of a reasonable frontier

3.1 The concept of reasonable frontier

Once they have accepted the need to set up an incentive scheme that limits the moral risk problem, the health authority and hospital managers can negotiate the levels of effort that are required under the scheme. This requires the identification of the reference set for each hospital in the context of the incentive scheme; i.e., the "reasonable frontier" needs to be defined. The health authority is interested in setting targets very close to their potential level, as defined by the transformation function. The agents will want the targets to be less demanding. The result of the negotiation process between principal and agent is an agreed reasonable frontier. Hospitals whose performance cannot be matched must not belong to this frontier, since their presence may distort the system of optimal targets of effort on which the incentive scheme is based.

Here we propose an iterative methodology for efficiency assessment, efficiency being defined in terms of the optimal effort required from the agent. The first step in this methodology is the calculation of Andersen and Petersen's (1993) super-efficiencies for all hospitals using *LP [9]*. Hospitals are ranked in increasing order of super-efficiency. The hospital with the lowest value of $\hat{q}^{i**}$ is thus identified. Next, we assess the impact that this hospital, *i*, has on the definition of the frontier. Since it has the lowest value of $\hat{q}^{i**}$, the exclusion of hospital *i* from efficiency calculations would produce the largest shift in the position of the frontier. Following this reasoning, we remove hospital *i* from the sample and re-calculate efficiencies. If the change in efficiency values is significant, we exclude hospital *i* from the target setting mechanism, as its presence may set targets that cannot be achieved. The decision to exclude hospital *i* from the target setting mechanism gives the health authority further degrees of freedom when negotiating with managers and facilitates the achievement of a compromise.

3.2 Identification of the reasonable frontier

Imagine that, having solved *LP [9]*, hospital $r,\ r \in (1,\dots,I)$ is associated with the lowest value of the efficiency index $\hat{q}^{i**}$.[3] We need to test whether this hospital should be labelled as influential. We do this as follows. We remove *r* from the sample and re-calculate the efficiencies for all remaining hospitals by solving the following *LP*:

$$\begin{aligned} &\underset{\hat{q}^{\circ},\hat{l}^{i}}{\mathrm{Max}}\ \ \hat{q}^{\circ}_{-r} \\ &\text{s.t.}\quad \sum_{\substack{i=1 \\ i\neq \circ,r}}^{I} \hat{l}^{i}_{i} y^{i}_{j} \geq \hat{q}^{\circ}_{-r} y^{\circ}_{j} \quad ; \quad j=1,\dots,m \\ &\qquad\ \ \sum_{\substack{i=1 \\ i\neq \circ,r}}^{I} \hat{l}^{i}_{i} x^{i}_{k} \leq x^{\circ}_{k} \qquad ; \quad k=1,\dots,n \end{aligned} \qquad [13]$$

Where sub-index "$-r$" indicates that observation *r* has been left out of the calculations. To assess if *r* is an influential observation, we compare efficiency indexes obtained under *LP [9]*, excluding the one associated with r, $\hat{q}^{**} = \left(\hat{q}^{1**},\dots,\hat{q}^{r-1**},\hat{q}^{r+1**},\dots,\hat{q}^{I\,**}\right)$, and efficiency indexes obtained under *LP [13]*, $\hat{q}^{**}_{-r} = \left(\hat{q}^{1**}_{-r},\dots,\hat{q}^{I**}_{-r}\right)$. We can say that r is an influential observation if the distribution of $\hat{q}^{**} = \left(\hat{q}^{1**},\dots,\hat{q}^{r-1**},\hat{q}^{r+1**},\dots,\hat{q}^{I\,**}\right)$ is statistically different from the distribution of $\hat{q}^{**}_{-r} = \left(\hat{q}^{1**}_{-r},\dots,\hat{q}^{I**}_{-r}\right)$. If a statistically significant difference is found, *r* is removed from the sample and the procedure is repeated with a new sample that excludes r, $i = (1,\dots,r-1,r+1,\dots,I)$. This iterative process ends when the null hypothesis of equality of distributions cannot be rejected.

Diagnosing whether an observation is influential or is not, obviously depends on the ability to detect statistically significant changes in the distribution of parameter $\hat{q}^{**}$. We are thus assuming that *DEA* efficiency is a random variable with a statistical distribution function, perhaps because the original data contains errors. The *F*

[3] Had program *[9]* been input-oriented, then we would have chosen the hospital with biggest efficiency coefficient (superefficient *DMUs* would have efficiency coefficients bigger than one and inefficient *DMUs* lower than one).

distribution is a possible statistical distribution for *DEA* efficiency; see Banker (1993).[4] Several statistics have been used in the *DEA* literature to test hypotheses on the distribution of efficiency. The Wilcoxon-Mann-Whitney Rank Sum test has been popular to compare the frontiers of two different types of *DMUs*; see, for example, Valdmanis (1992). Other tests, such as Banker's (1993) and Kolmogorov-Smirnov; Giokas (2001), have also been used. Kittelsen (1993) argues that, if two measures of efficiency are based on the same *DMUs*, the correct test to use is the Wilcoxon Signed-Rank Test.[5]

In this particular case we use the non-parametric Wilcoxon (matched-pairs) Signed Rank Test for related samples. There are two reasons to choose this test. First, we cannot make assumptions about the distribution of efficiency as calculated by *LP [9]*. Second, we cannot use a test for unrelated samples, such as the one due to Kruskal-Wallis, since the same decision unit belongs to both populations under consideration.

Wilcoxon's test checks if two probability distributions share the same median by calculating the following statistic:

$$z = \frac{T_{+} - \dfrac{n\cdot(n+1)}{4}}{\sqrt{\dfrac{n\cdot(n+1)\cdot(2n+1)}{24}}} \qquad [14]$$

where n is the number of observations in the sample. This number will change in every successive iteration. In the first iteration $n = I - 1$, since we are dealing with the complete sample after removing observation r, whose influence we are trying to establish. T_{+} is the lower of two values: the sum of positive ranks or the sum of negative ranks. The calculation procedure to be followed is as follows. We start by calculating the difference between the values of $\hat{q}^{**} = \left(\hat{q}^{1**},...,\hat{q}^{r-1**},\hat{q}^{r+1**},...,\hat{q}^{I**}\right)$ and $\hat{q}_{-r}^{**} = \left(\hat{q}_{-r}^{1**},...,\hat{q}_{-r}^{I**}\right)$. Next, we order the absolute values of the differences. We assign ranks, from 1 to n, 1 is associated with the smallest difference and n with the highest difference[6]. We add the ranks of all the cases associated with positive differences to obtain the

---

[4] Without this assumption, the use of statistical methods is not justified.

[5] But, if models are nested, this *test* is not significant because every change in the efficiency coefficients is the same for the whole sample. Pearson or Spearman correlation coefficients can be applied with nested models.

[6] If there are equal differences, the same average rank value is assigned to the units.

sum of positive ranks, and we proceed in the same way with the ranks of negative differences in order to obtain the sum of negative ranks.

Hospital r is considered to be an influential observation if the null hypothesis is rejected. If r is identified as influential, it should not be included in the reference set when designing the incentive scheme. The final result after the successive removal of all influential observations is what we call the *reasonable frontier*. It should be easier to achieve a compromise between the health authority and hospital managers on the basis of this frontier because it sets reasonably moderate targets. The incentive scheme will reward managers as a function of the distance that separates the current position of the agent and the relevant point on the reasonable frontier. Under Holmström's (1982) scheme, agents would obtain the following reward:

$$\hat{s}^{\circ}\left[\hat{q}^{\circ}_{-r}\right] = \begin{cases} v^{\circ}\left(\hat{e}^{\circ**}_{-r}\right) & \Leftrightarrow \quad \hat{q}^{\circ**}_{-r} \leq 1 \\ 0 & \Leftrightarrow \quad \hat{q}^{\circ**}_{-r} > 1 \end{cases}, \qquad [15]$$

Under this scheme, the reward received by agents will depend on the distance that separates the current position- that reflects past decisions- and the reasonable frontier. This incentive scheme has valuable properties: 1) it solves the problem of how to motivate efficient hospitals, 2) it is not influenced by the presence of efficient hospitals whose performance cannot be reached, and 3) it solves the problem of influential observations that are masked by unambiguous outliers. We also know that if there are changes in the observations that define the frontier, or if one of the observations is removed, no significant changes will take place in the rest of the sample, since neighbouring observations are relatively close to each other.

4 Data, variable definition and results

The data used in this case study were obtained from the *EESRI* survey of the Spanish Ministry of Health (Estadística de Establecimientos Sanitarios con Régimen de Internado, Ministerio de Sanidad y Consumo, 1995). The survey contains information on all public hospitals in Spain. This study is limited to general hospitals (specialized hospitals and nursing homes are excluded) with more than 200 beds, thus restricting the number of observations to 137.

The concept of efficiency used in this case study is in line with the theoretical discussion in the sections about agency theory. As in Chillingerian and Sherman (1990), our case study deals with hospital management efficiency, but the analysis could be extended to the efficiency of medical practitioners in the provision of health services.[7] The variables selected in order to assess efficiency were as follows:

y_1: *Total number of discharges,* number of discharges from internal medicine, medical specialties, general and digestive surgery, traumatology, obstetrics, and gynaecology.

y_2: *Total number of bed-days,* number of bed-days in internal medicine, medical specialties, general and digestive surgery, traumatology, obstetrics, and gynaecology.

y_3: *Total number of out-patient appointments,* medical assistance to out-patients for diagnostic purposes, for treatment, and for the follow-up of clinical cases.

The productive resources considered were:

x_1: *Medical staff and other professional workers*; medical staff (full-time equivalent at 40 hours per week); chemist, medical staff on hospital duty, and other staff with higher degrees.

x_2: *Other staff;* Nurses, physiotherapists, auxiliary staff, managers, administration staff.

x_3: *Beds*: Number of hospital beds as a proxy for investment in plant, and as a measure of size of hospital, building volume, and plant complexity.

x_4: *External purchases,* purchases of medicines, sanitary material, food and drinks, instruments, clothes, and other consumables..

[TABLE 1 ABOUT HERE]

Table 1 shows the descriptive statistics for the input and output variables. Table 2 summarises the results of applying our proposed methodology to the sample of hospitals. The first five columns of the table contain information on the iterative process required for the identification of the reasonable frontier. There is an alternative procedure for outlier identification in *DEA* due to Wilson (1995). Wilson's procedure was also

---

[7] The EESRI questionnaire does not contain the required information for the evaluation of hospital activity in terms of health production. As pointed out by Murray (1992), if, in variable selection, we are interested in determining the final outcome of health services, there are enough reasons to prefer outcome variables (i.e., the number of patients treated weighted by their corresponding *Diagnostic Related Group*), rather than throughput variables (i.e.. the number of patient days or the number of cases treated).

applied: the next five columns of Table 2 contain relevant information on this alternative method. The final three columns of Table 2 compare the results of Wilson's method and our proposed methodology, and indicate the decision that we would take about the removal of the observation from the data set on which the reasonable frontier is based. The decision to remove an observation has to be based on the ordering of the extreme values under the outlier detection method used.

[TABLE 2 ABOUT HERE]

The first row of Table 2 deals with Hospital 22 (*H22*), which was considered for inclusion in the set of extreme and influential observations. *LP [9]* returns an efficiency of 45.92%, indicating that this hospital could reduce its *outputs* to about 46% of their current levels and still be efficient. If we look at the complete sample, this is the hospital that is most distant from the frontier on the super-efficiency side and, as such, it is the principal candidate for extreme observation status. When hospital *H22* is removed from the data set and the efficiency of all remaining hospitals is recalculated using *LP [13]*, the average efficiency level decreases from 110.66% to 110.43%. According to Wilcoxon´s test, this is a significant change at the 1% probability level. Based on the outlier removal procedure that we have suggested, this hospital should be removed from the sample. A similar conclusion would be reached if Wilson's method were used. Wilson proposes two outlier detection criteria. Under both of them *H22* takes second place in the ranking of extreme observations, the decision rule would suggest that *H22* be classified as an outlier. This is in agreement with our iterative procedure.

Differences between the two methodologies begin to appear after iteration 3. Iteration 3 identifies hospital 75 as an outlier under Wilcoxon's test procedure. Under Wilson's non-iterative procedure, *H75* would never be removed from the data set. It can be seen in Table 2 that its presence affects the efficiency of a single hospital, and its removal does not change the total efficiency. For another example, take iteration 10. In iteration 10 we see that our iterative process classifies *H19* as an influential extreme observation, but that this hospital would never be classified as an outlier under Wilson's procedure (it would be ranked as number 32 or 33 extreme observation).

Data masking problems appear to crop up after iteration 11, and are also present in iterations 12, 15, and 17. The methodology suggested here classifies hospitals *H87*, *H23*, *H35*, and *H64* as extreme observations but only after other observations, much more extreme than these, have been removed. This ability to cope with masking is absent from Wilson's (1995) methodology, since it concentrates on the observations that define the original frontier and does not contemplate the possibility that some observations may be revealed as extreme when other, more extreme observations, have been removed from the frontier. Only the methodology presented here can cope with the problem of masked extreme observations.

Table 2 gives examples of hospitals that would be classified as extreme observations under Wilson's method, but are not detected as such by our iterative method. Hospitals *H27* and *H30* are such cases. This reversal of fortunes may be due the removal of hospitals in previous iterations, that has an impact on the statistical test on whose significance we base outlier detection.

[TABLE 3 ABOUT HERE]

The application of the iterative procedure results in the removal of 18 hospitals, and in the construction of a final frontier that we describe here as a reasonable frontier. This frontier could form the basis for the implementation of an incentive scheme along the lines discussed in earlier sections. Table 3 shows that the reasonable frontier is defined by only 26 hospitals, about double the number of efficient observations in the first iteration. All efficiency levels have been reduced in the reasonable frontier: average super-efficiency is lower (and near 100%), the standard deviation of the efficiency coefficients is lower, and both minimum and maximum levels of efficiency are nearer 100%. These results ensure that each efficient hospital has a neighbouring point in the frontier, and this ensures that the efficiency of the entire sample does not depend much on the presence or absence of a particular hospital. In other words, what we describe as a reasonable frontier is statistically independent of the presence or absence of a particular efficient hospital, as the removal of any observation does not have a significant impact on the rest of the sample: the role of the hospital removed as a reference point is just taken up by another neighbouring efficient hospital.

5 Discussion

The conflict of interests between principal and agent is central to Agency Theory. The principal may attempt to force the agent to internalise his/her interests by making use of control mechanisms such as supervision or incentive schemes. This paper discusses an incentive scheme based on Agency Theory designed to motivate hospitals (agents) to maximise the achievement of targets set by the health authority (principal), in a situation where agents make decentralised decisions on the amount of effort they apply, where information is non-symmetric and cannot be verified by third parties, and where the production functions of individual hospitals are unknown.

The incentive scheme suggested here rewards the agent on the basis of the distance that separates observed output, the result of past decisions, and the potential output that would maximize the principal's benefit; Holmström (1982). But setting up such a scheme is challenging, since we have to estimate potential output and to do this we need to know the production function faced by each hospital - a knowledge we normally ignore. To overcome this difficulty we resort to *DEA*. This particular application of *DEA* identifies a frontier that sets *first best* effort levels; i.e., the level of effort attributed to those hospitals that are successful at delivering the greatest level of outputs for a given consumption of resources. A very delicate issue in this context is the selection of the hospitals that define the frontier. If the frontier contains hospitals whose performance cannot be matched, efficiency targets will be distorted for the remaining hospitals, and this may create disincentives rather than incentives, since the setting of an unrealistic target may result in the hospital we are trying to motivate giving up the challenge.

The identification of hospitals whose performance cannot be matched has been set within the context of influential observations detection. We propose an iterative procedure that has the advantage of identifying as outliers influential observations not previously on the frontier that become influential once the obvious outliers have been removed from the calculations. Removing both obvious outliers and outliers whose presence was previously masked facilitates the achievement of a compromise on the reward system in the negotiation between managers and the health authority. The actual algorithm that we use is based on a simple modification of the one proposed by Charnes et al. (1978). This algorithm is equivalent to the one suggested by Andersen and Petersen (1993) when constant returns to scale apply. This approach avoids the problem of undefined efficiency indexes, a problem found in the outlier detection method suggested by Wilson (1995).

5.1 Major findings

We have applied our methodology using a sample of hospitals taken from the Spanish public health service. 18 hospitals were identified as influential observations. They were excluded from the frontier used to set targets for other hospitals. Wilson's (1995) methodology was also applied to the same data set: It did not detect around 10% of the observations that were classified as very influential by our method. This is due to the problem of masking: some observations only appear as outliers when other, more extreme observations, are removed.

The result of the iterative procedure is the identification of a reasonable frontier made up of 26 hospitals – double the number of hospitals initially identified as efficient–. Negotiations with hospital managers can take place, on the basis of this reasonable frontier, to set up an incentive scheme. The level of potential output defined by this new frontier has three valuable properties. First, the target level of output can be reached, in the sense that target setting is not contaminated by the presence of hospitals whose performance cannot be matched (outliers). Second, the required level of output does not depend on a particular hospital, as the removal of a hospital does not modify the efficiency of the rest. And, third, the use of the super-efficiency concept of Andersen and Petersen (1993) has the advantage enabling the setting of incentive schemes for the managers of efficient hospitals, as it makes it possible to reward efficient hospitals on the basis of what they would produce on the frontier and what they are actually producing (which is less).

5.2 Study limitations

The methodology proposed in this paper would be useful in the design of incentives for the healthcare sector as well as other public sectors. The so-called "reasonable frontier" could be an excellent tool to help the setting of appropriate targets in an incentive-based payment scheme. However, our findings are dependent on available data. In the selection of variables, if we are interested in the final impact on hospital services, the number of patients treated weighted by its Group Diagnostic Related classification will be used; Murray (1992). When the focus is on the evaluation of the hospital activity rather than on the results of production of health, i.e. to identify the efficiency of hospital management, there are reasons to prefer indicators of throughput, such as the number of hospital-days; Chilingerian and Sherman (1990). Accordingly, to fulfill the main goal of this investigation (i.e. the design of incentives to improve management efficiency), and to due to availability of information, our attention is on throughput variables.

It would be of interest to identify efficiency improvements across time periods. In this case, a reimbursement scheme could depend on efficiency scores calculated by our reasonable frontier and on efficiency improvements across years (i.e. the distance between two reasonable frontiers). By introducing this second dimension in the incentive scheme, the inherent problem of relative efficiency could be solved. The frontier measures the distance from best practice, but this is relative to what other hospitals do. Since all hospital managers face the problem of moral hazard, it is to be presumed that moral hazard is present in all hospitals; it follows that we are not asking hospitals to behave in the best interests of the principal (i.e. the health authority), but just to minimise their greediness, to be as greedy as the less greedy managers in the sample. Unfortunately, our database is cross-sectional and it does not permit control of this possibility.

Another important feature of our iterative method is its conservatism. The identification of "masked outliers" depends on the previous elimination of other, more influential, outliers. Our intention here is to ensure that reliable data won't be catalogued as outliers. However, if this order generates problems, the process can be improved by considering the effects of additional exclusions; this could be similar to the non-iterative proposal made by Dusansky and Wilson (1994).

5.3 Implications

This paper is directed at those responsible for providing government services and those accountable for their delivery in an effective manner. It should encourage people to think about how more detailed and rigorous analysis of performance can assist in improving efficiency. These improvements will have repercussions on government budgets, and public service providers could render more and better service.

In most *OECD* countries, there is a concerted effort to make the public health system more efficient. Within the ambitious package of measures taken, managerial incentives play a central role in driving public hospitals to control expenditure, while producing health services in such a way that current needs are met.

Although the paper is based on a universal interpretation of principal-agent theory, incentive regulation of healthcare, as well as other public sectors, presents distinctive characteristics that demand innovative proposals. If healthcare technology was known, health services were marketed in the competitive arena, public authorities had clear and unconflicting preferences, or information was symmetric or verifiable by third-parties, innovative proposals would not be necessary; we could just apply traditional corporate governance mechanisms. But this is not the case. The contextual characteristics of the healthcare sector are: its products are not traded in organized markets; its complex technologies, political authorities have multiple and conflicting goals; and there are informative uncertainties related to future outcomes and agents (i.e. hospital managers) take decisions with more information than the principal (i.e. health authorities) has.

In this context, *DEA* is an excellent tool in the design of a system of incentives in the healthcare industry. Governments can use efficiency scores from *DEA* in several ways: to facilitate monitoring of managers and improve accountability within government; to promote "yardstick competition" by providing a means of comparing the performances of those responsible for similar aspects; to analyze the relationships between agencies and between programs, to allow governments to coordinate policy across agencies; and to assist the budgeting process by providing means of allocating funds on agreed plans for improved performance, rather than on the assumption that performance should equal part levels.

Attempting to measure performance provides a heightened awareness of data shortcomings for managers and policy makers. If data deficiencies are catalogued and advertised, the quality of data may be improved, thus enhancing the ability to measure performance. The reliability of the performance measures must be subjected to scrutiny. *DEA*, as other efficiency measurement techniques, is sensitive to outliers. Outliers can arise from either measurement or reporting error (maybe deliberate); or can reflect significant efficiencies being achieved by particular hospitals; hospitals which use different production technologies. In these cases, hospitals should be checked to determine whether they could be used with comparative purposes. This is the subject of the present article and the "reasonable frontier" its final outcome.

The "reasonable frontier" is a useful tool for policy makers. First, it is well suited for performance evaluations of multi-input—multi-output technologies, such as those of hospitals. Second, it is based on the concept of super-efficiency. From the pioneering work of Bogetoft (1994), it is demonstrated that the traditional *DEA* efficiencies provide incentives to reach the best-practices frontier, but not to surpass it. Thus enhancing the performance of the best performers is a challenge: They have no incentives to improve; extra effort is not accorded extra value, and yet the extra effort incurs extra cost. Super-efficiency solves this problem of motivating managers in efficient hospitals. Third, the "reasonable frontier" ensures that comparative performance reporting is not contaminated by the presence of hospitals whose performance cannot be matched. If such hospitals are included, performance inefficiencies are over-estimated and this could negatively affect managers' motivation. Additionally, removing these extreme hospitals facilitates the achievement of a compromise about the reward system when managers negotiate with the governmental authorities.

Finally, it is important to emphasize the complexity involved in the design of a regulation system. *DEA* could help with the setting of appropriate targets in an incentive-based payment scheme, but principal-agent conflict is just one of many relevant concerns. In future investigations, however, other topics such as organizational values, personal values and intrinsic motivation, team motivation, inter-temporal aspects will be taken into account.

References

Agrell, P.J., P. Bogetoft and J. Tind, 2002. Incentive Plans for Productive Efficiency, Innovation and Learning, *International Journal of Production Economics* 78, 1-11.

Andersen, P., and N.C. Petersen, 1993. A Procedure for Ranking Efficient Units in Data Envelopment Analysis, *Management Science* 39, 1261-1264.

Banker, R.D., 1993. Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation, *Management Science* 39, 1265-1273.

Banker, R.D., A. Charnes and W.W. Cooper, 1984. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science* 30, 1078-1092.

Banker, R.D., and R.M. Thrall, 1992. Estimation of Returns to Scale Using Data Envelopment Analysis, *European Journal of Operational Research* 62, 74-84.

Bogetoft, P., 1994. Incentive Efficient Production Frontiers: An Agency Perspective on DEA, *Management Science* 40, 959-968.

Bogetoft, P., 1995. Incentives and Productivity Measurements, *International Journal of Production Economics* 39, 67-81.

Breton, A., and R. Wintrobe, 1975. The Equilibrium Size of a Budget Maximizing Bureau: A Note on Niskanen's Theory of Bureaucracy, *Journal of Political Economy* 83, 195–207.

Byrnes, P., and V. Valdmanis, 1994. Analyzing Technical and Allocative Efficiency of Hospitals. In A. Charnes, W.W. Cooper, A.Y. Lewin and L.M. Seiford (Eds.), *Data Envelopment Analysis: Theory, Methodology and Applications*. Norwell, MA: Kluwer Academic Publishers, 129-144.

Charnes, A., W.W. Cooper and E. Rhodes, 1978. Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research* 2, 429-444.

Chilingerian, J.A., and H.D. Sherman, 1990. Managing Physician Efficiency and Effectiveness in Providing Hospital Services, *Health Services Management Research* 3, 3-15.

Dusansky, R., and P.W. Wilson, 1994. Technical Efficiency in the Decentralized Care of the Developmentally Disabled, *Review of Economics and Statistics* 76, 340-345.

Dusansky, R., and P.W. Wilson, 1995. On the Relative Efficiency of Alternative Methods of Producing Public Sector Output: The Case of Developmentally Disabled, *European Journal of Operational Research* 80, 608-628.

Ezzamel, M., and C. Mar-Molinero, 1990. The Distributional Properties of Financial Ratios in UK Manufacturing Companies, *Journal of Business Finance & Accounting* 17, 1-30.

Fox, K.J., R.J. Hill and W.E. Diewert, 2004. Identifying Outliers in Multi-Output Models, *Journal of Productivity Analysis* 22, 73-94.

Giokas, D.I., 2001. Greek Hospitals: How Well Their Resources Are Used, *Omega* 29, 73-83.

Gonzalez, B., and P. Barber, 1996. Changes in the Efficiency of Spanish Public Hospitals after the Introduction of Program-Contracts, *Investigaciones Económicas* 20, 377-402.

Gonzalez, E. and J. Ventura, 2004. Efficiency and Program-Contract Bargaining in Spanish Public Hospitals, *Annals of Public and Cooperative Economics*, forthcoming.

Holmström, B., 1982. Moral hazard in Teams, *The Bell Journal of Economics* 13, 324-341.

Kittelsen, S.A.C., 1993. Stepwise DEA. Choosing Variables for Measuring Technical Efficiency in Norwegian Electricity Distribution, University of Oslo, Department of Economics, mimeo.

Kuntz, L., and S. Scholtes, 2000. Measuring the Robustness of Empirical Efficiency Valuations, *Management Science* 46, 807-823.

Laffont, J.J., and J. Tirole, 1986. The Dinamics of Incentive Contracts, *Econometrica* 56, 614-641.

Leibenstein, H., 1966. Allocative Efficiency vs X-Efficiency, *The American Economic Review* 56, 392-415.

Murray, R., 1992. Measuring Public-Sector Output: The Swedish Report. In Zvi Griliches (Ed.), *Output Measurement in the Service Sectors*, National Bureau of Economic Research, The University of Chicago Press, 517-542.

Niskanen, W., 1971. *Bureaucracy and Representative Government.* Aldine-Atherton, Chicago.

Niskanen, W., 1975. Bureaucrats and Politicians, *Journal of Law and Economics* 18, 617–643.

Pastor, J.T., J.L. Ruiz and I. Sirvent, 1999. A Statistical Test for Detecting Influential Observations in DEA, *European Journal of Operational Research* 115, 542-554.

Timmer, C.P., 1971. Using a Probabilistic Frontier Production Function to Measure Technical Efficiency, *Journal of Political Economy* 79, 776-794.

Tirole, J., 2001. Corporate Governance, Econometrica 69, 1-35.

Tullock, G., 1965. *The Politics of Bureaucracy*. Public Affairs Press, Washington, D.C.

Verheyen, P., 1998. The Missing Link in Budget Models of Nonprofit Institutions: Two Practical Dutch Applications, *Management Science* 44, 787-800.

Wilson, P.W., 1995. Detecting Influential Observations in Data Envelopment Analysis, *The Journal of Productivity Analysis* 6, 27-45.

Wilson, P.W., and J.M. Jadlow, 1982. Competition, Profit Incentives, and Technical Efficiency in the Provision of Nuclear Medicine Services, *The Bell Journal of Economics* 13, 472-482.

Worthington, A.C., 2004. Frontier Efficiency Measurement in Health Care: A Review of Empirical Techniques and Selected Applications, *Medical Care Research and Review* 61, 135-170.

Table 1. Descriptive Statistics

		Mean	Stand. Dev.	Minimum	Maximum
Panel A. Initial sample. N = 137					
Inputs	x_1:*Medical staff*	569.83	361.97	200.00	1995.00
	x_2: *Other staff*	589.55	427.17	66.34	2447.96
	x_3: *Beds*	1597.57	1238.75	274.55	6878.60
	x_4: *External purchases*	3100886.18	2536931.12	397836.00	13322528.00
Outputs	y_1: *Discharges*	21081.72	13169.63	3775.00	67787.00
	y_2:*Bed-days*	171058.50	111330.92	31460.00	542702.00
	y_3: *Out-patients*	159007.64	98274.04	31143.00	480810.00
Efficiency	*CCR (LP [5])*	112.05%	9.51%	100.00%	146.38%
	A-P (LP [9])	110.66%	12.69%	45.92%	146.38%

		Mean	Stand. Dev.	Minimum	Maximum
Panel B. Final sample. N = 119					
Inputs	x_1:*Medical staff*	608.54	370.07	201.00	1995.00
	x_2: *Other staff*	642.45	431.18	165.30	2447.96
	x_3: *Beds*	1747.61	1255.14	430.89	6878.60
	x_4: *External purchases*	3399683.29	2579357.12	755371.00	13322528.00
Outputs	y_1: *Discharges*	22528.40	13376.00	7281.00	67787.00
	y_2:*Bed-days*	183040.67	113534.19	55526.00	542702.00
	y_3: *Out-patients*	162131.62	98767.35	31143.00	480810.00
Efficiency	*CCR (LP [5])*	108.66%	7.21%	100.00%	125.61%
	A-P (LP [9])	107.38%	9.12%	88.93%	125.61%

		Mean	Stand. Dev.	Minimum	Maximum
Panel C. Outliers. N=18					
Inputs	x_1:*Medical staff*	313.94	133.31	200.00	670.00
	x_2: *Other staff*	239.82	142.15	66.34	700.35
	x_3: *Beds*	605.58	376.25	274.55	1775.21
	x_4: *External purchases*	1125505.22	742500.47	397836.00	3350354.00
Outputs	y_1: *Discharges*	11517.50	5840.51	3775.00	30011.00
	y_2:*Bed-days*	91843.06	45116.69	31460.00	210177.00
	y_3: *Out-patients*	138354.61	95032.82	33520.00	385693.00

Table 2. Influential observations

ITERATIVE PROCEDURE					*WILSON*					*DECISION*		
Iteration	*Potentially influential hospital*	*Andersen-Petersen (CRS)*	*Average sample efficiency including r*	*Wilcoxon's Z*	*Potentially influential hospital*	*Andersen-Petersen (VRS)*	*DMUs whose efficiency changes when r is removed*	*Sum of efficiency changes / nº observations*	*Total efficiency change*	*Wilson*		*Our procedure*
	r	$\hat{q}^{r*}$	$E\left(\hat{q}^{i*}\right)$	*Z*	*r*	$\hat{a}^{r*}$	n^r	$\partial^r = \frac{\sum_{i, i \neq r} (\hat{a}_{-r}^{i*} - \hat{a}^{i*})}{n^r}$	$\overline{n}^r \times \partial^r$ ($\overline{n}^r = n^r + n.d.$)	*Ranking* $\hat{a}^{r*}$	$\overline{n}^r \times \partial^r$	*Removed*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
0	H22	45.92%	110.66%	-3.92***	*H22*	43.14%	17	4.42%	75.15%	2	2	Yes
1	H20	57.31%	110.43%	-3.29***	*H20*	n.d.	11	2.10%	29.36%	1	6	Yes
2	H25	71.29%	110.45%	-2.36**	*H25*	78.26%	11	1.27%	13.94%	6	15	Yes
3	H75	75.91%	110.73%	-2.36**	*H75*	78.93%	1	0.00%	0.00%	8	37	Yes
4	H98	80.60%	110.61%	-2.93***	*H98*	72.50%	24	0.84%	20.14%	4	11	Yes
5	H31	66.99%	110.59%	-4.45***	*H31*	88.07%	16	0.35%	5.53%	17	22	Yes
6	H107	79.64%	110.66%	-3.70***	*H107*	73.88%	10	1.37%	13.68%	5	16	Yes
7	H104	78.10%	110.67%	-4.10***	*H104*	84.07%	7	1.21%	8.45%	11	17	Yes
8	H28	83.44%	110.73%	-7.10***	*H28*	86.20%	48	1.05%	50.20%	15	4	Yes
9	H47	74.87%	110.07%	-9.10***	*H47*	82.46%	57	1.58%	89.90%	10	1	Yes
10	H19	83.30%	107.89%	-4.45***	*H19*	99.59%	3	0.15%	0.44%	33	32	Yes
11	H87	84.83%	107.73%	-3.35***								Yes
12	H23	84.24%	107.54%	-2.42**								Yes
13	H26	87.20%	107.67%	-4.37***	*H26*	98.02%	3	0.25%	0.75%	27	28	Yes
14	H50	87.76%	107.59%	-1.75*	*H50*	n.d.	2	0.00%	0.00%	1	37	Yes
15	H35	87.99%	107.71%	-6.14***								Yes
16	H3	88.98%	107.45%	-3.40***	*H3*	84.53%	26	1.30%	33.76%	12	5	Yes
17	H64	87.73%	107.34%	-2.66***								Yes
18	H82	88.93%	107.38%	-1.60								No

Notes:

***/**/* the null hypothesis can be rejected at the 1%/5%/10% level. $\hat{a}^{r**}$ y $\hat{a}_{-r}^{i**}$ are the equivalent results using variable returns to scale (VRS) to solutions using [8] y [12], $\hat{q}^{r**}$ y $\hat{q}_{-r}^{i**}$. *n.d.*: super-efficiency values not defined.

Table 2. Continuation

ITERATIVE PROCEDURE					*WILSON*					*DECISION*		
Iteration	*Potentially influential hospital*	*Andersen-Petersen (CRS)*	*Average sample efficiency including r*	*Wilcoxon's Z*	*Potentially influential hospital*	*Andersen-Petersen (VRS)*	*DMUs whose efficiency changes when r is removed*	*Sum of efficiency changes / nº observations*	*Total efficiency change*	*Wilson*		*Our procedure*
	r	$\hat{q}^{r*}$	$E\left(\hat{q}^{i*}\right)$	*Z*	*r*	$\hat{a}^{r*}$	n^r	$\partial^r = \frac{\sum_{i, i\neq r}(\hat{a}^{i*}_{-r} - \hat{a}^{i*})}{n^r}$	$\overline{n}^r \times \partial^r$ $(\overline{n}^r = n^r + n.d.)$	*Ranking* $\hat{a}^{r*}$	*Ranking* $\overline{n}^r \times \partial^r$	*Removed*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
					H17	99.82%	1	0.00%	0.00%	36	36	No
					H27	78.32%	33	0.75%	24.64%	7	9	No
					H33	99.80%	3	0.22%	0.65%	35	29	No
					H30	80.42%	42	1.70%	71.37%	9	3	No
					H118	68.92%	7	1.05%	7.34%	3	18	No
					H95	96.10%	39	0.64%	24.88%	25	8	No
					H51	89.32%	7	0.40%	2.82%	18	26	No
					H119	92.51%	1	0.60%	0.60%	21	30	No
					H37	96.00%	21	0.75%	15.84%	24	13	No
					H18	92.53%	14	0.42%	5.93%	22	21	No
					H16	98.56%	13	0.22%	2.92%	29	25	No
					H21	98.23%	11	0.31%	3.40%	28	24	No
					H24	99.63%	6	0.09%	0.51%	34	31	No
					H46	93.88%	4	0.92%	3.66%	23	23	No
					H48	99.37%	7	0.06%	0.39%	32	33	No
					H68	85.93%	5	4.41%	22.03%	14	10	No
					H74	90.64%	4	1.66%	6.63%	19	19	No
					H110	98.83%	4	0.07%	0.26%	30	34	No
					H111	99.96%	3	0.01%	0.04%	37	35	No
					H113	96.93%	3	0.27%	0.82%	26	27	No
					H114	85.10%	38	0.48%	18.29%	13	12	No
					H121	98.88%	33	0.18%	6.08%	31	20	No
					H122	87.69%	17	1.69%	28.65%	16	7	No
					H133	91.49%	18	0.79%	14.16%	20	14	No

Table 3. Reasonable frontier against initial frontier

Panel A. Initial frontier. N = 13					
		Mean	Stand. Dev.	Minimum	Maximum
Inputs	x_1:*Medical staff*	369.69	209.89	200.00	871.00
	x_2: *Other staff*	274.89	209.53	66.34	781.91
	x_3: *Beds*	667.63	453.45	274.55	1701.65
	x_4: *External purchases*	1482982.38	1375362.03	410986.00	5203863.00
Outputs	y_1: *Discharges*	12615.69	7213.68	6214.00	29533.00
	y_2:*Bed-days*	117113.85	75456.39	51005.00	300428.00
	y_3: *Out-patients*	139874.62	95081.02	33520.00	322663.00
Efficiency	*CCR (LP [5])*	85.34%	14.39%	45.92%	97.67%

Panel B. Reasonable frontier. N = 26					
		Mean	Stand. Dev.	Minimum	Maximum
Inputs	x_1:*Medical staff*	444.46	230.13	201.00	880.00
	x_2: *Other staff*	426.96	270.99	165.30	1184.29
	x_3: *Beds*	1083.38	624.34	430.89	2559.05
	x_4: *External purchases*	2240268.62	1908837.04	755371.00	7380664.00
Outputs	y_1: *Discharges*	17222.27	9286.69	7688.00	40250.00
	y_2:*Bed-days*	135638.46	76424.01	55526.00	300428.00
	y_3: *Out-patients*	149073.73	87199.78	48262.00	331801.00
Efficiency	*CCR (LP [5])*	94.12%	3.63%	88.93%	99.58%