Endogeneity Bias and Prospective Payment Systems

Romain Lesur
French Ministry of Finance and THEMA *
Very preliminary version. Do not cite or distribute.

December 19, 2003

1 Introduction

In the last decades, because of the increase in the health expenditures, many developed countries made reforms of their hospital financing system. Most of the reforms tend to shift from retrospective payments to more prospective ones.
Two types of solution had been proposed by regulators: the first one is known as the "global budget" funding, that consists in giving a budget to the hospital for their overall activities. The well-known second reform was proposed by Medicare and is named the Prospective Payment System (PPS), which relies on a medical classification (the classification in Diagnosis Related Groups). Medicare's PPS was declined in many countries as a way to efficiently finance hospitals regarding their activities.

The main interest of prospective financing is to give incentives to the providers to produce efficiently. However, each of these prospective financing systems faces criticisms :

- The DRG-based regulation does not succeed in lowering the increase in hospital care costs.
- Under the global budget, the budget of the hospital is determined *ex ante* by the regulator. The budget is not dependent on the number of

*Address: Romain Lesur - Direction de la Prevision et de l'Analyse Economique - Teledoc 652 - 139, rue de Bercy - 7572 Paris Cedex 12 - France - phone: (+33)153 18 86 27 - fax: (+33)153 18 36 24 - email: romain.lesur@dp.finances.gouv.fr

patients treated *ex post* and on their severities. In France, the global budget had replaced the previous retrospective reimbursement (a price per day) in 1984 for public hospitals (and later for nonprofit hospitals): the global budgets of the hospitals in 1985 were very close to the budgets of 1984. Year after year, the level of the global budgets were set in function of the historical budgets. In France, the problem of the global budget was the lack of commitment of the regulator to link the budgets to the activities of the hospitals. Twenty years after, the budgets are still correlated with the original budgets while the activities of the hospitals had been transformed. Some of the French hospitals receive rents while others are underfunded: there is a large inequity in the distribution of the whole budget. Theoretically, this kind of regulation can also give incentives to the hospitals to select patients. In France, there is no evidence of such a selection. This risk of selection is lower under the DRG-based regulation since the remuneration for a patient depends on his diagnosis.

A large reform of the French hospital financing system is launched in 2004: a DRG-based regulation is going to progressively replace the global budget regulation. The ultimate goal of this paper is to provide a theoretical framework to make simulations of changes in the hospital regulation (comparison of different DRG-based regulations, comparison of DRG-based with global budget,...). This work is in progress: in this version, this paper develops a theoretical model and the econometric estimations of this model. The simulations of the impact of changes in the regulation are not achieved.

The main purpose of this version is to test the existence of an endogeneity bias in the treatment cost estimation and to present its consequence on the optimal DRG-based regulation. The economic impact of this bias will be estimated when the simulations procedure will be completed.

DRG-based regulations are based on the definitions of prospective payments for each group of the DRG classification. For each of its patients, a hospital receives the "price" of the DRG in which belongs the patient. A key point of this paper is to examine the way that theses prices are set. Usually, the prospective payments are equal to the mean costs of each DRG (observed in identical hospitals): this permits to set the expected profit of each DRG to zero. This regulation is theoretically close to the optimum since it is prospective (it gives the right incentives to efficiency) and the rents are null. However, there are some problems with DRG-based reimbursements:

- Selection: if the hospital has more information on the diseases of the patient than the regulator, there may be selection ((Ma 1994), Chalkley and Malcomson(2002)) to avoid the most expensive patients within

a DRG. The incentives to select patients can be lowered by changing legislation, increasing the payments by DRG or lowering the cost sharing.

- DRG creep: the DRG creep relies on the nature of the information. A patient belongs to a DRG in function of the diagnosis declared by the hospital. However, information on diagnosis is not always verifiable or can be expensive to verify: Silverman & Skinner (2001) analyze upcoding from pneumonia to respiratory infections, that pays $2000 more to the hospital. They also precise that adverse publicity and lawsuits diminished the upcoding index.

- Transfers: reducing the length of stay by transferring patients is a mean to reduce the treatment cost. Solution are to lower the payments when the patients are transferred or to pay in function of the length of stay in case of transfer (Chalkley and Malcomson (2002)).

This paper points out another problem of the DRG-based reimbursement: many DRGs are defined using the treatment used to care for the patient. For instance, different DRGs exist for deliveries and depend on the realization of a cesarean section. McClellan (1997) explains that the use of the treatment choice in the formula of payment relativizes the prospective nature of PPS. The purpose of this paper is that the use of the treatment choice in the definition of the DRGs can imply a bias in the estimation of the opportunity cost of a treatment. When estimating the mean cost of C-sections, the regulator measures the treatment plus a selection bias of the patients: deliveries by C-sections are the most severe cases, and the selection bias arises if the treatment cost is correlated with the severity. This selection bias leads to overestimate the opportunity cost of a C-section and to give incentives to make more C-sections.

This paper is organized as follows: section 2 presents a simple example of an endogeneity bias and its regulatory consequences. Section 3 develops an economic model of treatment choice under different types of regulations. The data used for estimation are presented in the section 4.

2 A simplified approach

The following example presents the main purpose of the paper at hand that is the regulatory consequence of an endogeneity bias. Assume that two treatments are available for the same disease: L and H. Treating with L is costless than treating with H. However, the treatment H is more appropriate than

the treatment L to care for the most severe cases. In the empirical application, the treatment L is a vaginal delivery and the treatment H is a cesarian section.
Consider that the patients are uniformly distributed across three levels of severity denoted by -1, 0 and +1. The treatment costs for each level of severity are given by the following table:

Severity	Treatment L	Treatment H
-1	$\gamma_L - 1$	$\gamma_H - 1$
0	γ_L	γ_H
+1	$\gamma_L + 1$	$\gamma_H + 1$

The opportunity cost of the treatment H is assumed to be constant and equal to $\gamma_H - \gamma_L$ for every level of severity. The preceding assumptions can lead to a bias when estimating the opportunity cost of the treatment H with *ex post* realizations of the treatment costs. In this example, two situations can occur [1]:

- Patients with a severity level -1 are treated by L. The treatment H is used to care for the other patients (i.e. the patients with severity levels 0 and +1). In this case, the mean treatment cost of the treatment L observed *ex post* by the econometrician will be $\overline{c}_L = \gamma_L - 1$; and the mean treatment cost of the treatment H will be $\overline{c}_H = \gamma_H + \frac{1}{2}$.

- Patients with severity levels -1 and 0 are treated by L. The treatment H is used to care for the other patients (i.e. the patients with a severity level +1). In this case, one will observe $\overline{c}_L = \gamma_L - \frac{1}{2}$ and $\overline{c}_H = \gamma_H + 1$.

In both situations, the opportunity cost of the treatment H observed by the econometrician will be $\gamma_H - \gamma_L + \frac{3}{2}$, that overestimates the real opportunity cost. This overestimation comes from the correlation between the treatment choice and the residual, that is an endogeneity bias.
In the real life, regulators often decide to reimburse the mean treatment cost for each Diagnosis Related Group (DRG). In this example, if a regulator reimburses the mean treatment costs for L and for H, then any hospital caring for a patient with H instead of L would make a profit of $\frac{3}{2}$. Surprisingly, this regulatory choice seems to be less efficient that cost reimbursement that leads to a null profit of the hospital for any treatment choice. The main purpose of this paper is that reimbursing the mean treatment cost can distort the treatment choice. The full economic model is detailed in the next section.

[1] The trivial cases in which one of the treatments is not used are excluded of this analysis.

3 The economic model

3.1 Assumptions and definitions

This section develops a model of medical decision making. Two options (L and H) are available in order to treat the same disease. The treatment L is "light" and is the most appropriate one to care for the less severe patients. The treatment H is more expensive and is dedicated to the most severe cases. Any patient is described by the realization of a random vector $(\theta, x_1, \ldots, x_K, s_L, s_H, C_L, C_H)$. The definitions of these random variables are the following:

- θ is the severity of the disease. The random variable θ is assumed to be distributed in $]-\infty; +\infty[$. The density and the repartition function of θ are respectively denoted by f and F.
- $\mathbf{x} = (x_1, \ldots, x_K)'$ is a vector of K diagnosis.
- s_L (resp. s_H) measures the surplus that the treatment L (resp. H) provides to the patient. We extensively use the function $s(\theta)$ defined as:
$$s(\theta) = E(s_H - s_L|\theta).$$
- C_L (resp. C_H) is the casemix of the patient when treated with L (resp. H). The casemix influences the final treatment cost.

The treatment cost of a patient also depends on the hospital factors:

- Exogenous factors explain some of the variation of the treatment costs among hospitals (rural *vs.* urban hospitals, teaching hospitals). All the exogenous factors are captured by the parameter k, that depends on the hospital
- Endogenous factors, such as productivity can also influence the treatments costs. We assume that the treatment costs can be reduced by the manager of the hospital by exerting an effort $e \geqslant 1$. This effort is assumed to be chosen by the manager before the hospital provides any treatment. Exerting an effort e leads to a nonmonetary cost $\psi(e)$ for the manager. The function ψ respects the assumptions: $\psi(1) = 0$, $\psi', \psi'' > 0$ and $r = \psi'(1) > 0$.

The treatment costs c_L and c_H are assumed to have a multiplicative structure:

$$c_L = \frac{k}{e} C_L,\ c_H = \frac{k}{e} C_H$$

The main property of this cost function is that the relative costs are independent of the productivity effort e. In practice, this cost function is implicitly used by the regulators (McClellan [1997]).

The information structure is assumed to be the following:

- The patient does not know the realization of the random vector, he only knows that she needs a treatment.
- The parameter k and the effort e are only known by the manager of the hospital.
- The doctor belonging to the hospital only observes the severity of the patient θ and the diagnosis vector $\mathbf{x}$. The doctor determines the treatment to care for a patient with these elements.
- The regulator, who is also an econometrician, observes *ex-post* for any patient: the treatment used (L or H), the realizations of the diagnosis vector $\mathbf{x}$ and the final treatment cost (c_L or c_H).
- The distribution of the random vector $(\theta, x_1, \ldots, x_K, s_L, s_H, C_L, C_H)$ and the model is common knowledge.

In order to specify the model, we make the following assumptions:

A_Δ **:** There is a nonnegative constant Δ so that:

$$C_H = C_L + \Delta.$$

A_C **:** For any treatment $t \in \{L, H\}$, the expected casemix $E(C_t|\theta)$ is increasing with the severity θ.

A_s **:** The function s, that is the variation of the surplus provided by the treatment H regarding to the treatment L is strictly increasing in θ: $s'(\theta) > 0$. Moreover, there is a value θ^* so that $s(\theta^*) = 0$.

A_θ **:** The severity parameter θ is related to the random vector $\mathbf{x}$: θ conditional on $\mathbf{x}$ follows a normal distribution with mean $\beta'\mathbf{x}$ where $\beta = (\beta_1, \ldots, \beta_K)'$:

$$\theta = \beta'\mathbf{x} + u$$

with $u \sim N(0, \sigma_u^2)$.

The assumption A_Δ means that the cost of H is higher than the cost of L and that the opportunity cost of the treatment H does not depend on the patient. The assumption A_C is the main assumption that we focus on in this paper: A_C means that the casemix of a patient depends on the severity *independently* of the treatment choice. The assumption A_s seems to be realistic: A_s means that H is dedicated to the most severe cases. The assumption A_θ describes the information asymmetry in favor of the doctor: he observes θ and $\mathbf{x}$, but θ is more informative than $\mathbf{x}$. So, the medical decision making only depends on θ. The random variable u represents the information observed by the doctor that cannot be observed by the regulator.

The treatment is denoted by the variable $t \in \{L, H\}$. Wa also use the dummy d defined as follow :

$$d = 1 \text{ if } t = H$$
$$d = 0 \text{ if } t = L$$

From A_θ, the treatment choice only depends on θ. We define the function $t(\theta)$ in order to represent the medical decision making of the doctor. It is assumed that if the doctor is indifferent between both treatments, he chooses the less expensive one, i.e. L.

Defining $p_H = 1 - p_L = Prob\{t = H\} = E(d)$, we also define the expected surplus of the representative patient S:

$$S = Es_L + p_H E(s(\theta)|t(\theta) = H)$$

The casemix is defined by:

$$C = (1 - d)C_L + dC_H$$

It follows that the expected casemix $\overline{C}$ can be written:

$$\overline{C} = EC = p_L\overline{C}_L + p_H\overline{C}_H$$

with $\overline{C}_L = E(C_L|t(\theta) = L)$ et $\overline{C}_H = E(C_H|t(\theta) = H)$. $\overline{C}_L$ et $\overline{C}_H$ represent respectively the mean casemix of the patients treated with L and H. The mean treatment costs are denoted using lowercase: $\overline{c} = \frac{k}{e}\overline{C}$, $\overline{c}_L = \frac{k}{e}\overline{C}_L$ et $\overline{c}_H = \frac{k}{e}\overline{C}_H$. $\overline{c}$ represents the mean treatment cost of the hospital, $\overline{c}_t$ represents the mean cost of the treatment t.

From A_Δ, it comes :

$$C = C_L + \Delta d.$$

Then, the expected casemix can be written:

$$\overline{C} = E(C_L) + p_H\Delta$$

Defining $\Gamma_L = E(C_L)$, it comes:

$$\overline{C} = \Gamma_L + p_H \Delta$$

From its definition, Γ_L represents the mean casemix of the treatment L *if all the patients were treated by* L. We also define $\gamma_L = \frac{k}{e}\Gamma_L$: γ_L is the mean cost of the treatment L *if all the patients were treated by* L. Equivalent notations will be used for $\Gamma_H = \Gamma_L + \Delta$ and $\gamma_H = \frac{k}{e}\Gamma_H$. All these notations are compatible with those of the example presented in the previous section.

3.2 The objective of the hospital

The hospital is assumed to maximize an objective function U that depends on the surplus of the patients, its own profit and the cost of the effort. The number of patients treated in the hospital is denoted by N. The objective function of the hospital is given by:

$$U = q_1 NS + q_2 \Pi - \psi(e),\ q_1, q_2 > 0$$

where Π is the expected profit of the hospital that is linked with the regulatory scheme. With such a function, the objective of the hospital can be treated in generality.

The hospital has two instruments in order to maximize its objective: its treatment choice function $t(\theta)$ and its cost-reducing effort e.

Using A_s and $q_1 > 0$, we first show that the treatment choice function $t(\theta)$ is necessarily increasing in θ. $t(\theta)$ increasing in θ means that if $t(\theta) = H$ then $t(\theta') = H$ for any $\theta' \geqslant \theta$. The key property of any increasing treatment choice function is that one can bijectively define a level $\tilde{\theta}$ such that for any $\theta \leqslant \tilde{\theta}$ then $t(\theta) = L$ else $t(\theta) = H$.
The demonstration that the optimal treatment choice function is necessarily increasing is: let t be a treatment choice function and define $p_H = Prob\{t(\theta) = H\}$. Since A_s and $q_1 > 0$, the increasing treatment choice function defined by $\tilde{\theta} = F^{-1}(p_H)$ is necessarily preferred by the hospital. Then any treatment choice function is dominated by an increasing treatment choice function and we can restrict our analysis to this only set of functions.

Since the doctor makes an increasing treating choice, it comes the following result:

$$\overline{C}_L \leqslant \Gamma_L$$
$$\Gamma_L + \Delta = \Gamma_H \leqslant \overline{C}_H$$

These results come from A_C and A_Δ. It obviously follows:

$$\overline{c}_L \leqslant \gamma_L < \gamma_H \leqslant \overline{c}_H$$

This last result has to be compared with the example of the section 2. The goal of this paper is to test the assumption A_C. If the treatment cost is strictly increasing with the severity, then one gets:

$$\overline{c}_L < \gamma_L < \gamma_H < \overline{c}_H$$

But if the treatment cost is independent of the severity then there is no difference between $\overline{c}_t$ and γ_t:

$$\overline{c}_L = \gamma_L \text{ et } \overline{c}_H = \gamma_H$$

One remarks the key role played by the assumption A_C and the need to verify if the regulation suffers from such an endogeneity bias.

3.3 The hospital's behavior under different regulatory schemes

The goal of this subsection is to derive the optimal behavior of the hospital under the three main regulatory schemes empirically observed.

3.3.1 Cost reimbursement

If the hospital is *ex post* reimbursed for all its costs, then its expected profit is null. The objective of the hospital is summed up by the maximization of $U = q_1 NS - \psi(e)$. The first part of U is the surplus of the patients that only depends on the treatment choice function. The second part of U only depends on the cost-reducing effort. So, the optimal behavior of the hospital is to make the minimal effort $e = 1$ and the treatment choice has to verify:

$$t(\theta) = H \text{ iff } s(\theta) > 0$$
$$t(\theta) = L \text{ iff } s(\theta) \leqslant 0$$

The treatment choice function is increasing and $\tilde{\theta}$ is given by $s(\tilde{\theta}) = 0$, that means $\tilde{\theta} = \theta^*$. The main properties of the cost reimbursement is to maximize the surplus of the patients and to minimize the cost-reducing effort. This situation is obviously not socially optimal: maximizing the surplus of the patients cannot be the social optimum since it does not internalize the treatment costs ; minimizing the cost-reducing effort cannot be socially optimal since it does not internalize the benefit of the effort.

3.3.2 The global budget

The hospital receives a lump funding G in order to finance its activities. This global budget G is defined *ex ante*. Empirically, this kind of regulatory scheme suffers from the fact that the budget is historically based and that after many years this sum is not well related to the number of patients and the type of the treatments provided in the hospital. The profit of the hospital can be written:

$$\Pi \equiv G - N\frac{k}{e}(\Gamma_L + (1 - F(\tilde{\theta}))\Delta)$$

The objective function of the hospital has the following expression:

$$U = q_2 G + N q_1 E(s_L) + N q_1 \int_{\tilde{\theta}}^{+\infty} s(\theta) f(\theta) d\theta - N q_2 \frac{k}{e}(\Gamma_L + (1 - F(\tilde{\theta}))\Delta) - \psi(e)$$

The partial derivatives of U with respect to $\tilde{\theta}$ and e are:

$$\frac{\partial U}{\partial \tilde{\theta}} = N(k q_2 \frac{\Delta}{e} - q_1 s(\tilde{\theta})) f(\tilde{\theta})$$
$$\frac{\partial U}{\partial e} = N k q_2 \frac{\Gamma_L + (1 - F(\tilde{\theta}))\Delta}{e^2} - \psi'(e)$$

The main problem observed in France with the global budget is that some hospitals are rationed while others make rents. This comes from the correlation of the current budget with the past ones. In order to describe the problem of rationing that can occur, we solve the maximization program of the hospital:

$$\max_{\tilde{\theta}, e} \; U$$
$$\text{s.c. } \Pi \geqslant 0 \; (\lambda_1)$$
$$e \geqslant 1 \; (\lambda_2)$$

The first order conditions lead to:

$$s(\tilde{\theta}) = \frac{k}{e}\frac{(q_2 + \lambda_1)}{q_1}\Delta$$
$$\psi'(e) = N(q_2 + \lambda_1)\frac{k}{e}\frac{\Gamma_L + (1 - F(\tilde{\theta}))\Delta}{e} + \lambda_2$$

Under the global budget regulation, the hospital chooses a value $\tilde{\theta}$ greater than θ^*, that was the value found under cost reimbursement. That means that with the global budget, the hospital treats less patients with H than under cost reimbursement. Under the global budget, the hospital internalizes

a share of the treatment costs.
The cost-reducing effort e may be greater than 1, if this is optimal for the hospital. In the remainder of this paper, we will admit that under global budget, the hospital always make an effort greater than 1: we assume that $r < Nq_2k\Gamma_L$.

When the hospital is rationed ($\lambda_1 > 0$), the share of the patients treated with H is decreasing in λ_1, that is intuitive: if the hospital is rationed, it reduces the number of patients treated with H in order to respect its budget constraint. Rationing also increases the cost-reducing effort: the manager of the hospital lower the weight of the budget constraint by exerting a higher effort.

3.3.3 The DRG-based reimbursement

The regulator defines two levels of casemix R_L and R_H that are the basis to reimburse the hospital for having treated patients with L and H. In order to simplify the regulatory mechanism, we assume that the regulator takes into account the cost parameter k and reimburses $\frac{k}{e}R_L$ (resp. $\frac{k}{e}R_H$) for all the patients treated by L (resp. H). Then, the expected profit is given by:

$$\Pi = \frac{k}{e}\left(F(\tilde{\theta})(R_L - \overline{C}_L) + (1 - F(\tilde{\theta}))(R_H - \overline{C}_H)\right)$$

The regulator has to propose a contract (R_L, R_H) that satisfies the participation constraint of the hospital: $\Pi \geqslant 0$.
The formal resolution of the maximization program of the hospital leads to the following expressions for the cost-reducing effort and the treatment choice function :

$$\psi'(e) = Nq_2\frac{k}{e}\frac{\Gamma_L + (1 - F(\tilde{\theta}))\Delta}{e}$$

$$s(\tilde{\theta}) = -\frac{q_2}{q_1}\frac{k}{e}\left[R_H - R_L - \Delta\right]$$

The second first-order condition means that $\tilde{\theta}$ is decreasing in the difference between $R_H - R_L$ and Δ. If $R_H - R_L$ is greater than Δ, then the treatment choice threshold $\tilde{\theta}$ is lower than θ^*: the hospital uses the treatment H more frequently than in the cost-reimbursement regulation. In this case, the surplus of the patients is not maximized: the treatment H is used too frequently.

The theory of regulation shows that because of the shadow cost of public transfers, rents are distortive. In our context, the profit of the hospital can

be set to zero if R_L and R_H are such that:

$$R_L + (1 - F(\tilde{\theta}))(R_H - R_L) = \Gamma_L + (1 - F(\tilde{\theta}))\Delta$$
$$\Leftrightarrow F(\tilde{\theta})R_L + (1 - F(\tilde{\theta}))R_H = F(\tilde{\theta})\overline{C}_L + (1 - F(\tilde{\theta}))\overline{C}_H$$

It exists an infinity of contracts (R_L, R_H) such that $\Pi = 0$. However, we focus our analysis on two of them:

Contract A The contract A is usually used by the regulators in DRG-based reimbursements: they set R_L and R_H equal to the mean casemix: $R_L = \overline{C}_L$ and $R_H = \overline{C}_H$. The main property of this regulatory scheme is that it avoid cross-subsidies among treatments: both treatments are profitable. This contract reimburses the mean treatment cost for each treatment.

Contract B The contract B is defined by $R_L = \Gamma_L > \overline{C}_L$ and $R_H = \Gamma_L + \Delta < \overline{C}_H$. This contract makes cross-subsidies among treatments. For each treatment t, this contact reimburses its mean cost *if all the patients were treated* by t.

The contract A is such that $R_H - R_L > \frac{k}{e}\Delta$ and the contract B is such that $R_H - R_L = \Delta$. Then, the usual DRG-based reimbursement (using the contract A) implies a more frequent use of the treatment H than cost-reimbursement or global budget. The consequence of the usual DRG-based reimbursement on the overall treatment costs is unclear: the prospective payments give incentives to the hospital to produce efficiently but the contract A induces a more frequent use of the treatment H.

The main purpose of this paper that when the usual DRG-based payments were implemented (using the contract A), the threshold $\tilde{\theta}$ has decreased because of the difference between $R_H - R_L$ and Δ. This a new explanation of the observed increase in the casemix index associated with the implementation of the DRG-based regulations. Moreover, this contract does not maximize the patients' surplus: some cases do not need the treatment H but receive it.

The contract B that is described in this paper permit to implement the same threshold $\tilde{\theta}$ as under cost-reimbursement, i.e. θ^*. However, the contract B is based on the use of cross-subsidies among both treatments. The endogeneity bias that we want to test, is not only an econometric problem but is also quite important in term of regulation. If there is an endogeneity bias, the regulator is faced by a tradeoff between avoiding cross-subsidies and implementing an optimal regulation. Using cross-subsidies gives incentives to the hospital to select patients. But avoiding cross-subsidies among treatments implies a higher share of patients treated with H.

4 An application to deliveries

Our model was estimated using data on deliveries. Our dataset contains informations about 83,000 deliveries in years 1999 and 2000. This data come from public and nonprofit French hospitals.
For each delivery, the information we possess are: the treatment cost (using analytic account), the age of the mother, the diagnosis (using the ICD 10), the length of stay and the presence of complications and comorbidities (CC). The observations concern mothers from 15 to 45 years old. Transfers toward or from another hospital were excluded of the analysis to avoid any bias. In the French DRG classification, all these deliveries belong to 4 DRGs: DRG 530 (C-sections with CC), DRG 531 (C-sections without CC), DRG 539 (vaginal deliveries with CC) and DRG 540 (vaginal deliveries without CC). Our observations come from 26 hospitals in 1999, and from 28 hospitals in 2000. Tables 1 and 2 presents some descriptive statistics on the treatment costs.

Diagnosis Related Groups	*Analysis of variance*			Relative mean*	Weights used for regulation**
	Frequency	Mean	Standard Deviation		
C-sections w. CC	2.85%	24,424	14,533	1.853	2.155
C-sections w/o CC	12.93%	19,501	9,602	1.479	1.507
Vaginal deliv. w. CC	12.30%	15,884	9,704	1.205	1.158
Vaginal deliv. w/o CC	71.92%	13,183	5,364	*1.000*	*1.000*
All	*N=39,111*	*14,652*	*7,596*		
		R^2=*0.13*			

* 1.000=DRG 540 **Reference: Agence Technique de l'Information sur l'Hospitalisation

Figure 1: Treatment costs (in current French francs) by DRG in 1999

Diagnosis Related Groups	*Analysis of variance*			Relative mean*	Weights used for regulation**
	Frequency	Mean	Standard Deviation		
C-sections w. CC	2.77%	26,783	20,285	1.952	2.229
C-sections w/o CC	14.44%	19,755	10,980	1.440	1.431
Vaginal deliv. w. CC	12.95%	15,842	10,527	1.155	1.160
Vaginal deliv. w/o CC	69.84%	13,718	5,655	*1.000*	*1.000*
All	*N=43,974*	*15,226*	*8,589*		
		R^2=*0.11*			

* 1.000=DRG 540 **Reference: Agence Technique de l'Information sur l'Hospitalisation

Figure 2: Treatment costs (in current French francs) by DRG in 2000

The DRG classification explains only 11% and 13% of the variance of the treatment costs. The variance within each DRG can be explained by variations in productivity and efficiency of the hospitals (this is the factor $\frac{k}{e}$ in the previous section) and by variations in the casemix of the patients.
These descriptive statistics show that the cost of a C-section without CC was 47.9% higher than the cost of a vaginal delivery without CC in 1999. However, such an indicator can be biased if the hospitals present some differences in their productivity.

The proportion of C-sections in our sample is 15.78% in 1999 and 17.21% in 2000. The proportions of C-sections in France these two years are presented in tables 3 and 4). Table 3 shows that the proportion of C-sections

Diagnosis Related Groups	Our sample	All public and NFP hospitals	For-profit hospitals	All
C-sections w. CC	2.85%	2.45%	1.63%	2.17%
C-sections w/o CC	12.93%	13.36%	15.55%	14.11%
Vaginal deliv. w. CC	12.30%	12.29%	8.22%	10.89%
Vaginal deliv. w/o CC	71.92%	71.90%	74.60%	72.83%
All	39,111	464,997	243,857	708,854
Proportion of C-sections	*15.78%*	*15.81%*	*17.18%*	*16.28%*

Figure 3: The casemix of the French deliveries in 1999

Diagnosis Related Groups	Our sample	All public and NFP hospitals	For-profit hospitals	All
C-sections w. CC	2.77%	2.76%	1.67%	2.38%
C-sections w/o CC	14.44%	13.90%	16.07%	14.66%
Vaginal deliv. w. CC	12.95%	13.37%	8.30%	11.58%
Vaginal deliv. w/o CC	69.84%	69.97%	73.96%	71.38%
All	43,974	504,572	274,056	778,628
Proportion of C-sections	*17.21%*	*16.65%*	*17.75%*	*17.04%*

Figure 4: The casemix of the French deliveries in 2000

in our sample is representative of the patients of the French public and NFP hospitals in 1999. However, it seems that our sample is less representative in 2000. Testing the assumption A_C does not suffer from this problem of representativity. However, the simulations of the economic implications of the different regulatory schemes can be affected by this problem of representativity.

5 Specification of the model

We need to specify more precisely the model to estimate it. The cost of the effort is assumed to be defined as:

$$\psi(e) = \frac{r}{\alpha}(e^{\alpha} - 1)$$

with $\alpha > 1$. This specification verifies all the assumptions on ψ: $\psi(1) = 0$, $\psi'(1) = r$, et $\psi', \psi'' > 0$.

Many casemix functions respect the assumptions A_Δ and A_C. The empirical estimations are based on the following expression of C:

$$C = a + \Delta d + \rho\theta + v$$

where v is a random variable normally distributed with expectation 0 and variance σ_v^2. The covariance between u and v is noted σ_{uv} and v is assumed to be noncorrelated with $\mathbf{x}$.
The assumption A_θ leads to :

$$\Gamma_L = a + \rho E(\theta) = a + \rho E_{\mathbf{x}}(\beta'\mathbf{x})$$

The main problem using this formalization for C is that it theoretically implies negative values of the treatment cost. However, we show that in our empirical estimations the probability that the cost is negative is very low. So, the use of a normal distribution is not a strong assumption.
Let ϕ and Φ be respectively the density and the repartition functions of the normal distribution with expectation 0 and unit variance. The expressions of the expected casemix $\overline{C}_L$ and $\overline{C}_H$ become:

$$\overline{C}_L = a + \rho E_{\mathbf{x}}(\beta'\mathbf{x}|t = L) - (\rho\sigma_u + \frac{\sigma_{uv}}{\sigma_u})E_{\mathbf{x}}\left(\frac{\phi(\frac{\beta'\mathbf{x}-\tilde{\theta}}{\sigma_u})}{1 - \Phi(\frac{\beta'\mathbf{x}-\tilde{\theta}}{\sigma_u})}\middle| t = L\right)$$

$$\overline{C}_H = a + \Delta + \rho E_{\mathbf{x}}(\beta'\mathbf{x}|t = H) + (\rho\sigma_u + \frac{\sigma_{uv}}{\sigma_u})E_{\mathbf{x}}\left(\frac{\phi(\frac{\beta'\mathbf{x}-\tilde{\theta}}{\sigma_u})}{\Phi(\frac{\beta'\mathbf{x}-\tilde{\theta}}{\sigma_u})}\middle| t = H\right)$$

It follows:

$$\begin{aligned}\overline{C}_H - \overline{C}_L =& \Delta + \rho\left[E_{\mathbf{x}}(\beta'\mathbf{x}|t = H) - E_{\mathbf{x}}(\beta'\mathbf{x}|t = L)\right] \\ &+ (\rho\sigma_u + \frac{\sigma_{uv}}{\sigma_u})\left[E_{\mathbf{x}}\left(\frac{\phi(\frac{\beta'\mathbf{x}-\tilde{\theta}}{\sigma_u})}{\Phi(\frac{\beta'\mathbf{x}-\tilde{\theta}}{\sigma_u})}\middle| t = H\right) + E_{\mathbf{x}}\left(\frac{\phi(\frac{\beta'\mathbf{x}-\tilde{\theta}}{\sigma_u})}{1 - \Phi(\frac{\beta'\mathbf{x}-\tilde{\theta}}{\sigma_u})}\middle| t = L\right)\right]\end{aligned}$$

The assumption A_C hold if $\rho > 0$ or $\eta > 0$. The necessary condition to prove that there is no andogeneity bias is $\rho = \eta = 0$. However, if ρ or η is

nonnegative then the conclusion will be that the difference between the mean costs overestimate the opportunity cost of the treatment H.

The surplus function of the patient is specified using the joint recommendations of the WHO, UNESCO and UNFPA: the rate of C-sections in a population shall not be lower than 5%. However, it is extremely difficult to determine an upper bound for this rate, because the rate of C-sections is related to the medical practice and to the will of the patients. So, we use a surplus function that respect the following assumptions:

$$\begin{aligned} &\lim_{\theta \to -\infty} s(\theta) = -\infty \\ &s(\theta^*) = 0 \\ &\lim_{\theta \to \theta_{95}^-} s(\theta) = +\infty \end{aligned}$$

with θ_{95} is the 95^{th} percentile of the distribution of θ, and θ^* is the indifference threshold between a C-section and a vaginal delivery. In our simulations of the economic impact of the regulatory schemes, we present the results for different values of θ^*.

The function surplus used for the estimations is:

$$s(\theta) = \kappa.ln\left(\frac{\theta_{95} - \theta^*}{\theta_{95} - \theta}\right),\ \kappa > 0.$$

The previous properties hold for this function.

6 The estimation methods

Our data come from the observations of H hospitals, the index of the hospital is $h \in \{1, \ldots, H\}$. We observe all the deliveries in each of the hospitals of our sample. The number of deliveries in the hospital h is denoted by N_h. Within the hospital h, the index of the observation is $i \in \{1, \ldots, N_h\}$. The number of observations is denoted by N and we have $N = \sum_h N_h$.

We assume that the data are generated by the economic model. So, the econometric model associated with the economic model is:

$$\begin{aligned} c_{hi} &= \frac{k_h}{e_h}\left(a + \Delta d_{hi} + \rho\theta_{hi} + v_{hi}\right) \\ \theta_{hi} &= \beta' \mathbf{x}_{hi} + u_{hi} \\ \text{avec } &\left\{ \begin{array}{l} d_{hi} = 1 \text{ si } \theta_{hi} > \tilde{\theta}_h \\ d_{hi} = 0 \text{ si } \theta_{hi} \leqslant \tilde{\theta}_h \end{array} \right. \end{aligned}$$

The random vector (u_{hi}, v_{hi}) follows a bivariate normal distribution and is assumed to respect the following conditions:

$$\begin{aligned}
&E(u_{hi}) = E(v_{hi}) = 0 \\
&E(u_{hi}^2) = \sigma_u^2 \\
&E(v_{hi}^2) = \sigma_v^2 \\
&E(u_{hi}v_{hi}) = \sigma_{uv} \\
&E(u_{hi}v_{h'i'}) = E(u_{hi}u_{h'i'}) = E(v_{hi}v_{h'i'}) = 0 \text{ for } (h,i) \neq (h',i')
\end{aligned}$$

The semi-reduced form of this model is:

$$\begin{aligned}
c_{hi} &= \frac{k_h}{e_h}\left(a + \Delta d_{hi} + \rho(\beta' \mathbf{x}_{hi}) + \varepsilon_{hi}\right) \\
d_{hi}^* &= \beta' \mathbf{x}_{hi} - \tilde{\theta}_h + u_{hi}
\end{aligned}$$

with $\varepsilon_{hi} = \rho u_{hi} + v_{hi}$. Thus, $\sigma_{u\varepsilon} = \rho\sigma_u^2 + \sigma_{uv}$ and $\sigma_\varepsilon^2 = \rho^2\sigma_u^2 + \sigma_v^2 + 2\sigma_{uv}$. Moreover, one has:

$$\begin{aligned}
d_{hi} &= 1 \text{ si } d_{hi}^* > 0 \\
d_{hi} &= 0 \text{ si } d_{hi}^* \leqslant 0
\end{aligned}$$

Since the preceding system is triangular (d_{hi} is not present in the definition of d_{hi}^*,), the model can be estimated. However, this system is not identifiable. In the following of this paper, we use the following identifying restrictions: $a = \sigma_u = 1$.

In this semi-reduced model, the random vector $(\varepsilon_{hi}, u_{hi})$ follows a bivariate normal distribution. The coefficient of correlation of ε_{hi} on u_{hi} is denoted by η. So, one has: $\varepsilon_{hi} = \eta\frac{\sigma_\varepsilon}{\sigma_u}u_{hi} + e_{hi}$, where u_{hi} and e_{hi} are independent. Three types of estimation were performed:

- a Heckit two-step estimation (Probit followed by a NLS regression)
- a two-step limited information maximum of likelihood estimation (LIML)
- a full information maximum of likelihood estimation(FIML)

6.1 A two-step estimation: Probit followed by NLS

Following the resolution procedure introduced by Heckman, the conditional expectations of c_{hi} are:

$$E(c_{hi}|d_{hi} = 0, \mathbf{x}_{hi}) = \frac{k_h}{e_h}\left(1 + \rho(\beta' \mathbf{x}_{hi}) - \eta\sigma_\varepsilon \frac{\phi(\beta' \mathbf{x}_{hi} - \tilde{\theta}_h)}{1 - \Phi(\beta' \mathbf{x}_{hi} - \tilde{\theta}_h)}\right)$$

Similarly, one gets:

$$E(c_{hi}|d_{hi}=1,\mathbf{x}_{hi}) = \frac{k_h}{e_h}\left(1+\Delta+\rho(\beta'\mathbf{x}_{hi})+\eta\sigma_\varepsilon\frac{\phi(\beta'\mathbf{x}_{hi}-\tilde{\theta}_h)}{\Phi(\beta'\mathbf{x}_{hi}-\tilde{\theta}_h)}\right)$$

The two steps of the estimations are:

- A probit estimation of the treatment choice is first performed. Then we obtain estimations $\hat{\beta}$ and $\hat{d}^*_{hi}$ of β and d^*_{hi}.
- Defining $Z^1_{hi} = \hat{\beta}'\mathbf{x}_{hi}$ and $Z^2_{hi} = d_{hi}.\frac{\phi(\hat{d}^*_{hi})}{\Phi(\hat{d}^*_{hi})} - (1-d_{hi})\frac{\phi(\hat{d}^*_{hi})}{1-\Phi(\hat{d}^*_{hi})}$, a nonlinear least squares estimation (NLS) is used to estimate the equation:

$$c_{hi} = \frac{k_h}{e_h}\left(1+\Delta d_{hi}+\rho Z^1_{hi}+(\eta\sigma_\varepsilon)Z^2_{hi}\right)+\varepsilon'_{hi}$$

The problem of this method is that the residual of the last equation are heteroscedastic, but the estimations are quite easy using standard statistical softwares. Methods using the likelihood of the model do not suffer from this limit.

6.2 The limited information maximum of likelihood estimation (LIML)

The log-likelyhood of the model is:

$$\ln L = -\frac{N}{2}\ln 2\pi - N\ln(\sigma_\varepsilon) - \sum_h N_h \ln\left(\frac{k_h}{e_h}\right) + \sum_{h,i}\left[\ln\Phi\left(\omega_{hi}\frac{\eta\frac{\varepsilon_{hi}}{\sigma_\varepsilon}+\beta'\mathbf{x}_{hi}-\tilde{\theta}_h}{\sqrt{1-\eta^2}}\right) - \frac{1}{2}\left(\frac{\varepsilon_{hi}}{\sigma_\varepsilon}\right)^2\right]$$

avec $\varepsilon_{hi} = \frac{e_h}{k_h}c_{hi} - (1+\Delta d_{hi}+\rho\beta'\mathbf{x}_{hi})$ et $\omega_{hi} = 2d_{hi}-1$.

Proof: The model can be written:

$$c_{hi} = \frac{k_h}{e_h}(1+\Delta d_{hi}+\rho(\beta'\mathbf{x}_{hi})+\varepsilon_{hi})$$
$$d^*_{hi} = \beta'\mathbf{x}_{hi} - \tilde{\theta}_h + u_{hi}$$

The random vectors $(\varepsilon_{hi}, u_{hi})$ follow bivariate normal distribution with:

$$E(\varepsilon_{hi}) = E(v_{hi}) = 0$$
$$E(\varepsilon^2_{hi}) = \sigma^2_\varepsilon$$
$$E(u^2_{hi}) = 1$$
$$E(\varepsilon_{hi}u_{hi}) = \eta\sigma_\varepsilon$$
$$E(\varepsilon_{hi}u_{h'i'}) = E(\varepsilon_{hi}\varepsilon_{h'i'}) = E(u_{hi}u_{h'i'}) = 0 \text{ pour } (h,i)\neq(h',i')$$

The density of probability of the vector (ε, u) is given by:

$$\frac{1}{\sigma_\varepsilon}\phi_2(\frac{\varepsilon}{\sigma_\varepsilon}, u, \eta) = \frac{1}{2\pi\sigma_\varepsilon\sqrt{1-\eta^2}} e^{-\frac{1}{2}\frac{\frac{\varepsilon^2}{\sigma_\varepsilon^2}+u^2-2\eta\frac{\varepsilon u}{\sigma_\varepsilon}}{1-\eta^2}}$$

From the preceding expression, one founds the following equalities:

$$\int_{-\infty}^{\overline{u}} \frac{1}{\sigma_\varepsilon}\phi_2(\frac{\varepsilon}{\sigma_\varepsilon}, u, \eta)du = \frac{1}{\sigma_\varepsilon}\phi(\frac{\varepsilon}{\sigma_\varepsilon})\Phi(\frac{\overline{u}-\eta\frac{\varepsilon}{\sigma_\varepsilon}}{\sqrt{1-\eta^2}})$$

and

$$\int_{\underline{u}}^{+\infty} \frac{1}{\sigma_\varepsilon}\phi_2(\frac{\varepsilon}{\sigma_\varepsilon}, u, \eta)du = \frac{1}{\sigma_\varepsilon}\phi(\frac{\varepsilon}{\sigma_\varepsilon})\Phi(\frac{\eta\frac{\varepsilon}{\sigma_\varepsilon}-\underline{u}}{\sqrt{1-\eta^2}})$$

The likelihood for the observation (h, i) can be written :

$$\begin{aligned} L =& \frac{1}{\frac{k_h}{e_h}\sigma_\varepsilon}\phi(\frac{\varepsilon_{hi}}{\sigma_\varepsilon}) \left[\Phi\left(\frac{\tilde{\theta}_h - \beta'\mathbf{x}_{hi} - \eta\frac{\varepsilon_{hi}}{\sigma_\varepsilon}}{\sqrt{1-\eta^2}}\right)\right]^{1-d_{hi}} \left[\Phi\left(\frac{\eta\frac{\varepsilon_{hi}}{\sigma_\varepsilon} + \beta'\mathbf{x}_{hi} - \tilde{\theta}_h}{\sqrt{1-\eta^2}}\right)\right]^{d_{hi}} \\ =& \frac{1}{\frac{k_h}{e_h}\sigma_\varepsilon}\phi(\frac{\varepsilon_{hi}}{\sigma_\varepsilon})\Phi\left(\omega_{hi}\frac{\eta\frac{\varepsilon_{hi}}{\sigma_\varepsilon} + \beta'\mathbf{x}_{hi} - \tilde{\theta}_h}{\sqrt{1-\eta^2}}\right) \end{aligned}$$

with $\omega_{hi} = 2d_{hi} - 1$. **QED**

The LIML estimation proceeds in two steps:

- The first step is the same as the previous method: a probit is performed on the treatment choice. Then, one gets the estimations $\hat{\beta}$ and $\hat{d}^*_{hi}$ of β et d^*_{hi}.
- The following likelihood is maximized in $(\frac{k_h}{e_h}, \Delta, \rho, \eta, \sigma_\epsilon)$:

$$\begin{aligned} \ln L =& -\frac{N}{2}\ln 2\pi - N\ln(\sigma_\epsilon) - \sum_h N_h \ln\left(\frac{k_h}{e_h}\right) \\ &+ \sum_{h,i}\left[\ln\Phi\left(\omega_{hi}\frac{\eta\frac{\epsilon_{hi}}{\sigma_\epsilon} + \hat{d}^*_{hi}}{\sqrt{1-\eta^2}}\right) - \frac{1}{2}\left(\frac{\epsilon_{hi}}{\sigma_\epsilon}\right)^2\right] \end{aligned}$$

with $\epsilon_{hi} = \frac{e_h}{k_h}c_{hi} - (1 + \Delta d_{hi} + \rho\hat{\beta}'\mathbf{x}_{hi})$ and $\omega_{hi} = 2d_{hi} - 1$.

The main advantage of the LIML estimation is that the maximization is performed on few parameters and is less cumbersome to estimate.

6.3 The full information maximum of likelihood estimation (FIML)

This last method consist in the maximization of the likelihood on all the parameters of the model: $(\frac{k_h}{e_h}, \Delta, \rho, \eta, \sigma_\varepsilon, \tilde{\theta}_h, \beta)$.

7 The results

The three estimations were performed on each of the two years. The list of the regressors $\mathbf{x}$ is in appendix.

7.1 The confirmation of an endogeneity bias

The table 5 show that there is an endogeneity bias. The results of the LIML estimation seem to be less reliable than the results of the Heckit and FIML estimators that conclude to similar estimations and confirm the existence of an endogeneity bias.

	Year 1999			Year 2000		
	2-Step	LIML	FIML	2-Step	LIML	FIML
ρ	$0,0516^{**}$	$0,0219^{**}$	$0,0557^{**}$	$0,0338^{**}$	$-0,0018$	$0,0245^{**}$
η	-	$-0,0141$	$0,0414^{**}$	-	$-0,0223^{*}$	$0,0343^{**}$
σ_ε	-	$0,4443^{**}$	$0,4497^{**}$	-	$0,4422^{**}$	$0,4495^{**}$
$\eta\sigma_\varepsilon$	$0,0580^{**}$	-	-	$0,0352^{**}$	-	-

** 1% sign. * 5% sign.

Figure 5: Results on the endogeneity bias

7.2 Simulation of the economic impact of regulation

Work in progress

References

Ma, C.-t. A. (1994), 'Health Care Payment Systems: Cost and Quality Incentives', *Journal of Economics and Management Strategy* **3**(1), 93–112.

Silverman, E. & Skinner, J. (2001), Are For-Profit Hospitals Really Different? Medicare Upcoding and Market Structure. NBER Working Paper 8133 avail. at http://papers.nber.org/papers/w8133.pdf.

A List of the regressors

Code	Label
Variables	
AGE	Age of the mother
CC	Dummy of complication and comorbidities
MULTI	Dummy of multiple deliveries
Dummies of diagnosis (ICD 10)	
O13	Gestational [pregnancy-induced] hypertension without significant proteinuria
O14	Gestational [pregnancy-induced] hypertension with significant proteinuria
O22	Venous complications in pregnancy
O23	Infections of genitourinary tract in pregnancy
O24	Diabetes mellitus in pregnancy
O26	Maternal care for other conditions predominantly related to pregnancy
O32	Maternal care for known or suspected malpresentation of fetus
O33	Maternal care for known or suspected disproportion
O34	Maternal care for known or suspected abnormality of pelvic organs
O36	Maternal care for other known or suspected fetal problems
O41	Other disorders of amniotic fluid and membranes
O42	Premature rupture of membranes
O48	Prolonged pregnancy
O60	Preterm delivery
O61	Failed induction of labor
O62	Abnormality of forces of labor
O64	Obstructed labor due to malposition and malpresentation of fetus
O65	Obstructed labor due to maternal pelvic abnormality
O66	Other obstructed labor
O68	Labor and delivery complicated by fetal stress [distress]
O71	Other obstetric trauma
O75	Other complications of labor and delivery, not elsewhere classified
O86	Other puerperal infections
O87	Venous complications in the puerperium
O92	Other disorders of breast and lactation associated with childbirth
O98	Maternal infectious and parasitic diseases classifiable elsewhere but complicating pregnancy, childbirth, and the puerperium