THE UNIVERSITY *of York*

CHE
Centre For Health Economics

Optimal Contracts and Contractual Arrangements Within the Hospital: Bargaining vs. Take-it-or-leave-it Offers

CHE Research Paper 37

Optimal contracts and contractual arrangements within the hospital: bargaining vs. take-it-or-leave-it offers

Matteo M. Galizzi[1]
Marisa Miraldo[2]

Centre for Health Economics, University of York, UK and University of Brescia, Italy[1]
Centre for Health Economics, University of York, UK[2]

May 2008

Background

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The new CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Acknowledgements

This paper was presented at the Micro Group Seminar of the Department of Economics at the University of York; at the iHEA World Congress, University of Copenhagen; at the Department of Economics of University of Lugano and at the 10th National Health Economics Conference of the Portuguese Association of Health Economics. We thank all the participants for their comments

Further copies

Copies of this paper are freely available to download from the CHE website www.york.ac.uk/inst/che/publications/index.htm. Access to downloaded material is provided on the understanding that it is intended for personal use. Copies of downloaded papers may be distributed to third-parties subject to the proviso that the CHE publication source is properly acknowledged and that such distribution is not subject to any payment.

Printed copies are available on request at a charge of £5.00 per copy. Please contact the CHE Publications Office, email che-pub@york.ac.uk, telephone 01904 321458 for further details.

Centre for Health Economics
Alcuin College
University of York
York, UK
www.york.ac.uk/inst/che

Abstract

We study the impact of different contractual arrangements within the hospital on the optimal contracts designed by third party payers when severity is hospital's private information. We develop a multi-issue bargaining process between doctors and managers within the hospital. Results are then compared with a scenario where doctors and managers decide independently by maximizing their own profit, with managers proposing to doctors a take-it-or leave-it offer. Results show that, when the cost of capital is sufficiently low, the informational rent arising on information asymmetry is higher in a set up where managers and doctors decide together through a strategic bargaining process than when they act as two decision-making units.

***Keywords*:** Strategic Bargaining; Optimal Contracts; Hospitals; Asymmetric Information

JEL classification*:* I11, I18.

1. Introduction

It is often debated by health economists and policy makers which organizational frame could provide better incentives to the different actors within the health care sector. The relevance of such analysis is motivated in light of the rise of health care spending as a share of GDP. Indeed over the period 1960 to 1992, OECD countries have experienced a rise, on average, of more than 4% points, from under 4% to slightly more than 8% (Mossialos and Le Grand, 1999; OECD, 2000). Since public funding is the major funding source for most OECD countries, accounting for, in 2001, an average share of 72.6% (Mossialos and Le Grand, 1999; OECD, 2000) of total expenditure,[1] a steady growth of health expenditure might give rise to serious concerns about the sustainability of public budgets.[2] Therefore, controlling the growth of health expenditure has assumed increasing relevance in the public policy agenda of most OECD countries and has, therefore, lead to several financial reforms.

Financial policies have focused on, but have not been restricted to, mechanisms to control the financial flows, either from third party payers to providers, or between third party payers and patients. Indeed, on the supply side, changes within EU financing mechanisms have focused on hospital cost control (Mossialos, 2002).

Several theoretical studies have analysed the design of optimal payment systems to induce optimal behaviour by providers, i.e. induce optimal levels of quality, cost containment effort and the optimal number of treatments (see for example Chalkley and Malcomson, 1998a and 1998b; Ma, 1994; Rickman and McGuire, 1999; Ellis and McGuire, 1986; Ellis, 1998).

A common assumption, in this literature is that the hospital is considered as a *singleton*, a sort of *black box*. However, in practice, the hospital is a product of an array of different agents with different, and sometimes conflicting, objectives. For example in England, Foundation Trusts' *boards of governors* (with an average size of 33 members) are comprised of a diversity of members enabling patients, the public and staff members to elect representatives so that their interests and views are reflected on the organisations' governance. The *board of governors* will then interact with the *board of directors*, comprising non-executive and executive directors (with normally a maximum of 12 members), balancing the skills and experience to meet the organization needs (Department of Health, 2006).

Indeed, within the hospital, different staff groups will have different objectives. While doctors and consultants are normally seen as being interested in expanding the amount and quality of care supplied, managers are expected to be more interested in hospital surplus and breaking even (Crilly and Le Grand, 2004). Therefore, the relative strength of these different groups is likely to have an impact on resource allocation and, ultimately, on hospital performance (Pauly, 1978). Moreover, the behaviour of staff will be conditioned by the incentives they face, by the institutional arrangements within the hospital as well as by the management structures.

Therefore, as pointed out by Harris (1977), regulatory policy should account for its effects on the diverse decision units within the hospital. Indeed, our paper departs from most of the existing contributions by considering the *hospital* as the outcome of the interaction between different agents. On this matter Custer *et al* (1990) have analysed the effects of a prospective payment system on the production of the hospital, focusing on the relation between the hospital and its medical staff. Other studies that have accounted for different incentives within the hospital have modelled hospitals' internal relations by means of *take-it-or-leave* offers made by one of the parties, which amounts to assuming that only the proposer has decision power. Boadway *et al* (2004), in particular, propose a model with managers and doctors as decision makers and develop a two-stage agency problem in which contracts are designed to elicit

[1]Average of all OECD countries with the exception of Korea, Mexico and the United States.

[2]During the 1980s and 1990s we also observe a shift of costs from the public into the private sector that might reflect the financial reforms introduced in order to control health expenditure

information. Their paper has been extended to consider cooperative bargaining within the hospital by Miraldo (2000).

Our paper is closely related to these works and it generalizes Miraldo's (2000) model introducing strategic negotiations between doctors and managers within the hospital in the spirit of non-cooperative bargaining. The model considers two alternative scenarios of hospitals' organization: the *contract* and *bargaining* set-up. In both scenarios, a multistage game is modelled, and information is assumed to be symmetric between doctors and managers, while asymmetric between the hospital and the government. In both cases, at the first stage, the hospital signs a contract with the third party payer on fee for service, lump-sum transfers and hospital capacity, in a typical Principal-Agent problem. In the subsequent stages, managers and doctors decide on resource allocation within the hospital as well as the number of patients treated with different technologies, according to the decision-making process specific to the set-up. In particular, in the *bargaining* set-up, we model a multi-issue non-cooperative bargaining process, between doctors and managers, over both fees and treated patients. Outcomes from such negotiations are then compared with the *contract* scenario where doctors and managers decide independently by maximizing their own profit, with managers deciding on resource allocation and doctors on treatment allocation.

The results obtained comparing the two scenarios show that when the cost of capital is sufficiently low, the *informational rent is higher* in the set-up where managers and doctors decide together through a strategic bargaining process than when they split into two separate decision-making units.

In the next section we describe the assumptions of the model. In section 3 we develop the scenario of bargaining within the hospital, while section 4 follows with the case in which managers offer contracts to doctors. Government optimal contracts and results are discussed in section 5 and, finally, section 6 concludes.

2. The model

The main actors of our model are the *government* (*G*),[3] the *hospital managers* (*M*) and the *doctors* (*D*). They differ as far as their objective functions are concerned. In particular, hospital *managers* aim at maximizing the expected financial surplus of the hospital. On the other hand, *doctors* are interested not only in their personal income and exerted level of effort, but also in the improvement in health care status of their patients. Finally, *government*'s goal is to trade-off quality of care and health care expenditures.[4]

We consider two alternative organizational set-ups of hospitals. In both set-ups the government offers a contract to the hospitals, specifying a budget (a lump sum T), a capacity (K) and its fee per treated patient (g). The two scenarios only differ in the way decisions are made *within* the hospital. In the *contract* set-up, managers decide the fees per treated patient (h) to be paid to doctors and doctors decide the number (n) of patients to be treated. In the *bargaining* set-up, managers and doctors strategically bargain over both the fees (h) and the number (n) of treated patients. We will come back to the specific differences between the two set-ups in the next section. However, here it is important to underline that both set-ups share a bulk of common modelling assumptions, which we are now going to discuss.

2.1 Hospitals and patients

There are two types of hospital differing on patients' casemix, which we will denote as *type* i with $i = 1, 2$.

[3]It can also be thought of as regulator or third party payer.

[4]Our model could also apply to other secondary care configurations where the government/regulator decentralizes the procurement of health care services to a public organization as it happens, for instance, in England. In this case the government would represent the Department of Health, the managers the local Primary Care Trusts, and doctors would represent the hospital.

All hospitals serve the same total number of patients.
Patients of hospital of type i differ in the severity of illness s_i, which is randomly distributed over a unitary length interval according to a uniform distribution with mean α_i,

$$s_i \in [\alpha_i - \frac{1}{2}, \alpha_i + \frac{1}{2}]$$

Without loss of generality, *type 2* hospital is assumed to have a higher casemix, that is, $\alpha_2 > \alpha_1$. It is also assumed that the mass of patients with a particular realization s_i is equal to one.

The relation between the government and the hospital is modelled as a Principal-Agent problem with asymmetric information. Information on patients' severity is shared perfectly within the hospital but is not known by the government: the government has no information concerning the hospital's case mix of patients in terms of illness severity. The government, however, knows the probability p_i of a hospital being of type i, with $p_1 = 1 - p_2$.

Patients seek care in their local hospital and are offered one of two treatments, *high* technology treatment H or *low* technology treatment L. Alternatively, but equivalently, we can think of H being *treatment* and L *no treatment* decisions.

The benefit of using *high technology* therapy in the treatment of a patient increases with the patient's severity. In contrast, the health improvement when treatment occurs with therapy L is constant, not depending on patient's illness severity. The marginal benefit from the each treatment is given by:

$$q_H(s) = as$$
$$q_L(s) = b$$

with $0 \leq b < a \leq 1$.

Note that, while from a patient's perspective high technology therapy is a more effective treatment as long as $s > \frac{a}{b}$, from a cost containment perspective, its unlimited use may not always be efficient. In fact - as it will be clear once the agents' payoff functions are formally introduced - when direct and indirect costs are accounted for, and from a social welfare perspective, only high illness severity patients should be treated with high technology therapy H. That is, the use of high technology therapy is more beneficial when hospitals are of *type 2*.

For the sake of tractability we will conduct our analysis assuming that $b = 0$ and $a = 1$. This should not be seen as a restrictive assumption since, within the *treatment* versus *no treatment* taxonomy, it just implies that patients with a positive severity must be treated while the remaining should be sent home, what is indeed a realistic simplification of hospitals' decisions under financial and capacity constraints. However, it is important to underline that our main results remain qualitatively unaffected by instead working with generic values for a and b: only computed explicit solutions and comparisons get far more cumbersome.[5]

[5]Of course such formulation ignores induced demand issues, which are beyond the scope of this paper. Note that the main purpose of defining two treatments is simply to represent medical cases for which there exists a threshold of severity beyond which patients have a higher benefit from receiving one treatment rather than the other. This is contemplated within our model.

Depending on the set-up of the model, treatments will be assigned to patients either by hospitals (when doctors and managers act as a unique decision unit) or by doctors only. In general, patients with an illness severity above a certain level $\hat{s}_i$ will be treated with therapy H while the remaining receive treatment L. So the number of patients treated with high technology therapy n_i can be obtained by,

$$n_i \equiv \alpha_i + \frac{1}{2} - \hat{s}_i$$

which is clearly increasing in the casemix of the population. The total health status improvement $Q_i\left(n_i, \alpha_i\right)$ is then given by:

$$Q_i\left(n_i, \alpha_i\right) \equiv \int_{\alpha_i - \frac{1}{2}}^{\alpha_i + \frac{1}{2} - n_i} q_L\left(s\right) ds + \int_{\alpha_i + \frac{1}{2} - n_i}^{\alpha_i + \frac{1}{2}} q_H\left(s\right) ds =$$

$$= \left(\alpha_i + \frac{1}{2}\right) n_i - \frac{1}{2} n_i^2$$

Differentiating one can see how the total health status improvement is affected by n_i and α_i :

$$Q_{n_i} = \left(\alpha_i + \frac{1}{2}\right) - n_i, \quad Q_{\alpha_i} = n_i > 0, \quad Q_{n_i \alpha_i} = 1 > 0$$

Namely, the health status improvement is increasing in the average illness severity and on the number of patients treated with therapy H as long as $\left(\alpha_i + \frac{1}{2}\right) > n_i$, which is always verified as, by construction, $n_i < \alpha_i + \frac{1}{2}$. Finally, the marginal improvement in the health status by treating an extra patient with high technology therapy is higher for hospitals serving more severely ill populations.

2.2 Doctors' problem

In both scenarios, doctors' utility function is defined by,

$$\Pi_{Di} = Q_i(n_i, \alpha_i) + \left(h_i + g_i\right) n_i - \frac{n_i^2}{2} \qquad (1)$$

In other words, within a hospital of type i , doctors directly benefit from the improvement in patients health status $Q_i(n_i, \alpha_i)$ and from the fees paid by the government (g_i) and by the managers (h_i) for each patient treated with therapy H . This doctor reimbursement specification is close to some OECD countries that pay specialist physicians using fee-for-service. Specifically, in Austria, France, Mexico, and New Zealand, specialists employed in private hospitals are mostly paid by fee-for-service. But also in countries where specialists are mostly salaried, such as Australia and England, specialist physicians are paid a fee-for-service for treating private patients in public and private hospitals (Simoens and Giuffrida 2004).

Since high technology treatment requires more time and attention, it is also assumed that, when treating patients with therapy H , doctors should exert a higher effort than the one required under the low technology treatment L . Therefore, doctors also experience a disutility term $\frac{n_i^2}{2}$ which is quadratic in the number of patients treated with therapy H .

If doctors are allowed, as in our *contract* scenario, to decide the number of treated patients, they clearly set n_i at a level that maximizes their utility, trading off earned fees and marginal increase of patients' health status on the one hand, and cost of effort, on the other:

$$n_i^* = h_i + g_i + Q_{n_i}$$

which, as $Q_{n_i} = \left(\alpha_i + \frac{1}{2}\right) - n_i$ becomes:

$$n_i^* = \frac{\left(\alpha_i + \frac{1}{2}\right) + h_i + g_i}{2}$$

The optimal number of patients treated with therapy H in the *contract* scenario is increasing in the average illness severity of the patients' population and in the fees.

2.3 Managers' problem

We assume that the managers' objective is to maximize hospital surplus. This assumption may seem unrealistic for public hospitals operating in a publicly-funded health care system, since they are usually constrained on the distribution of profits. However, it should be noticed that public hospitals may add to their reserves the financial surplus obtained. For example, Foundation Trusts in England, despite being public organizations, are the residual claimants of their losses. With (elected) decision boards risking their position in case the hospital incurs significant losses, these hospitals are, therefore, considered to be profit maximizers (Department of Health, 2002a, 2002b). Alternatively hospitals might wish to maximize surplus in order to pursue other objectives, such as increasing physician staff, broadening the range of services, or even increasing managerial perks (see Chalkley and Malcomson, 1998a and 1998b; De Fraja, 2000; Dranove and White, 1994; Rickman and McGuire, 1999).

Managers receive a budget from the government, via a lump-sum transfer T_i . From this budget managers pay doctors a fee for service h_i for any patient treated with therapy H . Finally, we assume that managers derive disutility from a non efficient run of the hospital's capacity, $\left(n_i - K_i\right)^2 / 2$. Therefore, managers' surplus is given by:

$$\Pi_{Mi} = T_i - \frac{\left(n_i - K_i\right)^2}{2} - n_i h_i \qquad (2)$$

With the disutility term $\frac{\left(n_i - K_i\right)^2}{2}$ we aim at capturing, in a simple way, some further management concerns, such as *capacity* and *congestion* issues. In particular, we consider that there exists a technical capacity level K_i for the therapeutical treatment H . Capacity K_i, on the one hand, measures the maximum number of patients that can be effectively treated without further costs. On the other hand, it refers to the existence of some fixed costs necessary for the implementation of therapy H, such as the investment in a technology, the wages paid to technical staff and nurses, the administrative costs and so on. Therefore, it is assumed that whenever the number of patients treated with therapy H is different from the technical capacity K_i , managers support an extra (quadratic) cost. In fact, the idea is that if the number of treated patients is lower than K_i , the costs from the fixed technical and administrative structure are not optimally recovered. If, on the contrary, n_i is above the capacity, congestion arises in the supply of therapeutic treatment in forms of both direct (e.g. extra hours worked by the staff) and indirect costs, such as an undesirable increase in queues and waiting times.

Clearly, if the managers were called to choose the number of treated patients, they would set it at a level to maximize their own surplus, $n_i^* = K_i - h_i$. However, as discussed above, the decision of how many patients to treat with therapy H is fully in the doctors' hands within the *contract* scenario, while it is set through negotiations with the doctors in the *bargaining* scenario.

The disutility term $\frac{(n_i - K_i)^2}{2}$ exhibits a further intuitive interpretation in terms of the impact of capacity in the managers' surplus. In fact $\partial\Pi_{Mi} / \partial K_i = n_i - K_i$ is positive for $n_i > K_i$ and negative otherwise: if the number of treated patients exceeds capacity, managers' surplus increases with K_i, while it decreases when $n_i < K_i$. Indeed, if the hospital is already operating at an activity level above K_i, expanding capacity will clearly be beneficial as it allows reducing congestion. On the other hand, if the hospital is operating below capacity, then extra capacity will simply increase fixed costs with no counterpart on benefits: consider, for instance, a hospital investing in a second Magnetic Resonance Imaging (MRI) machine, requiring additional room, equipment, staff and maintenance, while the first MRI is still used far below its potential.

2.4 Timing and the contractual arrangements within the hospital

In the first stage, we model a typical Principal-Agent problem where government signs a contract with the hospital, represented by the managers. In particular, the government offers a limited *capacity*, K_i, which constrains the use of high technology treatment; a *lump sum transfer*, T_i, paid *to the managers*; and a *fee,* g_i, for each patient treated with high technology therapy, paid directly *to the doctors.*[6] The financial resources received by the manager are then distributed between doctors and managers in the hospital in the following stages.

As already discussed, the main difference between set-ups lies in the decision-making process within the hospital that can be of two types: contract and bargaining. In particular, in the *contract* set-up, given the government contract, managers decide the fee h_i which is paid to the doctors for any treated patient, while doctors decide the number of patients n_i to be treated. This organizational frame is the one usually adopted in the literature and it emphasizes the nature of decentralized decision-making by the different subjects in the hospital. We develop it for comparison purposes. In particular, we follow the majority of the literature in assuming that such decentralized decisions are taken sequentially and that the managers experience a first mover advantage.

In the *bargaining* set-up, managers and doctors strategically and simultaneously bargain over both the fees h_i and number of treated patients n_i. In this scenario, managers and doctors can be seen as a single decision making unit setting, through a bargaining process, the overall output (n_i, h_i).

Before moving into the details, it is worthwhile noticing that in either case the result of the overall strategic interaction may not be efficient from the government point of view, as its objectives partially diverge from the ones inherent in the hospital's internal decisions. In fact, when doctors and managers are considered as separate decision-makers, on the one hand, the government shares with doctors the aim of providing an effective health status improvement to the most severe patients. However, the costs of provision taken into account by the government are different from the ones considered by the doctors, and interests are directly conflicting as h_i is concerned.

On the other hand, the government shares with managers the objective to allocate patients to treatments in a cost-effective way in order to keep under control the fees and the overall costs of health care provision. However, their interests are also conflicting with respect not only to the lump sum transfer T_i but also for

[6]Note that under our setup the fee can be equivalently interpreted as being a fee-for-service but also a capitation fee.

the capacity level K_i for the high technology treatment.

Our set-up aims at reflecting the current trends in the contractual arrangements between governments and hospitals in the European Union. In fact, several governments in Europe, in accordance with their actual cost containment policies, attempt to design contracts with hospitals in a way to keep the use of some therapies under control and to avoid an over-utilization of expensive treatments. In spite of such supply side pressures, many crucial variables, however, can be exclusively controlled by managers and doctors, whose objectives may not encompass health care cost containment.

3. Bargaining scenario

Under the *bargaining* scenario, the second (and last) stage consists of a bargaining process between doctors and managers. The government allocation $\left(Ti, g_i, K_i\right)$ is taken as given.

3.1 Strategic negotiations between managers and doctors

We solve the second stage of the game using a non-cooperative approach to negotiations among doctors and managers. In particular, managers and doctors bargain *simultaneously* on both the number of patients to be treated n_i and the fee for service h_i. Their dynamic strategic negotiations are modelled as a multi-issue bilateral bargaining with random order of proposer following Osborne and Rubinstein (1990) and Muthoo (1999).

In fact, at any instant of time t, either managers or doctors can be randomly selected respectively with probabilities $(1-\beta)$ and β, to propose offers to the other party. An offer is a pair $\left(h_i, n_i\right)$ with $h_i \geq 0$ and $n_i \geq 0$. The other party can just accept or reject that offer. If it accepts, then each party $L = M, D$ takes his payoff $\Pi_i\left[L\right]\left(h_i, n_i\right)$ corresponding to that offer, with $\Pi_i\left[L\right]\left(h_i, n_i\right)$ given by (1) and (2).

If, instead, the responding party rejects the offer, both managers and doctors enter a further stage of negotiations. Bargaining effort and trade delays are costly and managers and doctors share the same inter-temporal discount factor $\delta \in \left(0,1\right)$, with $\delta \rightarrow 1$ describing the limit case of absence of impatience frictions.

At the beginning of each further bargaining stage, a new random selection of the party entitled to a make proposal is drawn. The game is repeated until managers and doctors agree on a pair $\left(h_i, n_i\right)$. If they reach an agreement on $\left(h_i, n_i\right)$ at any time t, the discounted final payoff would be $\delta^{t-1}\Pi_i\left[L\right]\left(h_i, n_i\right)$. On the other hand, clearly, if they perpetually disagree, their payoffs will be zero.

The described negotiation game is an infinite horizon dynamic game of complete information: in fact, players' payoff functions are common knowledge and at each move in the game the players know the full history of the play thus far. Therefore, in the analysis that follows, we will need to solve the game for its *subgame perfect Nash equilibria*. More precisely, we will look for those players' strategies, describing a complete plan of proposals in the price-offer phases and of decisions of either acceptance or rejection in the response phases, which generate a Nash equilibrium in the immediately subsequent price-response phases and which constitute a Nash equilibrium in every subgame.

Since this is an infinite horizon game, one cannot use the backward induction method to solve for the subgame-perfect Nash equilibria, so that any equilibrium solution of the present game would typically involve a high overall complexity. However, following a standard solution in strategic bargaining models, we will only focus on the subgame-perfect Nash equilibria in *pure* and *stationary* strategies *(PSSPNE)* satisfying *no delay* in the trade. Therefore, it is assumed that any trader always proposes the same price at every equivalent node where she has to make an offer, and she always behaves in the same way whenever facing identical proposals if responding to an offer. Moreover, the property of no delay guarantees that, whenever a player has to make an offer, her equilibrium offer is accepted by the other player.

We now characterize the unique subgame perfect Nash equilibrium of the (h_i, n_i) bargaining game between doctors and managers in pure and stationary strategies. Define $\left(h_i^*[L], n_i^*[L]\right)$ the *PSSPN*-equilibrium offer that agent L proposes whenever she is selected to make an offer. Furthermore, define $\delta V_i^*[L]$ the (discounted) expected equilibrium continuation payoff to player L from accessing a further round of negotiations.

Suppose *managers* have been selected to make offers. The equilibrium payoff from rejecting any offer for the doctors is, clearly, $\delta V_i^*[D]$. Perfection of the equilibrium solution requires that the doctors accept from the managers any proposal $\left(h_i[M], n_i[M]\right)$ such that $\Pi_i[D]\left(h_i[M], n_i[M]\right) \geq \delta V_i^*[D]$, while reject any offer $\left(h_i[M], n_i[M]\right)$ such that $\Pi_i[D]\left(h_i[M], n_i[M]\right) < \delta V_i^*[D]$. Note that we have taken advantage of a standard tie-breaking rule by which, if indifferent between rejecting and accepting an offer, a player accepts it. Furthermore, the property of no delay implies that in equilibrium $\Pi_i[D]\left(h_i^*[M], n_i^*[M]\right) \geq \delta V_i^*[D]$.
However, it is immediately reckoned that in equilibrium it can never be that:

$$\Pi_i[D]\left(h_i^*[M], n_i^*[M]\right) > \delta V_i^*[D]$$

as, otherwise, the managers could always increase their profit by offering the doctors a slightly lower fee: in fact, they could instead propose a pair $\left(h_i'[M], n_i^*[M]\right)$, still accepted, with $h_i'[M]$ slightly lower than $h_i^*[M]$, which improves their own payoff. Therefore, it must be that a managers' equilibrium offer $\left(h_i^*[M], n_i^*[M]\right)$ is such that:

$$\max \ \Pi_i[M]\left(h_i[M], n_i[M]\right)$$

$$s.t.\ \Pi_i[D]\left(h_i[M], n_i[M]\right) = \delta V_i^*[D]$$

A necessary implication of this is that in equilibrium managers choose a number of patients to be treated at which the *joint profits* of doctors and managers are at the highest possible level. This must hold as any potentially attainable surplus share can a fortiori be reached by managers when the total surplus to be divided between doctors and managers is maximum. Indeed, the maximum of $\Pi_i[M]\left(h_i[M], n_i[M]\right)$ given the constraint $\Pi_i[D]\left(h_i[M], n_i[M]\right) = \delta V_i^*[D]$ is, by construction, equivalent to the maximum of the joint profit $\Pi_i[M]\left(h_i[M], n_i[M]\right) + \Pi_i[D]\left(h_i[M], n_i[M]\right)$ valued in $\Pi_i[D]\left(h_i[M], n_i[M]\right) = \delta V_i^*[D]$. Of course, such argument assumes the second order conditions are satisfied for the joint profit function, which we will shortly show being indeed the case.

Hence, define the joint profit as,

$$J\Pi \equiv \Pi_{Mi} + \Pi_{Di}$$
$$= T_i - \frac{n_i^2}{2} - \frac{(n_i - K_i)^2}{2} + Q(n_i^*) + g_i n_i$$

and denote n_i^* as the unique solution of $\partial J\Pi / \partial n_i = 0$, i.e.

$$\left(\alpha_i + \frac{1}{2}\right) + g_i - n_i + K_i = 0$$

namely,

$$n_i^* = \frac{\left(\alpha_i + \frac{1}{2}\right) + g_i + K_i}{3} \quad (3)$$

Such solution turns out to be a global maximum of the joint profit function: in fact, the second order condition is always verified as $\partial^2 J\Pi / \partial^2 n_i < 0$. Notice that n_i^* is increasing with casemix α_i , i.e. given a fixed (g, K) pair, hospitals of *type* 2 always treat a larger number of patients than hospitals of *type* 1 . Therefore, in equilibrium it must be true that the managers always offer the doctors a proposed $n_i^*[M] \equiv n_i^*$ which maximizes their joint profits. Define,

$$J\Pi\left[n_i^*\right] = T_i - \frac{\left(n_i^*\right)^2}{2} - \frac{\left(n_i^* - K_i\right)^2}{2} + Q\left(n_i^*\right) + g_i n_i^*$$

the maximum attainable level of joint profits of doctors and managers, valued at the optimal number of admitted patients n_i^* . Hence the solution to the managers' program is given by a equilibrium pair $\left(h_i^*[M], n_i^*[M] = n_i^*\right)$ such that:

$$\Pi_i[M]\left(h_i^*[M], n_i^*\right) = J\Pi\left[n_i^*\right] - \delta V_i^*[D] \ldots\ldots (4)$$

If, on the other hand, *doctors* have been selected to make offers, by a symmetric line of arguments, the equilibrium offer proposed by the doctors satisfies:

$$\Pi_i[M]\left(h_i^*[D], n_i^*[D]\right) = \delta V_i^*[M]$$
$$n_i^*[D] = n_i^* \quad (5)$$
$$\Pi_i[D]\left(h_i^*[D], n_i^*\right) = J\Pi\left[n_i^*\right] - \delta V_i^*[M]$$

The next step, then, is to look closer at the players' continuation payoffs $V_i^*[L]$, $L = M, D$. Consider the doctors, for instance. Whenever in equilibrium they access a further stage of negotiations, they expect to go through a new random selection of the agent entitled to make a proposal.

As shown above, a *PSSPN* equilibrium, at any node of the game at which they have been called to propose offers, doctors always propose an equilibrium offer $\left(h_i^*[D], n_i^*\right)$ ensuring a surplus of $\Pi_i[D]\left(h_i^*[D], n_i^*\right)$. On the other hand, at any node of the game at which the managers have been selected to make proposals, they always offer a *PSSPN* equilibrium pair $\left(h_i^*[M], n_i^*\right)$ which, by construction, delivers to the doctors a payoff $\delta V_i^*[D]$. Therefore, we can immediately specify the doctors' continuation payoffs in a *PSSPN* equilibrium as:

$$V_i^*[D] = \beta\Pi_i[D]\left(h_i^*[D], n_i^*\right) + (1-\beta)\delta V_i^*[D]$$

which, by direct substitution of $\Pi_i[D]$, is equivalent to:

$$V_i^*[D] = \beta\left[J\Pi\left[n_i^*\right] - \delta V_i^*[M]\right] + (1-\beta)\delta V_i^*[D] = \frac{\beta J\Pi\left[n_i^*\right]}{1-\delta(1-\beta)} - \frac{\beta\delta V_i^*[M]}{1-\delta(1-\beta)} \quad (6)$$

Analogous arguments allow characterizing the managers' continuation payoffs in a *PSSPN* equilibrium as:

$$V_i^*[M] = \frac{(1-\beta)J\Pi\left[n_i^*\right]}{1-\beta\delta} - \frac{(1-\beta)\delta V_i^*[D]}{1-\beta\delta}$$

These equations have clearly a unique solution: in equilibrium managers and doctors expect the continuation values:

$$V_i^*[M] = \frac{J\Pi\left[n_i^*\right](1-\beta)^2}{\beta(\delta+1)(\beta-1)+1}, \quad V_i^*[D] = \frac{J\Pi\left[n_i^*\right]\beta^2}{\beta(\delta+1)(\beta-1)+1}$$

As it can be noticed, the equilibrium continuation values depend on both β and δ. In the limit scenario when the game converges to the case of perfectly patient agents, i.e. when $\delta \to 1$, the equation payoffs tend to: $V_i^*[M] = J\Pi\left[n_i^*\right](1-\beta)^2 / \left(\beta^2 + (1-\beta)^2\right)$ and $V_i^*[D] = J\Pi\left[n_i^*\right]\beta^2 / \left(\beta^2 + (1-\beta)^2\right)$, which, crucially, depend on the bargaining power β. When managers and doctors are equally likely to be selected to make offers, i.e. $\beta = 1/2$, they expect to split equally the surplus of the hospital, obtaining each $J\Pi\left[n_i^*\right]/2$. On the other hand, when doctors (managers) have all the bargaining power, $\beta = 1$ ($\beta = 0$) doctors (managers) earn the whole hospital surplus. This last case represents a *take-it-or-leave* scenario.

Finally, the equilibrium *fee for service* offers can be easily worked out as the $h_i^*[L]$, $L = D, M$ which, given the above continuation values, solve the system:

$$\Pi_i[M]\left(h_i^*[M], n_i^*\right) = J\Pi\left[n_i^*\right] - \delta V_i^*[D]$$

$$\Pi_i[D]\left(h_i^*[D], n_i^*\right) = J\Pi\left[n_i^*\right] - \delta V_i^*[M]$$

This returns the two, asymmetric, equilibrium *fee for service* offers by managers and doctors:

$$h_i^*[M](\delta, \beta, \alpha_i; K_i, T_i, g_i) \neq h_i^*[D](\delta, \beta, \alpha_i; K_i, T_i, g_i)$$

with the exact expressions for $h_i^*[M]$ and $h_i^*[D]$ in Appendix A.

Equilibrium is then reached at the first round with immediate acceptance. It is easy to check that there are no profitable deviations and therefore this pair of strategies is indeed the unique *PSSPN* equilibrium.
It is interesting to notice some salient properties of such equilibrium.

First, in the unique *PSSPN* equilibrium of the multi-issue bargaining game, the equilibrium agreed *number of patients* n_i^* maximizes the joint surplus between managers and doctors $T_i - \frac{\left(n_i^*\right)^2}{2} - \frac{\left(n_i^* - K_i\right)^2}{2} + Q(n_i^*) + g_i n_i^*$. This implies that the bilateral negotiation between doctors and managers is Pareto efficient in the equilibrium number of treated patients.

Second, the equilibrium *fee for service* is $h_i^*[D]$ if the doctors are selected to make offer at the first round, and $h_i^*[M]$ if the managers make proposals at the first round. Contrary to the number of treated patients, equilibrium offers for the fee are thus different. However, notice that, as $\delta \to 1$, i.e. if the game approaches the limit case of perfectly patient players, both proposals for the fee for service converge to the same equilibrium value[7]

$$h_i^*[M]_{\delta\to 1} \equiv h_i^*[D]_{\delta\to 1} \to \overline{h_i^*}(\delta, \beta, \alpha_i; K_i, T_i, g_i)$$

This is clearly due to the fact that, when frictions due to impatience are negligible, the advantage experienced by the first player who has been selected to make offers disappears and the corresponding asymmetries in the proposed equilibrium fees for service also vanish.

Notice, moreover, that the equilibrium level of the fee for service $\overline{h_i^*}(.)$ corresponds to the one found by Miraldo (2000). This is due to the well-known property by which the equilibria of a non-cooperative bargaining game approximate the cooperative Nash bargaining solution as inter-temporal frictions disappear and $\delta \to 1$. Since such expression for $\overline{h_i^*}$ is fully corresponding to the Nash solution and can be directly compared with the equilibrium fee for service under a *contract* scenario (in which impatience does not play any role), hereinafter we will refer to $\overline{h_i^*}$ as the unique *equilibrium fee* for service emerging under the *bargaining* scenario.

The equilibrium of our non-cooperative negotiations game also shows an intuitive interpretation: when bargaining within the hospital, doctors and managers agree on setting the number of treated patients at a level which maximizes hospital profits, and then use the fee for service as an instrument to divide the generated profits.

3.2 Government problem

We now characterize the menu of contracts proposed to hospitals by the government. However, before doing so, it is helpful to provide, as a benchmark, the so-called *full information* contracts. These would be chosen by the government if it were aware of each hospital's average illness severity α_i.

In such a case, for a given α_i, the government chooses $[T_i, g_i, K_i]$ to maximize the difference between the sum of the improvement in patients' health $Q_i(n_i^*, \alpha_i)$ plus the hospital profits $\Pi_{Mi} + \Pi_{Di}$, and the total costs of health care, $-(1+\lambda)(rK_i + T_i + g_i n_i^*)$, subject to the hospitals' participation constraints - IR_i - by which $\Pi_{Mi} + \Pi_{Di} \geq 0$. In fact, when formulating its optimal contract, the government accounts for, within the total costs, the lump-sum transfers T_i, the total fees for service $g_i n_i$, the cost of capital r borne for any unit of capital invested in capacity K_i and the direct and indirect distortions caused by the tax revenue raised to cover health expenditure: $\lambda \in [0,1]$ represents the shadow cost of public funds and captures the deadweight loss per unit of tax revenue.

[7]Exact expression for $\overline{h_i^*}(\delta, \beta, \alpha_i; K_i, T_i, g_i)$ is in Appendix A.

Hence, for each hospital $i=1,2$, the government problem is characterized by:

$$\underset{K_i,g_i,T_i}{Max}\, W = Q_i\left(n_i^*,\alpha_i\right)+\Pi_{Mi}+\Pi_{Di}-\left(1+\lambda\right)\left(rK_i+T_i+g_i n_i^*\right)$$

$s.t.$

$$\left\{IR_i\right\}\ :\quad \left[T_i-\frac{n_i^2}{2}+Q_i\left(n_i^*;\alpha_i\right)-\frac{\left(n_i^*-K_i\right)^2}{2}+g_i n_i^*\right]\geq 0$$

Given that public funds are costly then $\partial W/\partial T_i<0$ and, consequently, T_i will be chosen such that the participation constraints are just satisfied:

$$T_i=\frac{\left(n_i^*\right)^2}{2}-Q_i\left(n_i^*;\alpha_i\right)+\frac{\left(n_i^*-K_i\right)^2}{2}-g_i n_i^*$$

Substituting T_i into W and maximizing with respect to $\left\{K_i,g_i\right\}$, the full information (*first-best*) contracts must satisfy the following first order conditions:

$$\left\{K_i\right\}\ :\quad \frac{\partial W}{\partial K_i}\equiv\left[\frac{\partial Q_i}{\partial K_i}+\frac{\partial J\Pi_i}{\partial K_i}\right]-\left(1+\lambda\right)\left[r+\frac{\partial T_i}{\partial K_i}+g_i\frac{\partial n_i^*}{\partial K_i}\right]=0 \qquad (7)$$

$$\left\{g_i\right\}\ :\quad \frac{\partial W}{\partial g_i}\equiv\left[\frac{\partial Q_i}{\partial g_i}+\frac{\partial J\Pi_i}{\partial g_i}\right]-\left(1+\lambda\right)\left[\frac{\partial T_i}{\partial g_i}+g_i\frac{\partial n_i^*}{\partial g_i}+n_i^*\right]=0$$

where $J\Pi_i=\Pi_{Mi}+\Pi_{Di}$ is the total hospital surplus and n_i^* is the equilibrium number of patients treated, outcome of the negotiations between doctors and managers (3). Note that, in the optimal contract, K_i and g_i will be chosen such that the social marginal benefits (respectively $\partial Q_i/\partial K_i+$ $\partial J\Pi_i/\partial K_i$ and $\partial Q_i/\partial g_i+$ $\partial J\Pi_i/\partial g_i$) equal the social marginal costs (respectively $r+\partial T_i/\partial K_i+g_i\partial n_i^*/\partial K_i$ and $\partial T_i/\partial g_i+g_i\partial n_i^*/\partial g_i+n_i^*)$ weighted by the shadow cost of public funds. Hence, in a full information contract, the government chooses g_i and K_i to optimally trade-off health outcomes and health care costs. Namely, the third party payer will weigh the improvement in patients' health status against the costs of treatment, ensuring that every population is served by one provider. When weighing the financial burden, in particular, the government takes also into account that g_i and K_i directly affect the equilibrium number of treated patients n_i^* as negotiated within the hospital at the last stage of the game. Moreover, the government also explicitly considers the cost of capital r and the indirect effects that both g_i and K_i exert on the levels of the lump-sum transfers T_i necessary to induce hospitals' participation.

It is worthwhile to notice, moreover, that g_i and K_i affect, in a positive and symmetric way, the equilibrium number of treated patients and are, therefore, perfect substitute instruments for the government in this respect. In fact, $\partial n_i^*/\partial g_i=\partial n_i^*/\partial K_i=1/3$. Such a symmetric and positive impact characterizes also the effects of g_i and K_i on the improvement in patients' health status $Q_i\left(\alpha_i,n_i^*\right)$. In fact,

$$\frac{\partial Q_i\left(\alpha_i, n_i^*\right)}{\partial g_i} = \frac{\partial Q_i\left(\alpha_i, n_i^*\right)}{\partial K_i} = \frac{1}{3}\left(\alpha_i + \frac{1}{2} - n_i^*\right) > 0$$

Computing explicit derivatives with respect to K_i and g_i and substituting in the above first order conditions, the policy instruments in the optimal contracts are as follows.[8]

$$g_i^{b,pi} = \frac{3+2\alpha_i}{2\left(2\lambda+3\right)}, \quad K_i^{b,pi} = \frac{\left(\alpha_i + 1/2\right)\left(2+\lambda\right)}{2\lambda+3} - \frac{r\left(4+3\lambda\right)}{2\lambda+3}$$

The full information fee per patient is increasing in casemix and decreasing with the shadow cost of the public funds. The capacity offered by a full information contract, on the other hand, is increasing with the casemix and decreasing with the cost of capital. As a consequence, the equilibrium number of treated patients agreed in the subsequent stage is:

$$n_i^{*b,pi} = \frac{\left(\alpha_i + 1/2\right)\left(2+\lambda\right)}{2\lambda+3} - \frac{r\left(1+\lambda\right)}{2\lambda+3} \qquad (8)$$

It can been noticed that, in the full information benchmark, the number of patients treated by the hospital only differs from the contractual capacity by a positive (but small) term related to the cost of capital.[9] Intuitively, this is due to the fact that the hospital's interests with respect to the number of treated patients are not perfectly aligned with the government's goals which also encompass the cost of capacity r .
Finally, by comparing the number of treated patients across hospitals of different types,

$$n_2^{*b,pi} - n_1^{*b,pi} = \left(\alpha_2 - \alpha_1\right)\frac{2+\lambda}{2\lambda+3}$$

it is noticed that hospitals with higher casemix treat more patients and this difference is directly increasing with the distance in the average illness severity between the two populations.

Of course, the full information policy cannot be implemented when the government is not aware of the hospital's type. Indeed, as in a standard asymmetric information problem, the hospital with the less severe casemix has an incentive to claim of being a hospital of *type* 2 , in order to benefit from the more generous bundle $\left[T_2, g_2, K_2\right]$. In other words, the full information contracts are not incentive compatible. Therefore, the government should design its menu of contracts to induce the hospitals to truthfully reveal their true type. This is equivalent to ensuring that each hospital is better off by choosing a specific contract $\left[T_i, g_i, K_i\right]$ designed for its type rather than mimicking the other type and selecting its contract i.e. IC_i are satisfied. Contracts should also satisfy Individual Rationality constraints (IR_i) for both types so that hospitals' payoffs are large enough for them to operate:

$$\underset{K_i, g_i, T_i}{Max}\, W = \sum_{i=1}^{2} p_i \left[Q_i\left(n_i^*, \alpha_i\right) + \Pi_{Mi} + \Pi_{Di} \right] -$$
$$\left(1+\lambda\right)\sum_{i=1}^{2} p_i \left(rK_i + T_i + g_i n_i^*\right)$$
$$s.t.$$

[8] Note that the subscript b, pi has been used to identify the perfect information benchmark case within the bargaining scenario.

[9] In fact, $\left(1+\lambda\right)/\left(4+3\lambda\right) < 1$ is verified as $\forall \quad \lambda > 0$

$$\{IR_i\} \ : \quad \left[T_i - \frac{\left(n_i^*\right)^2}{2} + Q_i\left(n_i^*;\alpha_i\right) - \frac{\left(n_i^* - K_i\right)^2}{2} + g_i n_i^* \right] \geq 0$$

$$\{IC_i\}: \quad \left[T_i - \frac{\left(n_i^*\right)^2}{2} + Q_i\left(n_i^*;\alpha_i\right) - \frac{\left(n_i^* - K_i\right)^2}{2} + g_i n_i^* \right]$$

$$\geq \left[T_{-i} - \frac{\left(\widehat{n}_{-i}^*\right)^2}{2} + \widehat{Q}_{-i}\left(\widehat{n}_{-i}^*;\alpha_i\right) - \frac{\left(\widehat{n}_{-i}^* - K_{-i}\right)^2}{2} + g_{-i}\widehat{n}_{-i}^* \right]$$

where, T_{-i} , K_{-i} and g_{-i} stand for the elements of the contract intended for hospitals of the type opposite to $i = 1,2$, while

$$\widehat{n}_{-i}^* = \frac{\left(\alpha_i + \frac{1}{2}\right) + g_{-i} + K_{-i}}{3}, \qquad \widehat{Q}_{-i} = \left(\alpha_i + \frac{1}{2}\right)\widehat{n}_{-i}^* - \frac{1}{2}\left(\widehat{n}_{-i}^*\right)^2$$

denote the number of patients treated and corresponding health improvement, by a *type* $i = 1,2$ hospital mimicking one of the opposite *type* $-i$.

Before proceeding with the optimal contract it is useful to investigate the properties of the hospital iso-profit curves. In particular, it is possible to check that a standard *Single Crossing Condition* holds, by which hospitals' indifference curves get steeper as patients' casemix gets more severe. In fact, given the number of treated patients decided by doctors and managers, one can write the hospitals' indirect utility function $v_{H_i}\left(K_i, g_i, T_i;\alpha_i\right) + T_i$ as:

$$v_{H_i}\left(K_i, g_i, T_i;\alpha_i\right) + T_i \equiv T_i - \frac{\left(n_i^* - K_i\right)^2}{2} + Q_i(n_i^*, \alpha_i) + g_i n_i^* - \frac{\left(n_i^*\right)^2}{2}$$

Hence, by computing the total differential and the partial derivatives of $v_{H_i}\left(K_i, g_i, T_i;\alpha_i\right) + T_i$ it is possible to write the marginal rate of substitution between any pair of K_i , g_i and T_i . For instance, it can be seen that:

$$\frac{\partial\left(v_{H_i} + T_i\right)}{\partial g_i} = -\left(n_i^* - K_i\right)\frac{\partial n_i^*}{\partial g_i} + \frac{\partial Q_i}{\partial n_i^*}\frac{\partial n_i^*}{\partial g_i} + g_i\frac{\partial n_i^*}{\partial g_i} + n_i^* - n_i^*\frac{\partial n_i^*}{\partial g_i}$$

$$= \frac{K_i + g_i + \alpha_i + \frac{1}{2}}{3}$$

The marginal rate of substitution among T_i and g_i can then be computed from $\frac{\partial\left(v_{H_i} + T_i\right)}{\partial T_i} dT_i + \frac{\partial\left(v_{H_i} + T_i\right)}{\partial g_i} dg_i = 0$ as

$$\frac{dT_i}{dg_i} = -\frac{K_i + g_i + \alpha_i + \frac{1}{2}}{3}$$

Therefore, as expected, hospitals' indifference curves are downward sloping in the $\left(T_i, g_i\right)$ space. Thus, by analysing the derivative of the above expression with respect to α_i , it is possible to study how such rate of substitution varies with the average patients' illness severity:

$$\frac{\partial}{\partial\alpha_i}\frac{dT_i}{dg_i}\bigg|_{v_{H_i}+T_i}<0$$

We find that, given $\alpha_2>\alpha_1$, the marginal rates of substitution is monotonic, so that the *Single Crossing Condition* holds.

By analogous arguments, it can also be shown that:

$$\frac{\partial}{\partial\alpha_i}\frac{dK_i}{dg_i}\bigg|_{v_{H_i}+T_i}=\frac{12K_i}{\left(2\alpha_i+1-4K_i+2g_i\right)^2}>0$$

$$\frac{\partial}{\partial\alpha_i}\frac{dT_i}{dK_i}\bigg|_{v_{H_i}+T_i}=-\frac{1}{3}<0$$

From the *Single Crossing* conditions, it follows that the IC_2 constraint is always automatically satisfied as long as constraints IR_2 and IC_1 are verified. This standard result is justified in view of the fact that only hospitals of *type 1* have incentives to mimic hospitals of the other type in order to be allowed higher budgets.

Moreover, given that public funds are costly, constraint IC_1 will bind at the optimum: the informational rent to be paid to hospitals of *type 1* will be minimized, leaving such hospitals just indifferent between revealing their true type and mimicking hospitals of *type 2*. That is, the government will choose T_1 so that the following equality would be satisfied:

$$\{IC_1\}\ :\quad T_1+Q_1\left(n_1^*;\alpha_1\right)+g_1n_1^*-\frac{\left(n_1^*\right)^2}{2}-\frac{\left(n_1^*-K_1\right)^2}{2}=$$

$$=T_2-\frac{\left(\widehat{n}_2^*\right)^2}{2}+\widehat{Q}_2\left(\widehat{n}_2^*;\alpha_1\right)-\frac{\left(\widehat{n}_2^*-K_2\right)^2}{2}+g_2\widehat{n}_2^*$$

In fact, the expression on the right hand side:

$$R\left(K_2,g_2;\alpha_1\right)=\left\{T_2-\frac{\left(\widehat{n}_2^*\right)^2}{2}+\widehat{Q}_2\left(\widehat{n}_2^*;\alpha_1\right)-\frac{\left(\widehat{n}_2^*-K_2\right)^2}{2}+g_2\widehat{n}_2^*\right\}\qquad(9)$$

stands for the *informational rent* accruing to hospitals of *type 1*. Clearly, the lump sum remuneration of managers T_1 will be higher than in a full information case in order to induce hospitals to choose the contract intended for their type. Since this rent is positive, the participation constraint for hospital of *type 1* (IR_1) is also automatically verified.

On the other hand, in the contract intended for hospitals of *type 2* there will be no informational rent: T_2 will be chosen such that the hospital participation constraint just binds:

$$\{IR_2\}\ :\quad T_2=\frac{\left(n_2^*\right)^2}{2}+\frac{\left(n_2^*-K_2\right)^2}{2}-Q_2\left(\widehat{n}_2^*;\alpha_2\right)-g_2n_2^*$$

In order to investigate the role of K_i , g_i and T_i in the government contract, it is useful to have a closer look at the *informational rent*. Substituting T_2 in $R^b(K_2, g_2; \alpha_1)$, comparative statics' analysis shows that the impact of both K_2 and g_2 on the above informational rent is negative and equal in magnitude. In fact:

$$\frac{\partial R^b(K_2, g_2; \alpha_1)}{\partial K_2} = \frac{\partial R^b(K_2, g_2; \alpha_1)}{\partial g_2} = -\frac{1}{3}(\alpha_2 - \alpha_1) < 0$$

Therefore, the capacity and the fee for service are perfect substitutes as policy instruments targeted at the minimization of the informational rent.

However, when looking at the design of the menu of contracts, we need to keep in mind that the government objective of provision of cost effective health care involves not only the minimization of the informational rent, but also the maximization of patients' total health improvement and the control of health treatment's costs. Therefore, the policy instruments in the optimal contracts will reflect the trade-off between all these elements.

Maximizing W, with respect to K_i and g_i, it can be seen that the optimal contract for the hospitals of *type 1* is described by the following first order conditions,

$$\{K_1\}: \quad \frac{\partial W}{\partial K_1} \equiv p_1\left[\frac{\partial Q_1}{\partial n_1^*}\frac{\partial n_1^*}{\partial K_1} + \frac{\partial J\Pi_1}{\partial K_1}\right] -$$
$$-(1+\lambda)\,p_1\left[r + n_1^*\frac{\partial n_1^*}{\partial K_1} + (K_1 - n_1^*)\left[1 - \frac{\partial n_1^*}{\partial K_1}\right] - \frac{\partial Q_1}{\partial n_1^*}\frac{\partial n_1^*}{\partial K_1}\right] = 0$$

$$\{g_1\}: \quad \frac{\partial W}{\partial g_1} \equiv p_1\left[\frac{\partial Q_1}{\partial n_1^*}\frac{\partial n_1^*}{\partial g_1} + \frac{\partial J\Pi_1}{\partial g_1}\right] -$$
$$-(1+\lambda)\,p_1\left[n_1^*\frac{\partial n_1^*}{\partial g_1} - (K_1 - n_1^*)\left(\frac{\partial n_1^*}{\partial g_1}\right) - \frac{\partial Q_1}{\partial n_1^*}\frac{\partial n_1^*}{\partial g_1}\right] = 0$$

where $\partial J\Pi_1 / \partial K_1$ and $\partial J\Pi_1 / \partial g_1$ capture the impact of K_1 and g_1 on the overall hospital surplus while $(\partial Q_1 / \partial n_1^*)(\partial n_1^* / \partial K_1)$ and $(\partial Q_1 / \partial n_1^*)(\partial n_1^* / \partial g_1)$ stand for the effects of the policy instruments on the improvement on patients' health status, an objective that the government is able to control only indirectly, through the influence on the equilibrium number n_1^* of patients negotiated within the hospital. On the other hand, the negative term in the second brackets multiplied by $(1+\lambda)$ represents the overall costs associated with the effect of the policy instruments on the equilibrium number of treated patients: in particular, concerning capacity, the costs are net of the beneficial effect on the increase in patients' health status $(\partial Q_1 / \partial n_1^*)(\partial n_1^* / \partial K_1)$, and include the cost of capacity r, the induced cost $n_1^*(\partial n_1^* / \partial g_1)$ of treated patients in terms of higher fees for services paid by the government, and the indirect cost $(K_1 - n_1^*)\left[1 - (\partial n_1^* / \partial K_1)\right]$ by sub-optimal utilization of the medical technology, arising anytime the number of treated patients diverges from the allowed capacity and as long as their relative rates of adjustment are different. Concerning the fees for service, as there is no related cost of capacity, costs, net of the beneficial health status improvement $(\partial Q_1 / \partial n_1^*)(\partial n_1^* / \partial K_1)$, only include the induced cost $n_1^* \partial n_1^* / \partial g_1$ of treated patients in terms of higher fees paid by the government, and the indirect cost

$\left(n_1^* - K_1\right)\left[\partial n_1^* / \partial g_1\right]$ by congestion.

The first order conditions for the contract intended for hospitals of *type 2* are analogous to the ones for hospitals of *type 1*, while showing some supplementary terms standing for the impact of g_2 and K_2 on the informational rent $R^b\left(K_2, g_2; \alpha_1\right)$:

$$\begin{aligned}\{K_2\}: \quad & \frac{\partial W}{\partial K_2} \equiv p_2\left[\frac{\partial Q_2}{\partial n_2^*}\frac{\partial n_2^*}{\partial K_2} + \frac{\partial \Pi_{H2}}{\partial K_2}\right] - \\ & -\left(1+\lambda\right) p_2\left[r + n_2^*\frac{\partial n_2^*}{\partial K_2} + \left(K_2 - n_2^*\right)\left[1 - \frac{\partial n_2^*}{\partial K_2}\right] - \frac{\partial Q_2}{\partial n_2^*}\frac{\partial n_2^*}{\partial K_2}\right] - \\ & -\left(1+\lambda\right) p_2 \frac{\partial R^b\left(K_2, g_2; \alpha_1\right)}{\partial K_2} = 0\end{aligned}$$

$$\begin{aligned}\{g_2\}: \quad & \frac{\partial W}{\partial g_2} \equiv p_2\left[\frac{\partial Q_2}{\partial n_2^*}\frac{\partial n_2^*}{\partial g_2} + \frac{\partial \Pi_{H2}}{\partial g_2}\right] - \\ & -\left(1+\lambda\right) p_2\left[n_2^*\frac{\partial n_2^*}{\partial g_2} - \left(K_2 - n_2^*\right)\left(\frac{\partial n_2^*}{\partial g_2}\right) - \frac{\partial Q_2}{\partial n_2^*}\frac{\partial n_2^*}{\partial g_2}\right] - \\ & -\left(1+\lambda\right) p_2 \frac{\partial R^b\left(K_2, g_2; \alpha_1\right)}{\partial g_2} = 0\end{aligned}$$

Hence, it can be immediately seen that the allocation of g_i and K_i will differ across different types. In particular, solving explicitly the above first order conditions, it can be seen that the allocation of both g_1 and K_1 - targeted at hospitals with lower casemix – will not be distorted with respect to the full information benchmark:

$$\begin{aligned} g_1^b &= g_1^{b,pi} = \frac{2\alpha_1 + 2r + 1}{2\left(2\lambda + 3\right)} \\ K_1^b &= K_1^{b,pi} = \frac{\left(2+\lambda\right)\left(\alpha_1 + 1/2 - r\right) - 2r\left(1+\lambda\right)}{2\lambda + 3} \end{aligned} \qquad (10)$$

As a consequence, the number of patients of *type 1* treated in equilibrium corresponds to the one treated in first best:

$$n_1^b = n_1^{b,pi} = \frac{\left(\alpha_1 + 1/2\right)\left(2+\lambda\right)}{2\lambda + 3} - \frac{r\left(1+\lambda\right)}{2\lambda + 3}$$

This set of results is standard in asymmetric information problems, and is known as the property of *non distortion at the top.*

However, as both g_2 and K_2 are allocated taking into account the effects conveyed by the extra terms, the contract intended for hospitals of *type 2* clearly differs from the full information benchmark. In fact, the government will choose the optimal allocations g_2 and K_2 in such a way as to trade off between the allocative efficiency goal and the minimization of the informational rent.

More precisely, the exact choice of $\{K_2, g_2\}$ in the optimal contract will depend on the interaction between the impact of these policy instruments on the informational rent reaped by hospital of *type 1* and their direct and indirect effects on the other objectives envisaged by the government. Since, $\partial R / \partial g_2 = \partial R / \partial K_2 < 0$, the minimization of the informational rent implies that at least one of these two instruments should be over-provided by the government. However, the socially optimal care supply must also consider patients' health status, hospitals' surplus, and the number of patients treated within the hospital. From the government's perspective, the latter target should optimally trade off the benefit from the enhancement of patients' health status with the reduction of the overall costs in health care provision, including the cost of capacity, the fees paid to doctors and the costs due to sub-optimal utilization of medical technology or to congestion, as captured by the terms r , $n_2^* \partial n_2^* / \partial K_2$, $n_i^* \partial n_2^* / \partial g_2$, $\left(K_2 - n_2^*\right)\left(1 - \partial n_2^* / \partial K_2\right)$ and $\left(n_2^* - K_2\right)\left(\partial n_2^* / \partial g_2\right)$, respectively. Therefore, the optimal choice of g_2 and K_2 will ultimately depend on the complex interaction between all the above effects.

Solving explicitly the first order conditions the optimal menu of contracts can be characterized as,[10]

$$
\begin{aligned}
g_2^b &= \frac{2\lambda\left(\alpha_2 - \alpha_1\right)}{2\lambda + 3} + \lambda r + \frac{\alpha_2 + r + 1/2}{2\lambda + 3} \\
K_2^b &= \frac{2\lambda\left(\alpha_2 - \alpha_1\right) + \left(\lambda + 2\right)2\left(\alpha_2 + 1\right) - 4\left(\lambda + 2\right)\left(1 + \lambda\right)r}{2\left(2\lambda + 3\right)} \\
n_2^b &= \frac{2\left(2\alpha_2 - r\right) + \left(2 + \lambda\right) - 2\lambda\alpha_1}{2\left(2\lambda + 3\right)}
\end{aligned}
\tag{11}
$$

Comparing with the optimal contract under perfect information,

$$
g_2^b - g_2^{b,pi} = \frac{2\lambda\left(\alpha_2 - \alpha_1\right)}{3 + 2\lambda} + r\lambda > 0 \tag{12}
$$

it can be seen that the optimal contract's fee for patients is always higher than in a first best allocation. Since the effects on the (reduction of) informational rent and on (the maximization of) patients' health status prevail, hospitals with more severe casemix receive higher fees than under full information.
Moreover it can be seen that the difference between the capacity levels under asymmetric and full information depends on the cost of capacity r :

$$
K_2^b - K_2^{b,pi} = \frac{\lambda\left(\alpha_2 - \alpha_1\right)}{3 + 2\lambda} - r\lambda \tag{13}
$$

In particular, for sufficiently *low (high) cost of capacity*, namely if $r < (>) \left(\alpha_2 - \alpha_1\right)\left(3 + 2\lambda\right)$, the capacity allowed to *type 2* hospitals is greater (lower) than the one under perfect information, $K_2^b > K_2^{b,pi}$ ($K_2^b > K_2^{b,pi}$). As capacity becomes cheaper, the government is willing to over-provide capacity to hospitals with more severe casemix.

[10]The second order conditions are always satisfied. In fact, it is possible to see that the four leading minors are of alternating signs with the sign of the first leading minor being negative, i.e. $|H1| < 0$

The fact that both the allocations of K_2 and g_2 are distorted, is standard under asymmetric information contracts. However, it should be noticed that, even though both instruments are perfect substitutes (both in controlling the informational rent and in influencing the equilibrium number of patients and, thus, health status) the extent to which one will be preferred over the other will ultimately depend on the relative (social) cost of their usage. In that respect, when capacity becomes too expensive, then the government will substitute it away for g_2.

Finally, the above trade-offs also imply that the number of patients treated in high casemix hospitals is larger than under perfect information:

$$n_2^b - n_2^{b,pi} = \frac{\lambda(\alpha_2 - \alpha_1)}{2\lambda + 3} > 0$$

Notice, in fact, that this distortion, with respect to the first best, not only does not depend on the capacity cost but is also greater than the corresponding positive distortion in the capacity. Hence, hospitals with more severe casemix treat more patients, and such difference is greater than in the first best:

$$n_2^b - n_1^b \equiv (\alpha_2 - \alpha_1)\frac{2(1+\lambda)}{3+2\lambda} >$$

$$> (\alpha_2 - \alpha_1)\frac{2+\lambda}{3+2\lambda} \equiv n_2^{b,pi} - n_1^{b,pi}$$

4. Contract scenario

In the contract scenario, doctors and managers act as two separate units, deciding n_i and h_i respectively. Doctors, given managers and government decisions, decide on the number of patients to be treated with therapy H by maximizing (1). The optimal number of treated patients is the level $n_i^* = \arg\max\left\{\Pi_{D_i} \equiv Q_i(n_i, \alpha_i) + (h_i + g_i)n_i - \frac{n_i^2}{2}; 0\right\}$

As doctors treat patients with technology H until the treatment's marginal benefit equals its marginal cost, it follows that

$$n_i^* = \frac{(\alpha_i + \frac{1}{2}) + h_i + g_i}{2}$$

Notice that the equilibrium number of treated patients n_i^* is increasing with casemix α_i . Also notice that, under a contract scenario, n_i^* is independent of the capacity K_i contracted by the government, while it clearly depends not only on g_i but also on the fee h_i set by managers. Indeed, the fees for service paid by the managers and by the government are now perfect substitutes to the extent at which they affect the number of treated patients decided by doctors: $\partial n_i^* / \partial g_i = \partial n_i^* / \partial h_i > 0$. On the other hand, contrary to the bargaining scenario, the impact of capacity and fee for services offered by the government are no longer symmetric. Indeed, while the capacity has no impact on doctors' choices ($\partial n_i^* / \partial K_i = 0$), fees for service affect positively the number of patients treated and this effect is greater than the analogous under bargaining scenario. The same holds with respect to health benefits:

$$\frac{\partial Q_i}{\partial g_i} = \frac{1}{2}\left(\alpha_i + \frac{1}{2} - n_i^*\right) > 0, \qquad \frac{\partial Q_i}{\partial K_i} = 0$$

Given doctors' decision, managers will then offer the contract on h_i that maximizes their own profits subject to the doctors' participation constraint:

$$\underset{h_i}{Max}\ \Pi_{Mi} = T_i - \frac{\left(n_i^* - K_i\right)^2}{2} - n_i^* h_i$$

$$s.t.\ \Pi_{Di} \geq 0$$

The contract works as a *take-it-or-leave-it* offer and therefore, as long as their non-negative profit constraint is satisfied, doctors will accept any fee for patient offered by managers. It can be checked that since $\partial \Pi_{Mi} / \partial h_i < 0$ and $U\big|_{h_i=0} > 0,$ then the manager will optimally set $h_i^* = 0$ [11].

Analogously to the bargaining scenario, in a perfect information scenario the menu of contracts offered by the government is be given by[12]:

$$g_i^{c,pi} = \frac{\alpha_i + 1 - 2r\left(1+\lambda\right)}{2\lambda+3},\quad K_i^{c,pi} = \frac{\left(2+\lambda\right)\left(\alpha_i+1\right)}{2\lambda+3} - r$$
$$n_i^{c,pi} = \frac{\left(2\alpha_i+1\right)\left(\lambda+2\right) - 2r\left(1+\lambda\right)}{2\left(2\lambda+3\right)},\quad n_1^{c,pi} - n_2^{c,pi} = \left(\alpha_1-\alpha_2\right)\frac{2+\lambda}{2\lambda+3} \tag{14}$$

As the policy that would be optimal in case of full information cannot be implemented when the government is not aware of the hospital's type, the menu of contracts is designed by the government in order to maximize the total welfare function W, subject to the usual *Individual Rationality* (IR_i) and *Incentive Compatibility* (IC_i) constraints for both types $i = 1,2$:

$$\underset{K_i, g_i, T_i}{Max}\, W =$$
$$\sum_{i=1}^{2} p_i \left[Q_i\left(n_i^*, \alpha_i\right) + \Pi_{Mi} + \Pi_{Di} \right] - \left(1+\lambda\right) \sum_{i=1}^{2} p_i \left(rK_i + T_i + g_i n_i^*\right)$$

$$\{IR_i\}\ :\quad \left[T_i - \frac{\left(n_i^*\right)^2}{2} + Q_i\left(n_i^*;\alpha_i\right) - \frac{\left(n_i^* - K_i\right)^2}{2} + g_i n_i^* \right] \geq 0$$

$$\{IC_i\}:\quad \left[T_i - \frac{\left(n_i^*\right)^2}{2} + Q_i\left(n_i^*;\alpha_i\right) - \frac{\left(n_i^* - K_i\right)^2}{2} + g_i n_i^* \right]$$
$$\geq \left[T_{-i} - \frac{\left(\widehat{n}_{-i}^*\right)^2}{2} + \widehat{Q}_{-i}\left(\widehat{n}_{-i}^*;\alpha_i\right) - \frac{\left(\widehat{n}_{-i}^* - K_{-i}\right)^2}{2} + g_{-i}\widehat{n}_{-i}^* \right]$$

where $\widehat{n}_{-i}^*$ and $\widehat{Q}_{-i}$ denote the treated patients' number and health improvement, respectively, for the hospital of *type* $i = 1,2$ mimicking one of the opposite *type* $-i$.

Note that also in the contract scenario the *Single Crossing Conditions* are verified (in Appendix B).

[11]Notice that such a "corner" solution is also common to Boadway et al (2004) and serves us mainly as a benchmark scenario.

[12]The subscript c, pi has been used to identify the perfect information contract scenario.

In the optimum the contract for hospitals of *type 1* is not distorted with respect to the perfect information benchmark (14) [13] ($g_1^c = g_1^{c,pi}, K_1^c = K_1^{c,pi}$ and $n_1^c = n_1^{c,pi}$) while the one for the hospitals of *type 2* is such that:

$$
\begin{aligned}
g_2^c &= \frac{2\lambda(\alpha_2 - \alpha_1) + \alpha_2 + 1/2 - 2r(1+\lambda)}{2\lambda + 3} \\
K_2^c &= \frac{(\alpha_2 - \alpha_1)(5 + 2\lambda) + (\alpha_2 + 1/2 - r(1+\lambda))(\lambda + 2)}{2\lambda + 3} \\
n_2^c &= \frac{2(2\alpha_2 - r) + (2 + \lambda) - 2\lambda\alpha_1}{2(2\lambda + 3)}
\end{aligned} \tag{15}
$$

Finally, hospitals serving populations with higher average illness severity treat more patients than hospitals with lower average illness severity,

$$n_2^c - n_1^c = \frac{2(\alpha_2 - \alpha_1)(1+\lambda)}{2\lambda + 3} > 0$$

Interestingly, this difference is the same under the bargaining and the contract scenario.

Comparing the fee for *type 2* with the perfect information benchmark,

$$g_2^c - g_2^{c,pi} = 2\lambda \frac{\alpha_2 - \alpha_1}{2\lambda + 3} > 0$$

It turns out that, as in the bargaining set-up, hospitals of *type 2* are given higher fees than in perfect information. However, given (12), this distortion is smaller in the contract than in the bargaining set-up. In particular, the distortion is smaller by a factor $r\lambda$ which is directly related to the cost of capacity. Indeed, while under bargaining scenario g_2 and K_2 were found to be perfect substitutes and the government was able to directly substitute away K_2 with g_2 when the cost of capacity was high, this is no longer the case under the contract scenario.

Proceeding in a way analogous to the bargaining scenario, it can be shown that, even though both K_2 and g_2 have a negative impact on the informational rent,

$$\frac{\partial R^c(K_2, g_2; \alpha_1)}{\partial K_2} = \frac{\alpha_1 - \alpha_2}{2} < 0, \quad \frac{\partial R^c(K_2, g_2; \alpha_1)}{\partial g_2} = \frac{\alpha_1 - \alpha_2}{4} < 0$$

capacity K_2 is twice more effective in reducing the informational rent with respect to the fee for service g_2 : $\left|\partial R^c / \partial K_2\right| > \left|\partial R^c / \partial g_2\right|$.

Comparing capacity levels:

$$K_2^c - K_2^{c,pi} = \frac{\lambda}{2(2\lambda + 3)} \left\{ \frac{2\lambda + 5}{2\lambda + 3}(\alpha_2 - \alpha_1) - 2r \right\}$$

Hence, the capacity allowed to hospital of *type 2* is the same as under perfect information only in the special

[13] The expressions for the lump sum transfers are cumbersome. Given that their role is not so informative we do not explicitly state them in the paper.

case in which the cost of capacity assumes the punctual value $r = 2(\alpha_2 - \alpha_1)(2\lambda + 5)2(2\lambda + 3)$. For sufficiently low (high) capacity prices the *distortion* in the allowed capacity is *positive* (negative): $K_2^c > K_2^{c,pi}$ $(K_2^c < K_2^{c,pi})$. That is, the outcome and the distortion with respect to the perfect information case are qualitatively the same under the contract and bargaining scenarios.

5. Discussion

In the last sections, we have presented the optimal financing contracts offered by a third party payer under bargaining and contracts within the hospital. In this section we further discuss those findings.

As a benchmark, it may be helpful to start with a direct comparison between the government's contracts that would be offered in the two scenarios under *perfect information*. Regarding the fees for service it can be seen that, under perfect information, they are higher in the bargaining than in the contract scenario: $g_i^{b,pi} - g_i^{c,pi} = r > 0$. Notice that the difference equals the cost of capacity. Intuitively this is related to the different role of g_i under the two scenarios. In fact, as discussed above, the fee paid by the government to the doctors is a policy instrument that in both set-ups can be used to impact on the different elements that determine social welfare: the informational rent, the number of patients treated and their health benefit, and the social and private costs of treatment.

Regarding the informational rent, clearly it is not an issue in the perfect information case. With respect to the monetary incentives the two scenarios differ. First, while in the bargaining scenario the equilibrium fee for service h_i^* is set such that the hospital's surplus is shared equally between doctors and managers, in the contract scenario h_i^* is set so that doctors are just provided the minimum monetary payment to induce their participation. Therefore, the government's scope to integrate monetary payments and to induce optimal doctors' behaviour is clearly higher in the contract than in the bargaining scenario. Therefore, given that, from doctors' perspective, h_i and g_i are perfect substitutes, one should expect higher fees paid by the government to doctors when managers choose h_i unilaterally than when it is decided through bilateral negotiations. However, the different way in which the number of patients is set in the two scenarios may counterbalance and even crowd-out this effect. In fact, in the bargaining scenario the number of patients is also negotiated by both managers and doctors, and is, therefore, set, in equilibrium, at a level encompassing the interests of both parties, and taking into account the direct and indirect costs of the hospital, including the costs associated with inefficient use of capacity. On the other hand, under the contract scenario, the number of treated patients is decided exclusively by the doctors, who choose it in an opportunistic way disregarding the impact of their decision on the overall costs of the hospital. In fact they are more likely to strategically manipulate the number of treated patients in order to compensate the loss in their revenues due to the zero fees paid by managers, with artificially high revenues earned by inflating the number of treated patients above the optimal level. The complete split between the decision-making of doctors and managers within the hospital, in the contract scenario, would thus generate not only a failure of coordination on the optimal level of patients to be treated (from the hospital's point of view), but also a negative externality on the government in the form of higher fees to be paid to doctors. To counterbalance this effect, the government reduces its fees g_i to compensate or even to dominate the first effect.

With respect to the (socially) optimal number of patients treated, the government must trade-off increases in overall costs of health care provision against improved health outcomes. As discussed in the previous sections, an increase in the fee for service g_i induces a higher number of treated patients in both contractual scenarios. The absolute size of such positive effect, however, is higher in the contract scenario ($\partial n_i^* / \partial g_i^c > \partial n_i^* / \partial g_i^b$) and, therefore, a smaller level of the fee in the contract scenario is, therefore, sufficient to induce the same number of treated patients in the bargaining scenario and this may further explain why $g_i^{b,pi} > g_i^{c,pi}$. Moreover, in the contract scenario, since doctors decide n_i^* without internalizing the overall costs for the hospital, the number of treated patients does not depend on the capacity offered and, therefore, only g_i can be used to control such number. On the contrary, in the bargaining scenario both g_i and K_i can be used in direct and perfect substitution to control the equilibrium number of treatments. Hence, under bargaining, the government can use g_i instead of K_i as an instrument to affect n_i^* whenever capacity is a relatively more expensive policy tool. Given that the relative price of fees with respect to capital is $1/(1+r),$ whenever, in the bargaining scenario, the government substitutes away one unit of capacity K_i with a unitary increase in the fee g_i, it is saving r, which also explains why $g_i^{b,pi} - g_i^{c,pi} = r$.

Furthermore, notice that, under perfect information, the government would allow exactly the same capacity under both scenarios $K_i^{b,pi} = K_i^{c,pi}$. In fact, with perfect information, capacity is chosen to maximize overall efficiency and to minimize the total costs of health care provision, including the costs of capacity which are identical in both scenarios. Therefore, also the number of treated patients is identical under both scenarios.

We finally compare the contracts offered in both scenarios under *asymmetric information*. With respect to the contracts offered to *type 1* hospitals, given the *non-distortion-at-the-top* result, the difference between bargaining and contract scenarios exactly corresponds to the difference under *perfect information*. On the other hand, perfect information contracts targeted at higher casemix hospitals are distorted under asymmetric information in order to minimize the informational rent.

Given that, in the bargaining scenario, the fee for service g_2 and the capacity K_2 are perfectly substitutable in reducing the informational rent, while in the contract K_2 is twice more effective,

$$\left|\frac{\partial R^c}{\partial K_2}\right| = \frac{\alpha_1 - \alpha_2}{2} > \left|\frac{\partial R^b}{\partial K_2}\right| = \left|\frac{\partial R^b}{\partial g_2}\right| = \frac{\alpha_1 - \alpha_2}{3} > \left|\frac{\partial R^c}{\partial g_2}\right| = \frac{\alpha_1 - \alpha_2}{4}$$

then, *ceteris paribus*, the distortion on the fee for service for the hospitals of *type 2* should indeed be larger in the bargaining than in the contract set-up: $g_2^b - g_2^c = (\lambda + 1) r > 0.$

Indeed, as the use of capacity is costly, when the government substitutes away one unit of K_2 with a unitary increase in g_2 under the bargaining scenario, it still experiences the same envisaged reduction in the informational rent while it also benefits from a saving equal to r.

Therefore, when the cost of capacity, $r,$ is low, the capacity allowed to hospitals of *type 2* is greater than it would be optimal in perfect information under both contractual scenarios: $K_2^b > K_2^{b,pi}$, $K_2^c > K_2^{c,pi}$. However, the effect is not the same in magnitude across the two contractual set-ups: in particular, for any given level of the costs of capacity, the capacity allowed to hospitals of *type 2* is always smaller in the bargaining than in the contract: $K_2^b - K_2^c = \frac{\lambda}{2}(\alpha_1 - \alpha_2) < 0.$ Indeed, for a given cost of capacity r, K_2 is a less cost-effective policy in the bargaining than in the contract. Therefore, since the optimal level of capacity in the perfect information case is the same under both scenarios, even for low capacity costs, the over-provision of capacity, with respect to the optimum, is less pronounced under the bargaining than in contract scenario.

Finally, it is worthwhile to underline that under both scenarios the government manages to induce the perfect information number of treated patients for both types of hospitals, and that such a level is equal under both the bargaining and the contract scenarios. Nevertheless, the policy instruments used by the government to achieve such a result differ between the two scenarios.

The last result we are going to compare across different scenarios is the informational rent. It is shown that when the costs of capacity are sufficiently low (high), the rent allowed to hospitals of *type 2* is lower (higher) in the contract scenario than in the bargaining set-up (comparisons in Appendix C).

The ultimate reason beyond such a result is related to the interaction and trade-offs among all the described effects in action in both scenarios. In the *contract* scenario, in fact, the government faces more than one trade-off when designing the menu of contracts to offer hospitals with more severe casemix. If, on the one hand, an increase in the capacity K_2 is twice more effective in reducing the informational rent than an equivalent increase in fees paid to doctors g_2, on the other, capacity K_2 is also more costly, as a policy instrument, than the fees paid to doctors. Hence, in the control of informational rents, the relative use of these two instruments will be decided on cost-effectiveness grounds. In particular, when r is low enough, the allowed capacity K_2 is much more cost-effective in reducing the informational rent than the fee g_2 and, therefore, the government uses more intensively K_2 than g_2 to reduce the informational rent.

Note that, while K_2 and g_2 are (imperfect) substitute instruments vis-à-vis the informational rent, in the contract scenario they can not be used interchangeably to control the number of treated patients, and to balance total costs of health care against patients' improved health. Indeed since doctors decide independently n_i^* (without internalizing managers' objectives) then their decision is not affected by capacity levels $\left(\partial n_2^* / \partial K_2\right)^c = \left(\partial Q_2^* / \partial K_2\right)^c = 0$. Moreover, also the fee paid by managers does not affect doctor's decisions since in equilibrium it is set at $h_i^* = 0$. Therefore, under the contract scenario, the fee for service g_2 is the only available instrument to control the number of treated patients and will be set in order to optimally trade-off between the direct and indirect costs of health care provision and the enhancement of patients' health status. Therefore, it is as if the separation of decisions units within the hospital, typical of the contract scenario, induces the government to use its main instruments K_i^c and g_i^c in a similar "dedicated" way, by clearly separating their role and their impact: on the one hand, the capacity K_i^c is used to minimize the informational rent, provided that the cost of capacity r is low enough; on the other hand, the fee for service g_i^c is set to affect and to control the doctors' decision on the number of treated patients and to optimally trade-off the improvement in health status with the total costs of health care.

Notice also that, in the *bargaining* scenario there is a trade-off in the use of K_i^b against g_i^b. In fact, the government knows that the decision-making within the hospital is not separate but is an integrated one, in which doctors and managers negotiate simultaneously both on the number of treated patients and the fees for doctors. Hence, it can no longer easily identify which agent is ultimately responsible for an objective and which instrument is more appropriate to induce desired changes in the behaviour and the relevant variables. In fact, recall that K_2 and g_2 were found to be *perfect* substitutes in their impact on the rent and on the number of treated patients inducing the government to use the instrument associated to the lowest cost of utilization, namely, g. Therefore, in a sense, the integrated decision-making within the hospital induces the government to also treat its different policy targets in an integrated way and to use almost a unique policy instrument for all targets.

Therefore, in the bargaining scenario g_i^b is a less specialized and less accurate instrument curbed to the achievement of two main government's objectives. This may represent an explanation of the above described occurrence of higher informational rents under the bargaining scenario. In fact, as long as the cost of capacity r is low, K_i^c is used as a dedicated instrument to minimize the informational rent under the contract scenario, while g_i^c is exclusively targeted at the control of the optimal number of treated patients. Hence the fact that, in the contract scenario, there are two more precise, tailored and specialized instruments (instead of only one as in the bargaining scenario) explain the achievement of a lower informational rent under the contract scenario, when the cost of capacity is sufficiently low.

6. Conclusion

We have studied the impact of different hospital organizational structures on the optimal contracts designed by third party payers when severity is hospital's private information. Namely, we have compared a multi-issue bargaining process between doctors and managers within the hospital with a scenario where doctors and managers decide independently, with managers proposing to doctors a *take-it-or leave-it offer*. Results found are of interest to policy makers. Indeed, we have shown that, not only, the government instruments, settled within contracts signed with providers, must be tailored according to the hospitals' organizational structure and decision-making procedures, but also that their effectiveness in eliciting information differs across the different contractual arrangements within the hospital.

In particular, our results support the idea that designing radical organizational reforms within the hospitals must be based on thorough and articulated analysis of the induced policy implications. In fact, if on the one hand a scenario by which all the strategic variables are negotiated within the hospital between managers and doctors can lead to less opportunistic and less biased unilateral behaviour regarding the key financial and medical decisions (namely, the monetary incentives to doctors and the number of treated patients) on the other hand, however, it can also lead to further distortions in the government ultimate objectives, and, at least when the cost of capacity is sufficiently low, to less accurate and less effective policy tools aiming at reducing informational rents accruing to secondary care providers.

In particular, policy insights ultimately depend on the cost of capacity. If we interpret the cost of capacity as the cost of capital, since public and private providers are likely to face different capital markets' conditions, the policy implications of our results will depend on the type of providers serving the market. Since public providers normally benefit from a rate of return below the commercial rate faced by their private counterparts,[14] we would expect a contract scenario to be more effective as an organizational structure if the secondary care market is served by public providers.

[14]For example in England the commercial rate of return faced by private providers is 6.1% while the analogous for public providers is of 3.5% (Mason *et al* 2007).

References

Boadway R, M Marchand, Sato M. An Optimal Contract Approach to Hospital Financing. *Journal of Health Economics* 2004;23;85-110

Chalkley M, Malcomson JM. Contracting for Health Services with unmonitored Quality, *The Economic Journal*, 1998a, 108, 1093-1110.

Chalkley M, Malcomson JM. Contracting for Health Services when patient demand does not reflect quality, *Journal of Health Economics*, 1998b, 17, 1-19.

Crilly T, Le Grand J. The motivation and behaviour of hospital trusts. *Social Science and Medicine* 2004, 58, no. 10, pp. 1809-1823

Custer WS, Moser JS, Musacchio RA, Willke RJ. The production of health care services and changing hospital reimbursement. The role of hospital-medical staff relationships. *Journal of Health Economics* 1990; 9; 167—192

De Fraja G. Contracts for Health Care and Asymmetric Information, *Journal of Health Economics*, 2000, 19(5), 663-677.

Department of Health. *A guide to NHS Foundation Trusts*, Department of Health: London, 2002a. Also available at http://www.dh.gov.uk/assetRoot/04/06/04/80/04060480.pdf

Department of Health. Reforming NHS *Financial flow: introducing Payment by Result.* Department of Health: London, 2002b.

Department of Health. *NHS Foundation Trusts: A sourcebook for developing governance arrangements*. Department of Health: London, 2006. Also available at http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4130231

Dranove D, White WD. Recent Theory and Evidence on Competition in Hospital Markets, *Journal of Economics and Management Strategy* 1994, 3, 169-209.

Ellis RP, McGuire TG. Provider Behaviour under prospective reimbursement: cost sharing and supply. *Journal of Health Economics* 1986; 5 : 129-51.

Ellis RP. Creaming, skimping and dumping: provider competition on intensive and extensive margins. *Journal of Health Economics* 1998; 17 (5); 537-55.

Harris JE. The internal organization of hospitals: Some economic implications. *Bell Journal of Economics* 1977, 8, 467-48

Ma, AC. Health care payment systems: cost and quality incentives, *Journal of Economics and Management Strategy* 1994, 3(1), 93-112.

Mason A, Miraldo M, Siciliani L, Sivey P, Street A. *Establishing a Fair Playing Field for Payment by Results*. Report to the Department of Health, October 2007.

Miraldo M, Hospital financing. Master Thesis, DEA -- UCL, Louvain-la-Neuve: UCL, 2000

Mossialos E, Le Grand J. *Health Care and Cost Containment in the European Union*. Aldeshot , Brookfield, USA : Ashgate; 1999.

Mossialos E. *Funding health care: options for Europe*. Open University Press; Buckingham; 2002. European Observatory on Health Care Systems series.

Muthoo A. *Bargaining Theory with Applications*, Cambridge University Press, 1999.

OECD Health Data 2000: *A Comparative Analysis of 29 Countries*. Paris: Organisation for Economic Co-operation and Development.

Osborne M, Rubinstein A. *Bargaining and Markets*. Academic Press Inc, San Diego, 1990.

Pauly M. Medical Staff Characteristics and Hospital Costs. *The Journal of Human Resources*, 1978, Vol. 13, pp. 77-111

Rickman N, McGuire A. Providers' reimbursement in the health care market, *Scottish Journal of Political Economy,* 1999, 46, 1, 53-71.

Simoens S and Giuffrida A. The impact of physician payment methods on raising the efficiency of the healthcare system. *Applied Health Economics and Health Policy* 2004, 3, pp. 39-46

Appendix

A Strategic bargaining

A.1 Fees expressions

The exact expressions for the equilibrium fee-for-service proposed by managers and doctors in the bargaining are, respectively,

$$h_i^*(M) = \frac{3\beta\delta Z - 4H}{12(1+2(\alpha_i + g_i + K_i))}$$
$$h_i^*(D) = \frac{3\delta(\beta - 1)Z - 1 + 4Q}{12(1+2(\alpha_i + g_i + K_i))}$$

where

$$\begin{aligned} Z &= 1+4\left[g_i(1+g_i)+\alpha_i(1+\alpha_i)+K_i\right]+8\left[\alpha_i(g_i+K_i)+K_i(g_i-K_i)\right] \\ &\quad +24T_i \\ H &= 1+K_i+2K_i(\alpha_i+g_i-K_i)+4\left[\alpha_i(\alpha_i+1)+g_i(g_i+1)\right]+8\alpha_i g_i \\ Q &= \left[2(K_i-\alpha_i g_i)+4K_i(\alpha_i+g_i+K_i)+18T_i-g_i-\alpha_i(1-\alpha_i)\right]. \end{aligned} \tag{16}$$

In the limit case of perfectly patient agents, $\delta \to 1$, both $h^*(M)$ and $h^*(D)$ converge to the identical expression $\overline{h}_i^*$

$$h_i^*(M)_{\delta\to 1} \equiv h_i^*(D)_{\delta\to 1} \to \overline{h}_i^* = \frac{3\beta Z - 4H}{12(1+2(\alpha_i + g_i + K_i))},$$

where Z and H are as in (16).

B Contract

B.1 Single crossing conditions

Under perfect information the optimal menu of contracts is such that:

$$g_i^{c,pi} = \frac{\alpha_i + 1 - 2r(1+\lambda)}{2\lambda+3},\ K_i^{c,pi} = \frac{(2+\lambda)(\alpha_i+1)}{2\lambda+3} - r$$
$$n_i^{c,pi} = \frac{(2\alpha_i+1)(\lambda+2) - 2r(1+\lambda)}{2(2\lambda+3)},\ n_1^{c,pi} - n_2^{c,pi} = \frac{(\alpha_1-\alpha_2)(1+\lambda)}{2\lambda+3}$$

Let $v_{H_i} + T_i$ stand for the hospital indirect utility function, given managers and doctors solutions,

$$v_{H_i}(K_i, g_i, T_i; \alpha_i) \equiv T_i - \frac{(n_i^* - K_i)^2}{2} + Q_i(n_i^*, \alpha_i) + g_i n_i^* - \frac{(n_i^*)^2}{2}$$

Computing the total differential and the partial derivatives of $v_{H_i}(K_i, g_i, T_i; \alpha_i) + T_i$ it is possible to write the marginal rate of substitution between any pair of K_i , g_i and T_i . For instance, take g_i and T_i by the implicit function theorem $\frac{\partial\left(v_{H_i}+T_i\right)}{\partial T_i}dT_i+\frac{\partial\left(v_{H_i}+T_i\right)}{\partial g_i}dg_i=0$ then:

$$\frac{dT_i}{dg_i}=-\frac{\partial\left(v_{H_i}+T_i\right)/\partial g_i}{\partial\left(v_{H_i}+T_i\right)/\partial T_i} \tag{17}$$

Analogously:

$$\frac{dT_i}{dK_i}=-\frac{\partial\left(v_{H_i}+T_i\right)/\partial K_i}{\partial\left(v_{H_i}+T_i\right)/\partial T_i},\qquad \frac{dK_i}{dg_i}=-\frac{\partial\left(v_{H_i}+T_i\right)/\partial g_i}{\partial\left(v_{H_i}+T_i\right)/\partial K_i} \tag{18}$$

Computing the partial derivatives,

$$\partial\left(v_{H_i}+T_i\right)/\partial g_i=\frac{g_i+\alpha_i+2K_i+1/2}{4}$$

$$\partial\left(v_{H_i}+T_i\right)/\partial T_i=1$$

$$\partial\left(v_{H_i}+T_i\right)/\partial K_i=\frac{g_i+\alpha_i+1/2}{2}$$

Plugging these expressions into (17) and (18), and simplifying we find:

$$\frac{dT_i}{dg_i}=-\frac{g_i+\alpha_i+2K_i+1/2}{4},\quad \frac{dT_i}{dK_i}=-\frac{g_i+\alpha_i+1/2}{2}$$

$$\frac{dK_i}{dg_i}=-\frac{2g_i+4K_i+2\alpha_i+1}{4\left(g_i+1/2+\alpha_i\right)}$$

Differentiating with respect to α_i ,

$$\frac{d}{d\alpha_i}\frac{dT_i}{dK_i}\bigg|_{v_{H_i}+T_i}=-\frac{1}{2}\qquad \frac{d}{d\alpha_i}\frac{dT_i}{dg_i}\bigg|_{v_{H_i}+T_i}=-\frac{1}{4}$$

$$\frac{d}{d\alpha_i}\frac{dK_i}{dg_i}\bigg|_{v_{H_i}+T_i}=\frac{K_i}{\left(\alpha_i+1/2+2g_i\right)^2}>0$$

it turns out that the marginal rates of substitution are monotonic in α_i. Therefore, the Single Crossing Conditions are verified.

C Comparison: bargaining vs. contract

C.1 Rent comparison

Substituting the optimal contract for the bargaining scenario given by (10) and (11) and the optimal contract for the contract scenario as described in (14) and (15) we obtain the informational rents under both scenarios, R^b and R^c . Computing their difference:

$$R^b-R^c=\frac{\alpha_1-\alpha_2}{48}\left[r-\left(\alpha_2-\alpha_1\right)\frac{3\lambda\left(\lambda+4\right)+2}{12\left(\lambda+2\right)\left(1+\lambda\right)}\right]$$

As $\alpha_2>\alpha_1$ then $R^b\geq R^c$ if and only if $r\geq\left(\alpha_2-\alpha_1\right)\frac{3\lambda(\lambda+4)+2}{12(\lambda+2)(1+\lambda)}$.