**INSTITUT DE STATISTIQUE**

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

**DISCUSSION PAPER**

**0216**

# ANALYSIS OF THE EXPENSES LINKED TO HOSPITAL STAYS : HOW TO DETECT OUTLIERS?

Cl. BEGUIN and L. SIMAR

http://www.stat.ucl.ac.be

# ANALYSIS OF THE EXPENSES LINKED TO HOSPITAL STAYS:

## HOW TO DETECT OUTLIERS?


Beguin C.*, Simar L.**[1]

* Centre d’Informatique Médicale, Cliniques Universitaires Saint Luc, Université Catholique de Louvain
** Institut de Statistique, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

Corresponding author:

C.Beguin

Cliniques Universitaires Saint Luc - Université Catholique de Louvain

Centre d’Informatique Médicale

10 avenue Hippocrate

1200 Bruxelles, Belgium

phone 32-2-7644711

fax 32-2-7644717

cbeguin@infm.ucl.ac.be



[1] Research support from « Projet d'Actions de Recherche Concertées » (No. 98/03--217) from the Belgian Government is acknowledged.

**Abstract**

Since the hospital financing is more and more based on a budget in function of the pathologies, it appears necessary to detect hospital stays presenting discrepancy between the resources they used and the medical characteristics they present. We propose the use of deterministic nonparametric frontier models to rank hospital stays in function of their expenses taking into account the severity level of the patient. In this case, the hospital stays with the lowest total expenses for a given level of severity are considered as efficient. But, deterministic models are very sensitive to the extreme stays so that some efficient stays could be in fact 'too' efficient and considered as outliers. We try to highlight these stays using the method of Simar (2001) which is based on the concept of order-m frontier. The idea is to define a 'benchmark' frontier that doesn't envelop all the data. Indeed, for a given level of severity, a reasonable minimal value of the expenses is estimated. The hospital stays with expenses lower than this reasonable value could be considered as 'outlier' and require further analysis. We compare our results with these obtained in Beguin (2001) which relies on a method proposed by Wilson (1995). As a conclusion, we recommend the use of both methods to detect the outlier and so to obtain a better ranking of the other hospital stays.

Key-words: deterministic nonparametric frontier models, free disposal hull, outliers

## 1. Introduction

Since 1995, the Belgian government introduced diagnosis-related groups (DRGs) to the hospital payment system (Fetter et al, 1980). The hospitals are paid not only as a function of the days of hospitalisation and of the reimbursement of medical acts, but also in function of length of stay (LOS) by DRG. Hospitals with mean LOS greater or less than the national mean are penalized or rewarded accordingly. However, the distribution of LOS by DRG is not symmetrical (Marazzi et al, 1998), so that the estimation of the mean could be biased. Until now the Ministry of Health estimates the mean LOS after excluding of outliers defined by rules based on the quartiles and the interquartile range. Each year, these outliers count for approximately 4% of the Belgian database and represent a non negligible number of hospital stays. This method is being extended to medical fees or to total expenses. Nevertheless, it does not take into account the characteristics of the patient and it does not explore other methods to point out abnormal stays.

In Beguin (2001), the use of deterministic frontier models is proposed to highlight hospital stays showing discrepancies between the total expenses, LOS and/or medical fees and the severity score of the patient. This application is somewhat different of that usually seen when frontier models are used in health care sector (Grosskopf et al 1987; Linna et al, 1998; Chilingerian, 1995; Puig-Junoy, 1998 ). In our application, the frontier model is used to rank hospital stays according to their total expenses by taking into account patient characteristics. In our framework, the frontier will correspond to the minimal achievable expenses for a particular stay of a patient having a given level of severity score. A stay on the frontier will be considered as 'efficient'. Deterministic frontier models are based on envelopment methods (DEA, FDH) so that the estimated frontier envelops all the data points.

Thus they are very sensitive to extreme points, which in our framework can be seen as 'efficient outlier'. The method developed by Wilson (1995) for detecting outliers has been adapted in Beguin (2001): the 'too efficient' outliers are eliminated before definitive estimation of the efficiencies of the hospital stays. It allows to point out the too efficient stays needing further scrutiny. It can be viewed as an exploratory analysis of data.

In this paper, we propose another method proposed by Simar (2001) to highlight the too efficient stays. This method is based on the concept of the minimum input function and the order-m frontier developed in Cazals-Florens-Simar (2002) (CFS). The results obtained by both methods will be compared.

In section 2, the frontier models and the FDH estimator are presented, we also summarize the two methods proposed to detect outliers in frontier models. The concept of the minimum input function and the order-m frontier is largely described. Section 3 describes the data. In a last section, results are presented and discussed.

## 2. Theoretical background

### 2.1. Definition of a frontier

The objective is to find the minimal achievable level of total expenses taking into account the characteristics of the patients such as age, number of dysfunctioning system, length of stay in intensive care unit (ICU) which can be considered as "proxy" for measuring the severity score of the patient, but also sex and emergency versus non-emergency admission, other qualitative characteristics of the patient which could be of interest. The minimal achievable level is considered as a "frontier" of a possibility set, which is the set of attainable combinations of total expenses with individual characteristics of the patient. For a specific level of the patient's characteristics, the minimal value of total expenses delimits the frontier enveloping all the possible values. The efficiency of an hospital stay will then be defined as the distance between the observed value of total expenses and the frontier. This type of problem appears in productivity analysis when looking for the minimal sets of inputs of a firm, which allow it to produce a given level of outputs. However, the terminology used in the context of productivity is not well adapted to the problem here. Our frontier model can be formalized as follows.

The set $\Psi$ of possibilities is the set of attainable points (x, y). It is defined by $\Psi = \left\{ (x, y) \in \Re_{+}^{p+q} \,\middle|\, (x, y)\ \textit{are attainable} \right\}$ where $x$ are the quantities relative to the total expenses of a stay and $y$ are the individual severity scores of the patient. For all possible values of y, the section of possible values of $x$ is the set defined as $X(y) = \left\{ x \in \Re_{+}^{p} \,\middle|\, (x, y) \in \Psi \right\}$ and its efficient boundary is the subset of $X(y)$ defined by :

$\partial X(y) = \left\{ x \,\middle|\, x \in X(y), \theta x \notin X(y),\ \forall\, 0 < \theta < 1 \right\}$. A measure of the efficiency of a particular stay $(x_k, y_k)$ can then be expressed by $\theta_k = \min\left\{ \theta : \theta x_k \in X(y_k) \right\}$. It is the radial distance from $x_k$ to the efficient boundary $\partial X(y_k)$. If $\theta_k = 1$, the stay $(x_k, y_k)$ is considered as being 'efficient' in the sense that it achieves the minimal attainable level of $x_k$ given $y_k$. The efficiency score $\theta_k < 1$ represents the feasible proportionate reduction of $x_k$ (total expenses) given the characteristics $y_k$ of the patient in order to be considered as being efficient. Nonparametric estimators of frontiers have been proposed in the literature in the framework of productivity analysis. These estimators rely on the idea that the attainable set is defined by the set of minimal volume containing all the observations. The DEA estimator (Data Envelopment Analysis) (Farrell, 1957) relies on convexity of the set $\Psi$ whereas the more recent FDH (Free Disposal Hull) (Deprins et al, 1984) does not impose such a restriction. We will use this more flexible estimator in what follows. Simar and Wilson (2000) propose a survey of the statistical properties of these envelopment estimators. The FDH estimator of $\Psi$ based on a sample of n observations $(x_i, y_i)$ is the free disposal closure of the reference set $\{(x_i, y_i) | i = 1,....,n\}$. It can be defined as follows: $\hat{\Psi}_{FDH} = \{(x, y) | x \le x_i, y \ge y_i, i = 1,....,n\}$. The efficiency score of a particular stay $(x_k, y_k)$ is given by $\hat{\theta}(x_k, y_k) = \min_k \{\theta | (x_k, y_k) \in \hat{\Psi}_{FDH}\}$. By construction, $\hat{\theta}(x_i, y_i) \le 1$ for all observed stays $(x_i, y_i)$ and a stay is efficient when $\hat{\theta}(x_k, y_k) = 1$.

**2.2. Detection of outliers with the method of Wilson**

The FDH estimator of the frontier is deterministic, as opposed to stochastic frontier models, in the sense that all observed stays are supposed to belong to $\Psi$. Consequently, it is highly sensitive to extreme observed stays since the FDH envelops all the data points. It is necessary to detect these extreme stays that pull the frontier away from the other observed stays so that the latter appear less efficient. This problem has been analyzed by Wilson (1993, 1995). In summary, the method to detect outliers consists of removing all stays one by one and re-estimating a new frontier on the remaining data (Wilson, 1993, 1995; Andersen and Petersen, 1989; Lovell et al, 1993). The efficiency $\hat{\theta}_{(i)}$ of the stay *i* when computed from a reference set which does not include the $i^{\text{th}}$ observation, may be greater than 1 . When $\hat{\theta}_{(i)}$ is greater than 1, the hospital stay is suspected to be an outlier. The second step is to estimate the impact of the exclusion of the efficient stay *j* on the efficiencies of the other stays. The efficient stay *j* is removed and also all other stays one by one. So for each of the other stays, we have two measures of the efficiencies. The impact of the efficient stay *j* is evaluated by computing the difference of the efficiency measures at step 1 when the reference set is the full set minus the stay *i* and the efficiency measures at step 2 when the reference set is the full set minus the stay *i* and the stay *j*. The difference is summarized by 2 statistics: the number of changes and the mean of these changes. The difficulty of this method is its non applicability to data sets with a great number of points.

A simplified adaptation of this method is proposed in Beguin (2001). In the latter, a frontier is estimated with all data. Then each efficient stay is removed one by one and the frontier is estimated once again on the N-1 remaining stays. It allows to detect hospital with an efficiency greater than 1. For the other hospital stays, the mean and the number of changes in efficiencies are calculated by comparison of the frontier estimated with all stays and the frontier estimated on N-1 stays. Nevertheless, the limit to be used to determine a stay as 'too efficient' remains arbitrary.

### 2.3. Robust non parametric estimators of the frontier and outliers

In the framework of deterministic models, Cazals, Florens and Simar (2002) propose a non parametric estimator of the frontier, robust to these extreme stays. It is based on the concept of the expected minimum input function of order-m.

The efficient frontier of $\Psi$, described above, can be expressed in another way using the distribution functions and the survival functions of $(x,y)$. Consider first the simplest univariate case, the lower boundary of X, the total expenses, can be defined by

$\varphi = \inf\{x \mid F_X(x) > 0\}$ where $F_X(x)$ is the distribution function of X. Equivalently, it can be rewritten by the survivor function so that $\varphi = \inf\{x \mid S_X(x) < 1\}$.

If we take into account the severity level *y*, it will be the lower level of the total expenses for the patients with at least a given level of severity. So, we have to introduce the concept of the conditional survivor $S_C(x \mid y) = P(X \geq x \mid Y \geq y) = \dfrac{S(x,y)}{S_Y(y)}$. In this case, the lower boundary of the expenses given $Y \geq y$ is now defined as $\varphi(y) = \inf\{x \mid S_C(x \mid y) < 1\}$. This defines a minimum expenses function as a function of *y*. CFS have proven that $\varphi(y)$ is monotone non decreasing in *y* and that $\varphi(y)$ is the largest monotone function which is smaller or equal to the efficient frontier $\partial X(y)$. In case of free disposability, the two functions coincide. Then, a nonparametric estimator of the frontier is given by the empirical counterpart of $S_C(x/y)$: $\hat{\varphi}_n(y) = \inf\{x \mid \hat{S}_{C,n}(x/y) < 1\}$ where $\hat{S}_{C,n}(x/y) = \dfrac{\hat{S}(x,y)}{\hat{S}_Y(y)}$ is the empirical conditional survival function of *x* conditionally to *y*. As pointed in CFS, this estimator $\hat{\varphi}_n(y)$ is the FDH estimator, the boundary of the free disposal hull of the observed data points.

Now in place of using a full frontier concept as above to determine the benchmark (the 'efficient frontier') to which we can compare each hospital stay, CFS propose a more flexible concept of frontier, the order-m frontier, which by construction does not have to envelop all the observed points and so will be more robust to extreme points. This benchmark will also be easy to estimate in a nonparametric way.

Consider $X^1, X^2, ......, X^m$ as the total expenses drawn from a sample of size m of hospital stays for patients having a severity score $Y \geq y$ . The order-m frontier is defined as $\varphi_m(y) = E\left[\min\left(X^1, X^2, ...., X^m\right) \mid Y \geq y\right]$. It represents the expected minimal value of the total expenses among a fixed number of m stays drawn from the population of stays with at least this level $y$ of severity. CFS show that this frontier can be calculated as follows: $\varphi_m(y) = \int_0^\infty \left[S_C(x|y)\right]^m dx$ and they show that for all possible values of y, $if\ m \to \infty,\ \varphi_m(y) \to \varphi(y)$, the full frontier. As in the case of the full frontier, a nonparametric estimator of $\varphi_m(y)$ is given by plugging the empirical conditional survival function of $X$ in the last formula: $\hat{\varphi}_{m,n}(y) = \int_0^\infty \left[\hat{S}_{C,n}(x/y)\right]^m dx$. An algorithm is proposed in CFS to compute it in practice. We note also that $\hat{\varphi}_{m,n}(y) \to \hat{\varphi}_n(y)\ as\ m \to \infty$.

The parameter m can be interpreted as a trimming parameter. Indeed for a fixed m, the estimator will not envelop all the data points and so will be less sensible to the extreme values than $\hat{\varphi}_n(y)$ was. If we look for a benchmark level of $x$ to see if the total expenses $x_i$ of a stay $(x_i, y_i)$ are too high, it is reasonable to compare this value $x_i$ to the value of $\hat{x}_i = \hat{\varphi}_{m,n}(y_i)$. The choice of m remains arbitrary but, for example, if m = 100, $\hat{\varphi}_{m,n}(y_i)$ could be a reasonable benchmark value since it is the estimated expectation of the minimum achievable expenses $x$ among 100 random hospital stays of patients having at least the level $y$ of severity score. The basic idea of Simar (2001) is to analyze the sensitivity of the frontier when varying the value of m. If m increases and if the stay $(x_i, y_i)$ is such that $x_i < \hat{\varphi}_{m,n}(y_i)$, even for large values of m, this stay is a potential outlier. In the next section, we describe how to apply this idea in practice.

### 2.4. Practical detection of outliers with the order-m frontier

The basic idea is that any point $(x_0, y_0) \in \Psi \subset R_+ \times R_+^q$ is outlying the cloud of points $\{(x_i, y_i) | i = 1,.....,n\}$ in the direction of the total expenses if $\hat{\theta}_{m,n}(x_0, y_0) > 1$ even if m increases, where here, since $x_0$ is univariate $\hat{\theta}_{m,n}(x_0, y_0) = \dfrac{\hat{\varphi}_{m,n}(y_0)}{x_0}$. In practice, for several reasonable values of m, $\hat{\theta}_{m,n}(x_i, y_i)$ is estimated for each data point. Of course, for finite m, $\hat{\theta}_{m,n}(x_i, y_i) \geq \hat{\theta}_{FDH,n}(x_i, y_i)$. So for every 'FDH-efficient' point, $\hat{\theta}_{m,n}(x_i, y_i) \geq 1$. So a value $\hat{\theta}_{m,n}(x_0, y_0) > 1$ does not automatically indicate a potential outlier, this depends on the value of m and on the extent to which $\hat{\theta}_{m,n}(x_0, y_0)$ is larger than one. For the latter, some reasonable threshold value has to be fixed to alert the existence of a potential outlier and for selecting m, a sensitivity analysis is proposed.

This is achieved in Simar (2001) through a semi-automatic procedure. The values of $\hat{\theta}_{m,n}(x_i, y_i)$ are computed for each data point with several values of m, say, 10, 25, 50, 75, 100, 150 (any other set of values could be chosen). Let $N_{input}(x_i, y_i)$ represent the number of stays in the sample analyzed with a severity level greater than $y_i$. Looking carefully at the results of $\hat{\theta}_{m,n}(x_i, y_i)$, along with $N_{input}(x_i, y_i)$ will help to detect potential outliers.

It is also necessary to choose several reasonable threshold values distant from 1 such as $1 + \delta$ where $\delta = 0.025, 0.05, 0.075, 0.10$. It is then helpful to produce a figure displaying the percentage of stays with $\hat{\theta}_{m,n}(x_i, y_i) \geq 1+\delta$ as a function of m, for the different values of $\delta$. These curves indicate the percentage of points outside the order-m frontier, as a function of m and $\delta$. If there is no outlier, the curves should decrease (approximately) linearly to 0 when m increases. Any strong deviation from linearity should indicate the existence of potential outliers. If the curves present an elbow, the points remaining outside the order-m frontier, for large values of m and for the chosen threshold $1+\delta$, are potential outliers. Then a careful analysis of these points should confirm if they are outlier or not. The procedure so helps to 'alert' the analyst of the presence of potential outlier but full-automatic data-driven method are never available in this kind of framework. In the application below, to simplify the presentation, we will choose $\delta = 0$, which can be viewed as a conservative approach.

## 3. Data and method

The objective is to work with data routinely collected in hospitals. At each patient discharge a minimal basic medical data set is collected. This set contains information such as age, sex, emergency versus non-emergency admission, destination of the patient after the hospital stay, length of stay in intensive care unit (ICU), diagnoses and procedures performed. Diagnoses and procedures are coded with the ICD-9-CM classification system allowing hospital stays to be grouped in Diagnosis Related Groups (DRG) (Fetter, 1980). As hospitals are not financed by global budgets but partly by a fixed day-price and partly by payment for each medical act, it is possible to establish a summary of the bill for each hospital stay. So the basic minimum medical data set of a hospital stay can be linked with the financial summary of the bill.

The database used in this study contains data from 34 hospitals collected during the second semester of 1988. DRG 122 (Acute myocardial infarction, discharged alive, with complicated cardiovascular diagnosis) was retained for further analysis. As there is no deceased patient on discharge in this DRG, censoring difficulties need not be taken into account in the statistical models. For this pathology, the variables analyzed are the total expenses and the characteristics of the patient: age, sex, admission in emergency, number of dysfunctioning systems and length of stay in ICU.

In order to define the efficient frontier, the variable to minimize is the total expenses taking into account age, the number of dysfunctioning systems and length of stay in ICU. One frontier is estimated for each level of the dichotomic variables: sex and emergency admission. The efficiency of each hospital stay is estimated by reference to the frontier of the corresponding level of the dichotomic variables.

We will estimate the full FDH frontier and the order-m frontier as described above. First, the frontier is estimated on the complete sample using the FDH estimator and minimizing *x*, the total expenses. As we are interested in outlying the too efficient stays, we continue the analysis on the efficient stays. So, for each efficient stay, the hospital stays with at least the same severity level ($Y \geq y$) are selected. For these stays, we intend to estimate $\hat{\varphi}_{m,n}(y) = \hat{E}\left[Min\left(X_1,\ldots,X_m \middle| Y \geq y\right)\right] = \int_0^\infty \left[\hat{S}_{C,n}(x/y)\right]^m dx$. CFS gives a formula to compute $\hat{\varphi}_{m,n}(y)$, but as shown in Simar (2001), the integral can be much more easily approximated as follows. B random samples with replacement of m stays are drawn among the selected stays having $Y \geq y$. For each of the B samples, the minimum of the total expenses is retained

$(\widetilde{X}_b)$. By the law of the large number, we know that $\frac{1}{B}\sum_{b=1}^{B}\widetilde{X}_b \approx \hat{E}\left(X_1, X_2, \ldots\ldots, X_B \middle| Y \geq y\right)$ as $B \to \infty$ so that we approximate $\hat{\varphi}_{m,n}(y)$ by $\frac{1}{B}\sum_{b=1}^{B}\widetilde{X}_b$. Finally, $\hat{\theta}_{m,n}(x_i) = \frac{\hat{\varphi}_{m,n}(y_i)}{x_i}$. If this efficiency is greater than 1 even for large values of m, the stay will be considered as being a potential outlier. In the analysis, B is fixed at 200 and m takes the values 10, 15, 20, 25, 30.

We will also detect the outliers with the same approach as in Beguin (2001) who used and adapted the method of Wilson (1995). This last method is called method B-W in the next paragraphs. After, we compare the results obtained with the order-m frontier to these of the method B-W.

## 4. Results

The sample contains 1643 hospital stays, among which 74.7% of males and 74.9% emergency admissions. The median age is 62 year (first quartile (q1) = 54, third quartile (q3) = 69). 86.5% of the patients have 1 dysfunctioning system and 1.7% have 3 or more dysfunctioning system. 44.7% of the patients were treated in ICU. The median total expenses is 92921BEF (q1 = 61729BEF and q3 = 144294 BEF). The length of stay in ICU is positively and significantly correlated with total expenses (Spearman r = 0.43, p = 0.0001). The same positive correlation is found for the number of dysfunctioning systems (Spearman r = 0.19, p = 0.0001). This shows that these proxy can be taken as a proxy of the severity score.

For the whole sample, the mean of efficiencies provided by the FDH model is 0.341 for total expenses and the proportion of efficient stays is 7.8% of all stays. For women, not admitted in emergency (n = 107), the mean of efficiencies is 0.479 and 20.6 % of the stays are efficient. For men, not admitted in emergency (n = 306), these values are respectively 0.279 and 9.2%. For women admitted in emergency (n = 311), the mean of the efficiencies is 0.460 and 9.3 % of the stays are efficient. For men admitted in emergency (n = 919), these values are respectively 0.306 and 5.4%. These figures for efficiency scores of hospital stays may be viewed as being too low, but this might be due to the presence of the too efficient stays. So it is certainly helpful to detect these potential outliers.

For the efficient stays defined by the FDH frontier model, we computed the order-m efficiency for m= 10, 15, 20, 25, 30. Table I shows for men not admitted in emergency the efficiency estimated using the order-m frontier when m takes the values 10, 20, 30. The second column displays $N_{input}(x_i, y_i)$, the number of stays with a severity level greater or equal to the severity of the stay analyzed. The last column shows the efficiency of the stay when the frontier is estimated without this stay as described in the B-W method.

Some stays, as stays 4 and 12, have an efficiency greater than 1, whatever the value of m in the order-m frontier and whatever the method used. These 2 stays have a high probability to be outliers. A careful analysis of the data shows they have very low value of the total expenses compared to the other stays with the same severity level, particularly for stay 12.

Other stays like stays 6, 8, 10, have an efficiency greater than 1 with the B-W method but not with the order-m frontier. For these stays, $N_{input}(x_i, y_i)$ is low which shows these points are really extreme compared to the observed cloud of points. For these points, very few stays have a greater severity level. With the B-W method, they have an efficiency greater than 1 confirming how extreme they are.

Inversely, some stays like stays 15, 16, 17, 18, have an order-m efficiency greater than 1 but their efficiency measured by the B-W method is equal to 1. A high number of stays have a severity level greater than the severity of these stays but with higher level of the total expenses. The B-W method fails to detect these points as being outlying.

Finally, a last group concerns the stays with an efficiency equal to 1 in both methods. One possibility is that these points are extreme but no similar points in their area allows to estimate their outlying character. Another possibility is that they could be just efficient point.

Note that the B-W method sometimes provides useful complementary information. For instance if we compare stay 14 and stay 26, both with $N_{input}(x_i, y_i)$ = 1, the B-W method indicates that stay 14 is a potential outlier but that no such clear conclusion can be drawn for stay 26.

Figure 1 displays the proportion of too efficient stays in function of the value of m. If m increases, the proportion of the too efficient stays diminishes. This decreasing is more important in subgroups with less stays (women, not admitted in emergency). More stability is observed for the values of m equal to 25 and 30, so we chose m = 30 for a reasonable benchmark frontier.

Table II shows the frequency of the too efficient stays detected by one or two of both methods. Globally (bottom row of Table II), 14 stays are too efficient in both methods, 27 stays are too efficient in the order-m frontier method but not in the B-W method, 37 stays are not too efficient in the order-m frontier method but well in the B-W method and 51 are not too efficient in both methods.

When removing the order-m outliers, m fixed at 30, the mean efficiency becomes 0.508 for women not admitted in emergency, 0.319 for men not admitted in emergency, 0.523 for women admitted in emergency and 0.411 for men admitted in emergency. The increase of the efficiencies is higher in the 2 groups of patients admitted in emergency. The proportion of efficient stays were respectively 20.2%, 8.0%, 9.9% and 4.8%. For the whole sample, the mean efficiency is 0.421 (standard deviation 0.270) and the proportion of efficient stays is 7.4%. In comparison, when removing all the stays highlighted as too efficient with the B-W method, the mean efficiency is 0.390 (standard deviation 0.310). In this case, the proportion of efficient stays is 10.9%. Table III shows the percentiles of the distribution of the efficiencies after the exclusion of the too efficient stays. With the order-m frontier, all the distribution shifts to the right. In the B-W method, the proportion of efficient stays is higher with an increase of the median and the mean but with less impact on the lower percentiles of the distribution.

## 5. Discussion

The increase in efficiencies after the exclusion of the too efficient stays is not surprising since the deterministic models are very sensitive to outliers. Therefore, the hospital stays considered too efficient must be detected. The order-m frontier is an interesting tool because it offers the possibility to compare an efficient stay to the stays with at least the same severity level. It is a kind of measure of the plausibility of the minimal value of the total expenses of this efficient stay. An efficiency score larger than 1 indicates a too low level of expenses and if we are too far from 1 (> 1), this low value seems not plausible. Nevertheless, the limit of the application concerns efficient stay with non other stay having at least the same severity level (very low $N_{input}(x_i, y_i)$). In this case, no measure of the plausibility can be realized. For these stays, the B-W method provides a useful complementary information.

The masking effect mentioned by Wilson seems less important in the order-m frontier. By masking effect, we mean the fact that if 2 outliers present close values of the variables analyzed and if only one stay is removed, the remaining stay shadows the outlier characteristic of the removed stay. The B-W method is more sensible to this aspect. The order-m frontier method is more robust to this aspect because the efficiency is a conditional mean of a minimum among a random sample of the stays with at least the same characteristics then the stay analyzed.

When using the nonparametric frontier model, we recommend the use of both methods for detecting outliers before a definitive estimation of the efficiencies. A good way of work would be using first the order-m frontier and second the B-W method when there is no stay with at least the same characteristics. This is a good method to highlight the too efficient stays and consequently it allows a better ranking of hospital stays according to their total expenses taking into account patient characteristics.

## References


Andersen P., Petersen N.C. A procedure for ranking efficient units in data envelopment analysis. Unpublished working paper, Department of Management, Odense University, Denmark, 1989.

Beguin C. Nonparametric frontier model as a tool for exploratory analysis of hospital stays. Methods of Information in Medicine 2001; 40: 241-247.

Cazals C., Florens JP., Simar L. Nonparametric frontier estimation: a robust approach. Journal of Econometrics, 2002; 106: 1-25.

Chilingerian J. Evaluating physician efficiency in hospitals: a multivariate analysis of best practices. European Journal of Operational Research 1995; 80: 548-574.

Deprins D., Simar L., Tulkens H. 'Measuring labor inefficiency in post offices' in M.Marchand, P.Pestieau and H.Tulkens (eds) *The Performance of Public Enterprises: Concepts and Measurements*, North-Holland, Amsterdam, 1984.

Farrell M.J. The measurement of productive efficiency. Journal of the Royal Statistical Society A 1957; 120: 253-281.

Fetter R., Shin Y., Freeman J.L., Averill R.F., Thompson J.D. Case-mix definition by Diagnosis Related Groups. Med Care 1980;18 (2): 53p.

Grosskopf S., Valdmanis V. Measuring hospital performance. A non-parametric approach. Journal of Health Economics 1987; 6: 89-107.

Linna M., Häkkinen U. and Linnakko E. An econometric study of costs of teaching and research in Finnish hospitals. Health Econ. 1998; 7: 291-305.

Lovell C., Walters L., Wood L. Stratified models of education production using DEA and regression analysis. In *Data Envelopment Analysis: Theory, Methods and Applications*. eds A.Charnes, W.Cooper, A.Lewinand L.Seiford, New York: Quorom Books, 1993.

Marazzi A., Paccaud F., Ruffieux C., Beguin C. Fitting the distribution of length of stay by parametric models. Med Care 1998; 36: 915-927.

Puig-Junoy J. Technical efficiency in the clinical management of critically ill patients. Health Econ.1998; 7: 263-277.

Resti A. Evaluating the cost-efficiency of the Italian Banking System: What can be learned from the joint application of parametric and non-parametric techniques. Journal of Banking and Finance 1997; 21: 221-250.

Simar L., Wilson P. Statistical inference in nonparametric frontier models: the state of the art. Journal of Productivity Analysis, 2000; 13, 49-78.

Simar L. Detecting outliers in frontier models. Discussion paper 0146, 2001, Institut de Statistique, Louvain-la-Neuve, Belgium.

Wilson P. Detecting outliers in deterministic nonparametric frontier models with multiple outputs. Journal of Business and Economic Statistics 1993; 11: 319-323.

Wilson P. Detecting influential observations in data envelopment analysis. The Journal of Productivity Analysis 1995; 6: 27-45.

Table I. Too efficient stays among men not admitted in emergency with FDH = 1

| Stays | $N_{input}(x_i, y_i)$ | M = 10 | M = 20 | M = 30 | B-W Method |
|---|---|---|---|---|---|
| 1 | 7 | 1.0007 | 1.0050 | 1 | 1.0339 |
| 2 | 4 | 1 | 1 | 1 | 1.0970 |
| 3 | 1 | 1 | 1 | 1 | 1.0131 |
| 4 | 16 | 1.1569 | 1.0438 | 1.0209 | 1.0139 |
| 5 | 5 | 1.5618 | 1 | 1 | 1.4077 |
| 6 | 2 | 1 | 1 | 1 | 1.7629 |
| 7 | 2 | 1 | 1 | 1 | 1.0311 |
| 8 | 5 | 1 | 1 | 1 | 1.3873 |
| 9 | 4 | 1.0406 | 1 | 1 | 1.0728 |
| 10 | 3 | 1 | 1 | 1 | 2.0955 |
| 11 | 7 | 1.0189 | 1.0379 | 1 | 1.1462 |
| 12 | 40 | 13.863 | 7.0451 | 8.1082 | 2.0787 |
| 13 | 12 | 1.4009 | 1.1420 | 1 | 1.2686 |
| 14 | 1 | 1 | 1 | 1 | 1.3080 |
| 15 | 60 | 3.4898 | 2.9444 | 2.9187 | 1 |
| 16 | 9 | 1.2265 | 1.0185 | 1.0440 | 1 |
| 17 | 41 | 1.3413 | 1.2970 | 1.2185 | 1 |
| 18 | 10 | 1.1397 | 1.0324 | 1.0324 | 1 |
| 19 | 19 | 1.0128 | 1.0230 | 1 | 1 |
| 20 | 10 | 1.0973 | 1.0312 | 1 | 1 |
| 21 | 5 | 1.0113 | 1 | 1 | 1 |
| 22 | 8 | 1.0138 | 1 | 1.0018 | 1 |
| 23 | 2 | 1 | 1 | 1 | 1 |
| 24 | 2 | 1 | 1 | 1 | 1 |
| 25 | 1 | 1 | 1 | 1 | 1 |
| 26 | 1 | 1 | 1 | 1 | 1 |
| 27 | 1 | 1 | 1 | 1 | 1 |
| 28 | 1 | 1 | 1 | 1 | 1 |

Legend of table I

$N_{input}(x_i, y_i)$ = the number of stays with a severity level greater or equal to the efficient stay analyzed.

M = 10, 15, 20, 25, 30: value of m chosen for the order-m frontier.

Method B-W = adaptation by Beguin of the Wilson's method.

Table II. Distribution of the too efficient stays in both methods

| | N | > 1 order-m<br>> 1 method B-W | = 1 order-m<br>> 1 method B-W | > 1 order-m<br>= 1 method B-W | = 1 order-m<br>= 1 method B-W |
|---|---|---|---|---|---|
| Women,<br>No emergency | 22 | 1 | 7 | 3 | 11 |
| Men,<br>No emergency | 28 | 2 | 12 | 5 | 9 |
| Women,<br>Emergency | 29 | 5 | 8 | 4 | 12 |
| Men,<br>Emergency | 50 | 6 | 10 | 15 | 19 |
| Total | 129 | 14 | 37 | 27 | 51 |

Legend of table II

N = the number of efficient stays with the FDH estimator is applied to all the data

> 1 order-m = efficiency greater than 1 when using the order-m frontier with m = 30

> 1 method B-W = efficiency greater than 1 when using the method B-W

Table III Distribution of the efficiencies after exclusion of the too efficient stays

| | All stays | Order-m frontier * | B-W Method |
|---|---|---|---|
| Min | 0.0029 | 0.0167 | 0.0029 |
| P1 | 0.0056 | 0.0374 | 0.0054 |
| P5 | 0.0109 | 0.0768 | 0.0109 |
| P10 | 0.0150 | 0.1042 | 0.0150 |
| P25 | 0.0964 | 0.2295 | 0.1278 |
| P50 | 0.255 | 0.3547 | 0.3237 |
| P75 | 0.4959 | 0.5657 | 0.5627 |
| P90 | 0.871 | 0.8966 | 1 |
| P95 | 1 | 1 | 1 |
| P99 | 1 | 1 | 1 |
| Max | 1 | 1 | 1 |

* when m = 30

Legend of figure 1

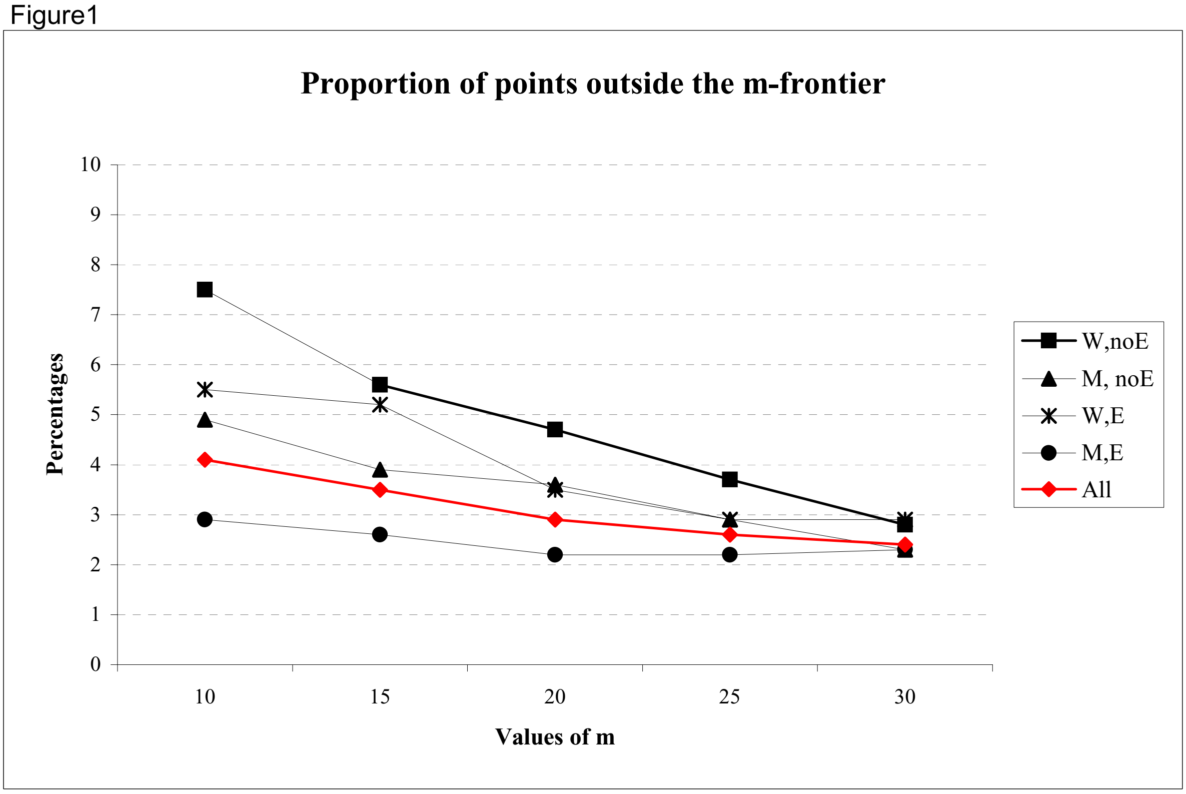
W, noE = women, not admitted in emergency (n = 107)

M, noE = men, not admitted in emergency (n = 306)

W, E = women admitted in emergency (n = 311)

M, E = men admitted in emergency (n = 919).

Figure1


Proportion of points outside the m-frontier
10
9
8
7
6
5
4
3
2
1
0
Percentages
10
15
20
25
30
Values of m
W,noE
M, noE
W,E
M,E
All