UNIVERSITA' DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA GESTIONALE
QUADERNI DEL DIPARTIMENTO[†]

Department of Economics and Technology Management

Working Paper

n. 06 – 2008

The impact of upcoding, cream skimming and readmissions on hospitals' efficiency: the case of Lombardy

by

Paolo Berta, Giuditta Callea, Gianmaria Martini, Giorgio Vittadini

[†] Il Dipartimento ottempera agli obblighi previsti dall'art. 1 del D.L.L. 31.8.1945, n. 660 e successive modificazioni.

The impact of Upcoding, Cream Skimming and Readmissions on hospitals' efficiency: the case of Lombardy

Paolo Berta[†], Giuditta Callea[§,*], Gianmaria Martini[§] and Giorgio Vittadini[+]

August 2008

[†] *CRISP, Italy,* [§] *University of Bergamo, Italy,* [+] *University of Milan Bicocca, Italy*

Abstract

This paper investigates the efficiency of Italian hospitals during the period 1998–2007 and consider the impact of some DRG distortions like uocoding, cream skimming and readmission. We apply both a stochastic frontier approach and a multilevel model. We show that readmissions are the most relevant distortion. Moreover, cream skimming has a negative impact for public and private hospitals and upcoding has a clear positive but little impact only for public hospitals. These results imply that the reaction of the different hospital types to the public policy aimed to reduced the DRG distortions is different. Private hospitals tend to reduce the DRGs while the public hospitals increase the readmissions. Last, private hospitals are less efficient than public and not–for–profit ones, once we take the DRG distortions (and other covariates) into account, both for the SF approach and the ML model.

JEL classification: I11, I18, L33
Keywords: DRG distortions, hospital efficiency, ownership forms, stochastic frontier, multi level.

[*]Correspondence to: G. Callea, University of Bergamo, Viale Marconi 5, 24044 DALMINE (BG), Italy, email: giuditta.callea@unibg.it. The authors are grateful to Luca Merlino and Carlo Zocchetti for the un valuable support in building the dataset and assessing the research project. Callea and Martini acknowledge financial aid support from the Italian Ministry of Education (MIUR) and the University of Bergamo. Giorgio Vittadini acknowledges financial support from the Italian Ministry of University (MUR) and Lombardia Informatica.

1 Introduction

In many industrialized countries the DRG system is a pillar of the health sector organization. There are two main reasons explaining its introduction: First, to guarantee (as the previous cost–based ex–post payment system) an easy and equal access to the health services to all citizens; Second, to improve the sector's efficiency, which was judged as quite low.[1] The latter is obtained, under the DRG system, through two factors: (1) the standardization of the various medical treatments and (2) a pre–determined payment system for each standardized treatment.

The DRG system, whose effects (and benefits) have been investigated by several contributions,[2] has some potential drawbacks (Barbetta *et al.* (2007), p. 82): (1) Early discharges. Since the payment is linked to the medical treatment and not to the lenght of the hospidalization, the management has the incentive to shrink the treatment period. (2) The readmission practice, so that the hospital receive for the same medical treatment more than one reimbursement. (3) The upcoding practice, which consists in registering the patient in more severe DRGs in order to receive higher reimbursements. (4) The well known (and highly debated) cream skimming strategy, i.e. the incentive to provide only the more lucrative health services. These practices are all under the hospital management's control, and they may have a substantial impact on the hospital efficiency. For instance, the early discharges strategy may increase one of the hospital output (i.e. the number of discharges),

[1]As stated by Barbetta *et al.* (2007), many contributions, especially those investigating the US market, point out that a reimbursement system based on incurred costs does not provide incentives to both cost containment and prive competition among hospitals.

[2]The evidence on the effects of the introduction of the DRG system in the US is mixed: Coulam and Gaumer (1991) show that it produced a reduction in the number of admissions and in the average lenght of stay, as expected. Dafny (2005) and Silverman and Skinner (2004) show that the introduction of the DRG system had a relevant impact on the case–mix of patient treated, with opportunistic behaviour exerted by private hospitals.

but the latter is misleading. The aim of our contribution is to estimate the magnitude of the above DRG distortions and to compute their impact on the hospital efficiency. The latter is crucial, since the above distortions may alter (i.e. artificially increasing) the output produced by the hospitals, and they are in direct control of the hospital's management.

To achieve this goal we investigate a sample of 133 Italian hospitals during the period 1998–2007. The sample covers the entire population of active hospitals during all the period in Lombardy, the most populated and rich Italian region.[3] The analysis is limited to an Italian region since national data concerning outputs, inputs and, above all, patients (an essential information to compute the proxies for the previous mentioned DRG distortions) at the hospital level are not public in Italy. Our access to the dataset of the Lombardy region is an exceptional opportunity for the health sector research in Italy, both for the magnitude and the quality of the information available and for the possibility to investigate some managerial practices, such as the distortions mentioned before.

The hospital efficiency is usually measured in terms of technical efficiency, i.e. the hospital's ability of producing the maximum feasible amount of output for a given level of input. The efficiency is obtained by estimating the distance between each hospital and a best practice. Such a benchmark can be given by a "frontier", e.g. a production or a cost function, or by a representative hospital, e.g. the average output–input relationship observed within the sample, taking random effects into account. The distance of each hospital

[3]The Italian health care system is governed by two levels of public authorities: the national government, which states the main guidelines of the health sector, and the regions, that control the reimbursement system and the organization of the sector within the region (e.g. each region may open the health sector also to private hospitals). The country is split in 20 regions: Lombardy is the most important one, looking both at the population (9,5 million, equal to 16% of the total), at the Italian GDP (25% of the total) and at the per capita income (the highest in Italy).

from the frontier can only be positive or zero (i.e. the hospital is fully efficient being exactly on the frontier); the distance from the representative hospital can be either negative or positive (i.e. the hospital may under–perform or over–perform the representative hospital). We evaluate these different distances using two methodological approaches: the stochastic frontier model and the multilevel model.[4]

In accordance with many contributions, the introduction of the DRG system in our sample has produced some well known outcomes: (1) a reduction in the number of admissions, (2) a reduction on the average length of stay and (3) an increase in the case–mix index. However our main finding is that all the four DRG distortions considered have a significant and positive impact on the hospitals' output, both under the stochastic frontier approach and the multilevel model. Hence the effective hospitals' productivity is lower than the observed one, and managers tend to utilize the available inputs to exploit some opportunistic behaviors allowed by the DRG system itself. The policy implication is that the regulator should take these effects into account when design the reimbursement system.

Moreover, we show that upcoding and cream skimming have a negative impact on the hospital's output, hence decreasing their overall efficiency; Second, readmissions have instead a positive impact, but this is due to a distortion not to an efficiency effect. Third, private hospitals are less efficient than public and not–for–profit ones, once we take the DRG distortions into account. Last, the regional system shows a trend of increasing efficiency during the period 1998–2007; this might be due to an effective health policy implemented by the regulator.

Related work. Our work is linked with several contributions that have investigated the impact of different DRG distortions on the hospitals' behaviour. France *et al.* (2005) show that in Italy the ownership has an impact

[4]A critical discussion about the two approaches is provided in Section 2.

on the average length of stay, because it is lower—after the introduction of the DRG system—in private ones. Silverman and Skinner (2004) are the closest contribution to our paper, since they tried to estimate whether the hospital's ownership has an impact on one of the four DRG distortions considered here, i.e. upcoding. They analyse only 4 DRGs (of which one more complicated than the others) and find that private hospitals do more upcoding than private ones. One paper, differently from them, consider all DRGs and 4 distortions.

The paper is organized as follows: in Section 2 we present the stochastic frontier model and the multilevel model, we show the proxies adopted to compute the DRG distortions and we supply some empirical evidence supporting them. The dataset is reported in Section 3, while Section 4 presents the results. The main conclusions and policy implications of the paper are reported in Section 5, which ends up our contribution.

2 Methodology

In this Section we present first the two statistical approaches adopted to compare each hospital with a best practice, then the proxies designed to compute the four DRG distortions and, last, the model on hospital's efficiency that we are going to estimate.

2.1 Estimation techniques

Among the different techniques provided by the literature to compare each hospital with a best practice we focus on two parametric approaches: Stochastic Frontiers (SF) and MultiLevel models (ML). Concerning SF, a fixed effect stochastic frontier model formulation is the following one:[5]

[5]See Greene (2005b) p.16. The same author (1997, 2005a, 2005b) proposes different fixed and random effects stochastic frontier models.

$$y_{it} = \alpha + \beta x_{it} + u_{it} + \epsilon_{it} \tag{1}$$

where y_{it} is the output of firm i in period t, α a constant, β a non random vector of parameters, x_{it} a non random vector of input adopted by firm i in period t, βx_{it} a function of inputs and of a time trend, ϵ_{it} a random vector of residuals, u_{it} a random vector of inefficiency, interpretable as the percentage deviation of observed performance from the firm's own frontier y^*_{it}, where:

$$y^*_{it} = \alpha + \beta x_{it} + \epsilon_{it} \tag{2}$$

and

$$u_{it} = y^*_{it} - y_{it} \tag{3}$$

The model assumes that (1) $[x_{it}, u_{it}, \epsilon_{it}]$ are all mutually uncorrelated; (2) u_{it} and ϵ_{it} have half normal and normal distributions, respectively; (3) u_{it} is not necessarily time invariant.

A ML model specification is instead the following one:

$$log(y_{it}) = \alpha + \beta x_{it} + u_t + \epsilon_{it} \tag{4}$$

where u_t is a random parameter describing how the hospital's average efficiency is affected by variations (around it) due to period t. The ML approach takes fully into account, differently from the SF approach, that in some cases as in health context (Leyland and Goldstein (2001)), the data may have a hierarchical structure, where, for instance, the behaviour of an hospital (which can be regarded as a low level, say level 2) is influenced by each year (the high level, say level 1). Under these circumstances the input/output relation at level 1 is nested in the same relation at level 2. As stated by Steele (2008), "longitudinal data are one example of a hierarchical two level structure" usually based on "repeated observations over time (level 1) nested within hospitals' observations (level 2)" (Steele (2008), p. 1).

Many studies on hospitals efficiency and relative effectiveness suggest that hierarchical models, and particularly ML models, offer better solutions for studying relationships between response variables and contextual variables in complex hierarchical structures (Thomas *et al.* (1994), Goldstein (1995), Epstein (1995), Leyland (1995), Normand *et al.* (1995), Morris and Christiansen (1996), Schneider and Epstein (1996), Goldstein and Spiegelhalter (1996), Rice and Leyland (1996), Leyland and Boddy (1998), Leyland and Goldstein (2001), Marshall and Spiegelhalter (2001), Grassetti *et al.* (2005), Vittadini and Minotti (2005)).

Moreover, when non hierarchical regression techniques treat the lower level units as the units of analysis, the variation among the lower units within the higher level ones is ignored. Consequently estimates of standard errors of means and regression weights are biased and the actual Type I or Type II error rates can be quite different from the nominal ones (Normand *et al.* (1995)).[6]

[6] We can also point out other reasons suggesting to use ML models in health efficiency investigations. First, the assumption of multinormality for random disturbances and random efficiency parameters is often too restrictive, as in case of qualitative or mixed outputs or no prior information on random efficiency parameters distributions (Longford and Lewis (1998), Marshall and Spiegelhalter (2001)). ML models, utilizing bayesian inference, can be instead estimated under several prior distributions for random efficiency parameters and different distribution hypotheses for random disturbances (Leyland and Boddy (1998), Verbeke and Molemberghs (2000), Leyland and Goldstein (2001), Pinheiro and Bates (2002), Vittadini and Minotti (2005)). Second, ML models overcome "small sample problems by appropriately pooling information across institutions, introducing some bias or shrinkage, and providing a statistical framework that allows to quantify and explain the variability in outcomes through the investigation of institutional level covariates." (Marshall and Spiegelhalter (2001, p. 128). Third, the relationships between outputs and inputs can be nonlinear. ML models can be proposed in a non linear version (Verbeke and

2.2 The proxies for the DRG distortions

The literature that investigates the effects of PPS underlines some possible distortions due to this reimbursement system (e.g. Coulam and Gaumer (1991)) but it fails to propose some quantitative indices to measure them (Silverman and Skinner (2004) is a notably exception for upcoding). This paper is an attempt to fill this gap.

As mentioned in the Introduction, we consider three DRG distortions: (1) Upcoding (henceforth UPCOD), (2) Re–admissions (READM) and (3) Cream skimming (CRSK). The fourth distortion we wanted to consider, i.e. the early discharges, is rather difficult to treat in a satisfactory way. In fact it is necessary to distinguish an early discharge due to an efficiency effect (i.e. the hospital discharges earlier because it is well managed) or to the treatment of simpler cases (i.e. the hospital discharges earlier because its patients require less treatments) from the distortion (the hospital discharges earlier because it aims to increase the number of cases in order to get more reimbursements). Hence it would be necessary to consider, in modelling the early discharges index, efficiency in health treaments and the patients' complexity. Since we are not able, at the moment, to compute these corrections we decide not to consider the early discharges in our investigations.

We have instead designed some proxies to compute the other three distortions. The upcoding distortion is defined as follow:

Molenberghs (2000), Ann Gillingan *et al.* (2001), Leyland and Goldstein (2001), Grassetti *et al.* (2005)). Fourth, several contributions (Longford and Lewis (1988), Goldstein (1995), Leyland (1995), Carlin *et al.* (1995), Goldstein and Spiegelhalter (1996)) show, by comparing studies based on observational data, that ML models obtain better and less distorted results, in comparison with other parametric and non parametric methods, when the available dataset has an observational nature (as with administrative data). Last, the relationship between outputs and inputs can be nonlinear and ML models can be proposed in such a version (Verbeke and Molenberghs (2000), Ann Gillingan *et al.* (2001), Leyland and Goldstein (2001), Grassetti *et al.* (2005)).

$$UPCOD_{it} = \frac{1}{CCI_{it}} \times \frac{COMPL_{it}}{COMPL_t} \tag{5}$$

where CCI_{it} represents the Charlson Comorbidity Index of hospital i at period t[7], $COMPL_{it}$ is the percentage of DRGs with complications in hospital i at period t and $COMPL_t$ is the percentage of DRGs with complications in all the hospitals considered. The ratio between $COMPL_{it}$ and $COMPL_t$ is corrected by CCI_{it}) to disentangle a reputation effect (the hospital may treat more complicated cases, in comparison with the regional average, because it has a better reputation than others) from the distortion we want to consider.[8]

The index of cream skimming is given by the following expression:

$$CRSK_{it} = \begin{cases} 1 \quad \text{if} \quad \frac{NDRG_{it}}{NWARD_{it}} \geq \left(\frac{NDRG_t}{NWARD_t}\right)^{90} \\ 2 \quad \text{if} \quad \left(\frac{NDRG_t}{NWARD_t}\right)^{10} < \frac{NDRG_{it}}{NWARD_{it}} < \left(\frac{NDRG_t}{NWARD_t}\right)^{90} \\ 3 \quad \text{if} \quad \frac{NDRG_{it}}{NWARD_{it}} \leq \left(\frac{NDRG_t}{NWARD_t}\right)^{10} \end{cases} \tag{6}$$

where $NDRG_{it}$ is the total number of DRG treated in hospital i at time t, $NWARD_{it}$ the total number of wards in hospital i at time t and $\left(\frac{NDRG_t}{NWARD_t}\right)^s$ is the s^{th} percentile of the ratio between these two indicators in the region in period t (with $s = \{10, 90\}$). The lower $CRSK$ the less cream skimming is observed in the hospital. The ratio between the number of DRG and wards takes into account the relation between the breadth of the hospital's inpatient activity and its size. The underlying idea is that the lower is the number of DRG per ward in a hospital the more the hospital is specialized. However

[7] In medicine comorbidity describes the presence in the patient of other diseases in addition to the primary one. The Charlson Comorbidity Index is the most widely accepted method to quantify it.

[8] Silverman and Skinner (2004) measure the upcoding only for pneumonia without taking not account the CCI, and so they may mix together the reputation effect and the opportunistic behavior.

the latter may be due both to health service concentration and to a selection of the more lucrative activities. To distinguish these effects we compare the hospital's number of DRG per ward with the regional distribution of the same ratio. Hence we assign an higher degree of cream skimming to those hospitals which have a very low ratio, i.e. less than the 10^{th} regional percentile; hospitals with the ratio higher than the 90^{th} regional percentile have instead a low degree of cream skimming.

The proxy for the strategy of readmission is given by:

$$READM_{it} = \frac{y_{it}^{45}}{y_{it}} \tag{7}$$

where y_{it}^{45} represents the total number of readmissions for the same MDC and within 45 days from the date of discharge, while y_{it} is the total number of admissions in hospital i at time t.[9] Descriptive statistics about these three distortions are provided in Section 4.

2.3 The model to estimate

Given the above proxies for the DRG distorsions, the model we estimate to compute the hospital's efficiency is the following one:

$$\begin{aligned} log(y_{it}^{*}) = \alpha &+ \beta_1 log(BEDS_{it}) + \beta_2 log(PHYS_{it}) + \beta_3 log(NURS_{it}) + \\ &+ \beta_4 log(NONSAN_{it}) + \gamma_1 UPCOD_{it} + \gamma_2 CRSK_{it} + \\ &+ \gamma_3 READM_{it} + \delta_1 OWN_{it} + \delta_2 EMERGE_{it} + \\ &+ \delta_3 MONO_{it} + \delta_4 TEACH_{it} + \epsilon_{it} \end{aligned} \tag{8}$$

where the βs are the estimated coefficients for the input variables, the γs for the DRG distortions, the δs for the hospital's characteristics, and ϵ are the errors, which are treated differently in the SF and ML models. The dependent

[9]The regional reimbursement system bears a reduction in the unit reimbursement in case of a readmission within the indicated time spell.

variable $log(y^*_{it})$ is the number of discharges adjusted for case–mix, with the number of discharges given by the following expression:

$$y_{it} = y^{IN}_{it} \times \left(1 + \frac{r^{DC}_{it} + r^{OUT}_{it}}{r^{IN}_{it}}\right)$$

with y_{it} being the total number of discharges, y^{IN}_{it} the total number of inpatient discharges, r^{DC}_{it} the day–care revenues, r^{OUT}_{it} the outpatient revenues and r^{IN}_{it} the inpatient revenues. The input variables regard beds, physicians ($PHYS$)[10], nurses ($NURS$) and non–sanitary personnel ($NONSAN$). The hospital's characteristics are dummy variables for ownership ($OWN = 0$ if the hospital is public, 1 if it is a not–for–profit organization and 2 if it is private), the presence of an emergency department ($EMERGE = 1$ if present), monospecialistic hospitals ($MONO = 1$ if the hospital is running only one department) and the teaching status ($TEACH = 1$ if the hospital has also a teaching activity).

3 The dataset

The data set used in this contribution is composed of information from the population of all the hospitals operating in the Lombardy Italian region during the period 1998–2007. We need data on inputs, such as the labor force separated according to different tasks, and as some proxy for the hospital capital stock, which is given by the number of beds. Moreover, for each hospital we need information to compute the output, which is given by the total number of discharges adjusted for the case–mix. Table 1 shows some descriptive statistics concerning inputs and outputs in the 133 hospitals that compose our dataset. The total number of discharges is reduced during the period, as well as the average length of stay; moreover, we observe an increase in the case–mix index. This evidence is conformed with the insights reported

[10] All labour variables are computed as employees full time equivalent.

in the literature concerning some general effects of DRG system. The total number of discharges adjusted for the case–mix is instead increased over the period. When we consider the inputs, the total number of beds is decreased over the period, signaling that the system was running in over–capacity at the start of the period. The workforce is instead increased in all the different categories, with the exception of the nurses.

	1998				2007			
	Mean	St Dev	Min	Max	Mean	St Dev	Min	Max
Case–Mix Adjusted Discharges	13,357	15,016	567	94,350	17,588	16,744	507	87,249
Discharges	10,872	11,830	143	86,647	9,565	9,023	11	48,169
ALOS	7.50	3.18	3.07	34.42	6.41	3.21	0.96	21.82
CMI	0.89	0.24	0.59	2.06	1.08	0.25	0.62	2.11
Beds	316	321	30	2,030	263	241	15	1,318
Physicians	141	159	1	808	157	156	3	778
Nurses	360	397	11	1,990	358	371	15	1,891
Non sanit Personnel	245	297	3	1,718	252	285	5	1,399

Table 1: Descriptive statistics for Italian hospitals, 1998–2007

In 2007, the average Italian hospital has a case–mix adjusted output equal to 17,588 (+32% in comparison with 1998), while the total number of discharges is 9,565 (-12%). The average length of stay is 6.41 days (-14%), and the case–mix index is 1.08 (+21%). Concerning capacity, the average hospital has 263 beds (-17%); the personnel is composed by 157 physicians (+11%), 358 nurses (-1%) and 285 non–sanitary employees (+3%).

4 Results

First we provide some empirical evidence concerning both the magnitude and the dynamics of the four DRG distortions in hospitals with different ownerships and then we presents the results of our econometric investigation. Figure 1 shows the dynamic of the four distortions during the period 1998–2007 in three different hospital's ownership types: public, not–for–profit and private (i.e. for–profit).[11] Each picture displays the yearly average distorsion per ownership type.

It is evident that the behaviour of the three ownership types regarding the distortions is heterogeneous. While the indices for upcoding and cream skimming are higher for private hospitals (even if the difference is much smaller for upcoding), the picture is reversed if we consider readmissions, where not–for–profit and public hospitals present higher values of the indices.

In more details, if we look at upcode, private hospitals show higher levels than public and not–for–profit ones only in 2003, as pointed out by Silverman and Skinner (2004). In the remaining years, instead, their behaviour is similar to that of the other hospitals. It is of interest to underline that we observe an increasing trend in this distortion and a convergence among the different ownership types.

The picture is different when we consider cream skimming: private hospitals have a much higher index for all the period, with an increasing trend. This new insight confirms some expectations among the profession (i.e. that private hospitals do adopt the cream skimming strategy), but it is the first attempt to quantify them. Another interesting evidence is that at the end of the period not–for–profit hospitals make less cream skimming than public

[11] The time spell for the upcoding distortion is reduced to the period 2000–2007 because the region has changed the method to compute the comorbidity index in 2000, and this modification does not allow to compare the statistics for 1998–1999 with the remaining years.

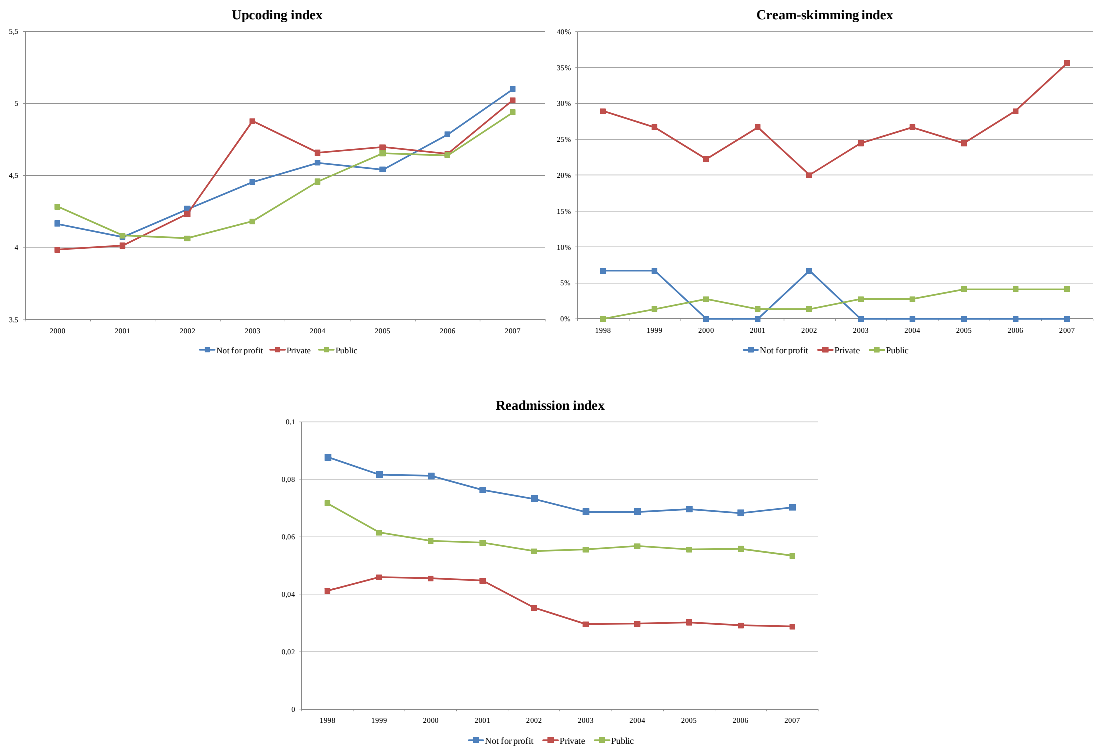

Figure 1: DRG distortions for ownership type

ones, while the opposite was true at the beginning.

Figure 1 also shows the general trend for readmissions. In this case not–for–profit hospitals produce a higher distortion, while the private ones show the lowest level. The not–for–profit index is more than the double of the private ones, while public hospitals tend to make twice the readmissions than private ones. We provide two possible explanations for this evidence: (1) more severe controls on the activity of private hospitals on this two distortions, which may be more easily checked by the region than the previous ones. (2) As mentioned before, we can think two effects having an impact on readmissions: an efficiency effect and the distortion. The efficiency effect may be stronger than the distortion if we consider not–for–profit and public hospitals. That is, not–for–profit and public hospitals make more readmissions because they have better reputation and so an highest share of less healthy patients, which require repeated and more frequent treatments.

Table 2 shows the regression results concerning the efficiency of the hospitals considered. First, the input variables are all significant under both the ML and SF models, with positive coefficients. It is interesting to point out that the two factors having more impact on the hospitals' output are the number of beds and nurses. Second, in the SF model, all the distortions are significant at 1% level, but with different signs. Upcoding and cream skimming have a negative impact on output, while readmissions have a positive impact. These findings imply that if an hospital makes more upcoding it will produce less output. The intuition is the following: this distortion yields an increase in the treatment's complexity. Since the output is computed as number of admissions corrected for the complexity, a higher upcoding should yield an higher complexity adjusted output unless it will reduce the number of admissions. The latter is exactly whan we observe in our investigation. Upcoding reduces the complexity adjusted output because hospitals shrink the number of admissions by treating more complex cases. Moreover, more cream skimming decreases the output since the hospital selects the more lucrative cases and refuses other admissions. Last, an higher readmission index increases the hospital's output, but due to a distortion. Among the three distortions considered readmission has the higher impact.

The analysis of the distortions under the ML model allows to point out an interesting difference from the SF approach, since upcoding is not significant. This is due to the following reason: while the ML model takes into account the nested structure of longitudinal health data, the SF treats time as a covariate and the data structure is not nested. Hence the difference in the significance of this covariate is consistent with its trend shown in Figure 1: the indices for different ownership types tend to assume the same value in the last spell of the time period, so that its significance, in the ML model, is

	ML	st error	SF	st error
Constant	4.27***	0.07	4.53***	0.07
BEDS	0.34***	0.02	0.55***	0.01
PHYS	0.13***	0.01	0.05***	0.01
NURS	0.50***	0.02	0.34***	0.01
NONSAN	0.08***	0.02	0.11***	0.01
UPCOD	0.003	0.002	−0.005***	0.001
CRSK	−0.11***	0.01	−0.08***	0.01
READM	0.97***	0.17	0.33***	0.09
OWN	−0.23***	0.01	−0.21***	0.02
EMERGE	−0.09***	0.02	−0.30***	0.02
T			0.04***	0.001
No. obs	1330	1330	1330	1330
Log–L			609.20	
-2 Log resid L	-119			
Sig Lev	***=1%	**=5%		

Table 2: Estimates for hospital efficiency

likely to be reduced. On the contrary in the SF model, each hospital has only one inefficiency value estimated for all the period (i.e. u_i), which is influence by the overall trend. Therefore since in the first spell of the period the values are rather different, upcoding is likely to have an higher significance under the SF approach. This highlights the usefulness of adopting ML models to complete the results obtained by means of the SF approach when there are longitudinal nested data.

The impact of the other two distortions under the ML model is similar to that obtained with the SF, i.e. cream skimming has a negative significant impact on output and readmission a positive (stronger) effect.

The dummy variable for the presence of an emergency unit has a negative significant effect under the both statistical techniques. The impact of ownership on output is negative and significant under both models, so that a public and a not–for–profit hospital produces less output than private ones. This evidence is consistent with the expectations that private hospitals, in order to maximize their return on investments, tend to treat more cases. The dummy variables MONO and TEACH are not shown because they are not significant. The analysis of the ownership impact has also been split by estimating equation (8) for each type of hospital. Table 3 displays the results if we consider only the not–for–profit hospitals.

	ML	st error	SF	st error
Constant	3.81^{***}	0.20	5.10^{***}	0.32
BEDS	0.10^{*}	0.06	0.36^{***}	0.04
PHYS	0.17^{**}	0.07	0.28^{***}	0.05
NURS	0.46^{***}	0.08	0.16^{*}	0.09
NONSAN	0.29^{***}	0.06	0.07	0.11
UPCOD	0.02^{*}	0.01	0.002	0.02
CRSK	0.24^{***}	0.05	0.07	0.06
READM	0.31^{***}	0.17	0.49	0.79
EMERGE	-0.16^{***}	0.04	-0.15	0.16
T			0.03^{***}	0.005
No. obs	150		150	
Log–L			113.40	
-2 Log resid L	-5.2			
Sig Lev	***=1%	**=5%	*=10%	

Table 3: Estimates for hospital efficiency—only NFP

Under the ML model cream skimming and readmissions have both a positive and significant impact on the output produced by the not–for–profit hospitals (larger for the latter), while upcoding has no effect. The positive sign of cream skimming implies that treating the most remunerative cases does not reduce the output for this type of hospitals, possibly for the presence of a reputation effect that we are not able, at the moment, to capture. The distortions have no effect under the SF approach, probably because it does not take into account the nested structure of longitudinal data the available observations are limited (there only 15 not–for–profit hospitals). Table 4 reports the regression estimates for the public hospitals. This subset of hospital shows that the differences within the results between SF and ML are more relevant with respect to upcoding. Indeed it has a positive significant impact under ML while it has a negative significant effect under SF. The other two distortions have the same effect for both models, i.e. cream skimming has a negative impact, while the effect of readmissions is positive.

	ML	st error	SF	st error
Constant	4.10^{***}	0.10	4.01^{***}	0.13
BEDS	0.61^{***}	0.03	0.73^{***}	0.02
PHYS	0.22^{***}	0.04	0.15^{***}	0.04
NURS	0.23^{***}	0.05	0.20^{***}	0.03
NONSAN	-0.05^{*}	0.03	-0.05	0.04
UPCOD	0.01^{***}	0.003	-0.004^{***}	0.001
CRSK	-0.11^{***}	0.02	-0.06^{***}	0.02
READM	0.90^{***}	0.26	0.34^{**}	0.17
EMERGE	-0.002	0.03	-0.09^{*}	0.05
T			0.04^{***}	0.001
No. obs	730		730	
Log–L			502.04	
-2 Log resid L	509			
Sig Lev	***=1%	**=5%	*=10%	

Table 4: Estimates for hospital efficiency—only public

Last, Table 5 shows the regression estimates when we consider only the private hospitals. All the distortions are significant under the SF approach, meaning that private hospitals reduces their output because of upcoding and cream skimming, and increase it for the readmission strategy. However, differently from the SF approach for the same methodological reasons above mentioned, under the ML model the only significant distortion is cream skimming, with a negative sign.

	ML	st error	SF	st error
Constant	4.46***	0.16	5.80***	0.23
BEDS	0.16**	0.05	0.23***	0.04
PHYS	0.08***	0.02	0.002	0.02
NURS	0.69***	0.04	0.40***	0.03
NONSAN	0.08**	0.03	0.14***	0.02
UPCOD	0.003	0.003	−0.001***	0.003
CRSK	−0.12***	0.03	−0.15***	0.02
READM	0.82*	0.31	0.90***	0.19
EMERGE	0.004	0.03	−0.14*	0.07
T			0.05***	0.003
No. obs	450		450	
Log–L			149.91	
-2 Log resid L	95.2			
Sig Lev	***=1%	**=5%	*=10%	

Table 5: Estimates for hospital efficiency—only private

Having considered all the three types of hospitals, we can now draw some considerations about their differences concerning the distortions (looking at the ML model). First, not–for–profit hospitals do not make "negative" cream skimming (the estimated coefficient is positive, i.e. it increases the output), while private hospitals have the higher level of this distortion. Second, public hospitals make more readmissions than not–for–profit ones, while this distortion is not significant for private hospitals. Third, if we consider upcoding, it has a significant positive impact only for public hospital, even its magnitude is very low.

Figure 2 shows the confidence intervals of the efficiency parameters of all the hospitals. It points out that during the period the regional system has increased the efficiency; hence we can draw the suggestion that the health policy implemented by the regulator during the period had a positive impact on the hospitals' effort to employ efficiently their inputs, coming from the same efficiency trend of all the different hospital types.

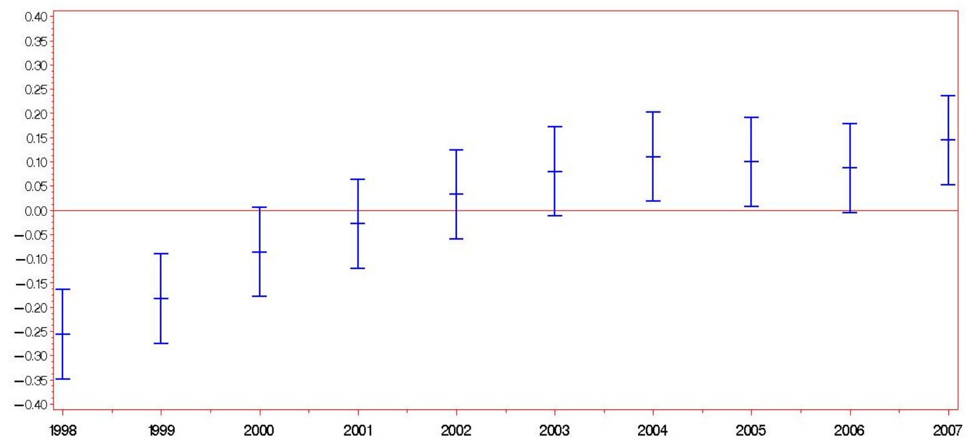

Figure 2: The dynamics of hospitals' efficiency

Table 6 reports the average inefficiencies (i.e. the u_i in the SF approach and the u_{it} under the ML model) according to the different ownership types. Not–for–profit hospitals tend to be more efficient, while the less efficient ones are the private hospitals. This evidence implies that, once we have taken into account of the distortions, private hospitals are less able to optimally use their inputs.

To sum up, we have obtained the following results: First, upcoding and cream skimming have a negative impact on the hospital's output, hence decresing their efficiency; Second, readmissions have instead a positive impact, but this is due to a distortion not to a real supply side effect. Policy makers should take these insight into account when they design their reimbursement

Ownership	ML	SF
NFP	0.10	0.26
Public	0.00	0.27
Private	-0.03	0.42

Table 6: Average inefficiency as function of hospital's ownership

policy. Third, private hospitals are less efficient than public and not–for–profit ones, once we take the DRG distortions into account. Last, the regional system shows a trend of increasing efficiency during the period 1998–2007; this might be due to an effective health policy implemented by the regulation.

5 Conclusions

This paper provides some indices to measure three typical DRG distortions: upcoding, cream skimming and readmissions. Moreover, these distortions are introduced as covariates to investigate the efficiency of Italian hospitals, by applying two different econometric techniques: the SF approach and the ML model. Our main results are the following: First, readmissions are the most relevant distortion, since they significantly increase the hospitals' output. They are particularly adopted by public hospitals. Second, cream skimming has a negative impact for public and private hospitals, and negative for not–for–profit ones. Third, upcoding has a clear positive but little impact only for public hospitals. These results imply that the reaction of the different hospital types to the public policy aimed to reduced the DRG distortions is different. The private hospitals tend to reduce the DRGs, while the public hospitals increase the readmissions. Not–for–profit hospitals, probably for their vocational mission, seem to be less involved in these distortions. Last, private hospitals are less efficient than public and not–for–profit ones, once we take the DRG distortions (and other covariates) into account, both for the SF approach and the ML model.

The paper has also shown, from a methodological point of view, that the ML model, in presence of longitudinal nested health data, complete the information given by the SF approach. Further research on this topic should deepen these methodological insights.

References

- Ann Gilligan, M.,–Kneusel, R.T.,–Hoffmann, R.G.,–Greer, A.L. and A.B. Nattinger, 2002, Persistent Differences in Sociodemographic Determinants of Breast Conserving Treatment Despite Overall Increased Adoption, *Medical Care*, 40, 181–189.

- Barbetta, G.P.,– Turati, G. and A.M. Zago, 2007, Behavioral differences between public and private not–for–profit hospitals in the Italian national health service, *Health economics*, 16, 75–96.

- Carlin, J.B.,–Wolfe, R.,–Brown, C.H. and A. Gelman, 2001, A Case Study on the Choice, Interpretation and Checking of Multilevel Models for Longitudinal Binary Outcomes, *Biostatistics*, 2, 397–416.

- Coulam, R.F.,–Gaumer, G.L., 1991, Medicare's prospective payment system: a critical appraisal, *Health Care Financing Review*, 13, 45–77.

- Dafny, L.S., 2005, How do hospitals respond to price changes? *American Economic Review*, 95, 1525–1547.

- Epstein, A., 1995, Performance Reports on Quality-Prototypes, Problems and Prospects, *N Engl J Med*, 333, 57–61.

- Goldstein, H., 1995, *Multilevel statistical models*, (2nd Ed.), Edward Arnold, Wiley, New York.

- Goldstein, H.,–Spiegelhalter, D.J., 1996, League Table and their Limitations: Statistical Issues in comparisons of Institutional Performances, *J. Royal Statistical Society*, 159, 385–443.

- Grassetti, L.,–Gori E. and S. Minotti, 2005, Multilevel flexible specification of the production function in health economics, *IMA Journal of Management Mathematics*, 16, 383-398.

- Greene, W., 1997, *Frontier Production Functions*, in M. H. Pesaran and P. Schmidt (eds.), *Handbook of Applied Econometrics*, Volume II: Microeconometrics. Oxford: Blackwell Publishers.

- Greene W., 2005a, Fixed and Random Effects in Stochastic Frontier Models, *Journal of Productivity Analysis*, 23, 7-32.

- Greene, W., 2005b, Reconsidering Heterogeneity and Inefficiency: Alternative Estimators for Stochastic Frontier Models, *Journal of Econometrics*, 126, 269–303.

- Leyland, A.H., 1995, Examining the relationship between length of stay and readmission rates for selected diagnoses in Scottish hospitals, *IMA J Math Appl Med Biol*, 12, 175–184.

- Leyland, A.H.,–Boddy, F.A., 1998, League tables and acute myocardial infarction, *Lancet*, 351, 555–558.

- Leyland, A.H.,–Goldstein, H. (eds.), 2001, *Multilevel Models of Health Statistics*, J. Wiley, London.

- Longford, I.,–Lewis, 1998, Outliers in Multivariate Data, *J. Roy. Statistical Society Series A*, 161, 121–160.

- Marshall, E.C.,–Spiegelhalter, D.J., 2001, *Institutional Performance*, in Goldstein H.,–Leyland A.H. (eds.), *Multilevel Modelling of Health Statistics*, Wiley, Chichester, 127–142.

- Morris, C.N.,–Cristiansen, C.L., 1996, *Hierarchical Models for Ranking and for Identifying Extremes*, in Bernardo, J.O.,–Berger Dawid, A.P. (eds.), *AFM In Bayesian Statistic 5*, Oxford University press, 277–297.

- Normand, S.L.T.,–Glickman, M.E. and T.J. Ryan, 1995, *Modelling Mortality Rates for Elderly Heart Attack Patients: Profiling Hospitals*, in Gatsonis, C. *et al.* (eds.), *The Co–Operative Cardiovascular Project. Case Studies in Bayesian Statistics*, Springer Verlag, New York, 435-456.

- Pinhero J.,–Bates, D., 2002, *Mixed–Effects Models in S and S–PLUS Series: Statistics and Computing*, New York, Springer.

- Rice, N,–Leyland, A, 1996, Multilevel Models: Applications to Health Data, *Health Serv Res Policy*, 3, 154–164.

- Schneider, E.C.,–Epstein, A.M., 1996, Influence of Cardiac–Surgery Performance Reports on Referral Practices and Access to Care, *New Eng J Med*, 335, 251–256.

- Silverman, E.,–Skinner, J., 2004, Medicare upcoding and hospital ownership, *Journal of Health Economics*, 23, 369–389.

- Steele F., 2008, Multilevel models for longitudinal data, *J. Royal Statistical Society*, A, 171, 5–19.

- Thomas, N.,–Longford, N.T. and J.E. Rolph, 1994, Empirical Bayes Methods for Estimating Hospital Specific Mortality Rates, *Stat Med*, 13, 889–903.

- Verbeke, G.,–Molenberghs, G., 2000, *Linear Mixed Models for Longitudinal Data*, Springer Verlag, New York.

- Vittadini, G.,–Minotti, S., 2005, A methodology for measuring the relative effectiveness of healthcare services, *IMA Journal of Management Mathematics*, 16 (3), 239–254.