

AN OPTIMAL CONTRACT APPROACH TO HOSPITAL FINANCING

Robin Boadway¹, Maurice Marchand² and Motohiro Sato³

January 2003

Abstract

Existing models of hospital financing advocate mixed schemes which include both lump-sum and cost-based payments. The doctor is generally the unique decision maker, which is unrealistic in a hospital setting where both managers and doctors are involved. This paper develops a model in which managers and doctors are responsible for different decisions within the hospital. In this model, public authorities who provide the financing, hospital managers who allocate resources within the hospital, and doctors who assign patients to either a low-tech or a high-tech therapy have information of increasing quality on the casemix of patients. The public authorities sign with hospital managers contracts specifying some lump-sum financing and some size of a high-tech equipment. In turn, managers, who know the broad mix of patients in the hospital, sign with hospital doctors contracts that specify the non-medical resources allocated to this facility as well as some remuneration. Doctors, who know each patient's illness severity, select the patients to be treated by the high-tech facility, and receive from public authorities some fee-for-service payment that is differentiated according to the low- or high-tech treatment used for curing their patients. What emerges is a two-stage agency problem in which contracts are designed to elicit information in the most efficient way.

JEL: D82, I18

⁰Queen's University. Email: boadwayr@qed.econ.queensu.ca

²CORE and IAG, Université catholique de Louvain. Email: mmarchand@core.ucl.ac.be

⁰Hitotsubashi University. Email: satom@econ.hit-u.ac.jp

This research was supported by the Prime Minister's Office, Science Policy Programming through a PAI-UAP research program and the research project entitled "Policy Instruments for Specialized Functions in Belgian Hospitals: Needs, Technology Assessment and Financing Mechanisms". Financial support from the Social Sciences and Humanities Research Council of Canada is also acknowledged. We are grateful to Jérôme Pouyet and the referees of the Journal for very helpful comments. We also thank I. Macho-Stadler, G. De Fraja, P. Gonzales, J.E. Askilden, P. Olivella and participants at seminars in Bergen and Barcelona. The scientific responsibility is assumed by the authors.

Corresponding author: Maurice Marchand, CORE, Voie du Roman Pays 34, 1348 Louvain-la-Neuve, Belgium. Tel.: 32-10-47 43 29; Fax: 32-10-47 43 01; email: mmarchand@core.ucl.ac.be

1 Introduction

The development of medical technologies has made the overutilization of expensive treatments a major concern for third-party payers. Their usage is largely determined by hospitals and their doctors. It can be controlled in part by influencing, through the financing system, the incentives of hospitals and doctors to assign patients to high-tech treatments.

This paper uses an optimal contract approach to investigate the design of hospital financing schemes when patients who differ in their illness severity can be treated by either a cheap low-tech therapy or an expensive high-tech one. In the setting developed in this paper, managers and doctors have specialized skills that are complementary, and they act as separate decision-makers within hospitals. Managers decide on the non-medical resources to make available to doctors for providing the expensive therapy, while doctors decide how many patients to treat with this therapy. For a given amount of non-medical resources, the higher this number of patients the greater the effort that doctors have to provide. To be realistic, doctors are assumed to have better information on their patients' illness severity than the manager. Moreover the manager and doctors do not share the same interests. The hospital manager aims, through his contract with doctors, to maximize the expected financial surplus of the hospital, while doctors are concerned with the quality of care they provide to patients and their own levels of effort and income. The main purpose of this paper is to see how this principal-agent relationship within the hospital affects the government's choice of the hospital financing mechanism.

The government is concerned with both the quality of care to patients and the level of public expenditures. We assume that health care is financed through taxes rather than voluntary insurance, and no co-payment is imposed on patients for hospital care. However, through its contracts with hospital managers, the government can restrain the use of the expensive treatment by controlling the equipment size of the high-tech facility. Such control is commonly used in numerous countries (like Belgium) to prevent the overutilization of some expensive treatments. In our setting, controlling equipment size does not, however, mean that the capacity for the high-tech treatment is under the full control of the government. That is because the amount of other non-medical resources (nurses, technicians, etc.) that are needed to operate the equipment is controlled by the hospital manager. Furthermore, doctors can adjust their effort to increase or decrease the number of patients treated in the high-tech facility. Before contracting with a hospital manager, the gov-

ernment has no information on the hospital's casemix of patients, that is, the distribution of their severity of illness. This is in contrast with the manager who has some information about his hospital's casemix, though less precise than that of the hospital's doctors.

The model therefore consists of a hierarchy of three levels of decision-makers – the government, the hospital manager, and the doctors – whose decisions are taken in that sequence. This gives rise to a two-tier hierarchy of principal-agent interactions. The top one involves the government as principal to the manager, and the second involves the manager as principal to the doctors. The doctors are treated as a single agent, which assumes that they behave cooperatively so that any conflicts of interest among them are suppressed. Information improves as one goes down the chain of decision-makers, so at each stage the principal is constrained by being less-well informed than the agent. Contracts at each stage are designed to extract information from the agent at the least cost to the principal. This gives rise to a potentially complicated two-stage game whose solution is a sub-game perfect Nash equilibrium. To make the problem manageable, we adopt as simple a setting as possible, focusing only on essential decisions. In particular, we abstract from the principal-agent relationship between the doctors and their patients by assuming that the latter are passive.

As mentioned above, a key feature of our setting is that the hospital manager does not share the government's objective function, nor the doctors that of the manager. This implies that the way in which the manager structures his menu of contracts with the doctors is not efficient from the point of view of the government. The government will therefore design its menu of contracts with the manager to extract information from the latter and also to correct for the inefficiency of the manager's contract with the doctors. However, in so doing, the government is restricted in its policy instruments: it can control hospital inputs only indirectly through the parameters of its contract with the manager.

We concentrate on the design of contracts for a single hospital even though it is implicit that there are a large number of such hospitals in the economy, each one catering to residents in its catchment area. This simplification allows us to focus on incentives arising from vertical interaction in the management of the hospital, thereby abstracting from horizontal interactions across hospitals, such as competition for patients or for doctors. In a contractual setting,

it also requires that the casemixes of hospitals are statistically independent.¹

Most of the recent literature on hospital financing has focused on the optimal mix of cost-based reimbursement and prospective payment, as illustrated by Ellis and McGuire (1986), Chalkley and Malcomson (1998) and McClellan (1997). This literature has not generally adopted an optimal contract approach such as those developed in Baron (1989) or Macho-Stadler and Pérez-Castrillo (1997), ch. 4. This is surprising in view of the asymmetries of information that characterize the hospital sector. It is true that diagnosis-related classification of patients has considerably reduced the asymmetries of information between third-party payers and hospitals. However, there remains some heterogeneity in the medical risks of patients classified in a given diagnosis-related group, and the motivation of the present paper comes from the fact that hospitals may differ in the illness severity of the patients they treat within a given diagnosis-related group. As a consequence, third-party payers as well as hospital managers may not share the same information as doctors regarding the medical risks of patients classified in a given diagnosis-related group. On the other hand, few papers in the literature have accounted for the interaction of doctors and manager as separate decision-makers within the hospitals. An exception is Custer et al (1990) who investigate how the move towards more prospective payment systems affects incentives within the hospital according to various kinds of cooperative and non-cooperative interactions between managers and doctors. The framework adopted in their paper differs from ours since it assumes full information at all levels.

The delegation of authority in a hierarchical organization as in this paper is commonly observed in practice. The literature discusses the pros and cons of delegation. Delegation can be beneficial for various reasons in environments involving transaction costs, incomplete contracts, lack of commitment or collusion.² First, it would reduce communication costs between the principal and his agents and enables the principal to spend his scarce time on more valuable tasks. Second, delegation could serve as a commitment device.³ Third, along

¹If they were dependent it would be possible to use the information revealed by the other hospitals' choice of contract to set the menu of contracts offered to an hospital.

²When these are absent, the revelation principle implies that a decentralized organizational form involving delegation cannot be superior to a centralized one in which all agents directly communicate and contract with a principal (Myerson (1982)).

³In the context of tax evasion, Melumad and Mookherjee (1989) claim that delegating the execution of ex ante announced enforcement policies to an agent can contribute to enhancing its credibility when the principal lacks commitment ability: the principal would be otherwise motivated to deviate from the policy ex post.

with limited communication or lack of monitoring, the possibility of collusion between an intermediate agent and an agent at the bottom of the hierarchy supports the case for delegation of authority to the former (Laffont and Martimort (1998)). At the same time, delegation involves costs. It may give rise to conflicts of interest when the agents who are given authority over some decisions pursue their own interests at the expense of that of the principal. In the presence of asymmetric information, the agents would try to exploit their informational rents.⁴ In this paper, we do not explore explicitly the optimal organizational form for hospital management, which is the concern of the above literature on delegation of authority. We take it as given. Since some important decision-making functions of hospital management are commonly assigned to the manager, we are concerned with how the contractual relationship should be designed between the government and the hospital, where the manager of the latter offers contracts to doctors.

The paper is organized as follows. In the next section, we explain the information setting underlying the model developed in the paper. We also explain in this section how patients can be treated and how the benefits they get vary with the low- or high-tech treatment provided and their illness severity. In Section 3, we present the nature of the contracts that are settled at each tier of the hierarchy. Section 4 investigates how doctors decide upon the number of patients to be treated in the high-tech facility according to the casemix of patients they observe and the contract they have signed with the hospital manager. In the next section, the way in which hospital managers design their contracts with doctors is studied. In particular we analyse how these contracts are affected by the upper-level contract that has been signed between the government and the hospital manager. In Section 6, we then turn to these upper-level contracts, and we explain how the asymmetry of information between managers and doctors within hospitals affects the design of the financing contracts adopted by the government. Finally, Section 7 is devoted to some conclusions.

2 The basic features of the hospital

As mentioned earlier, our model focuses on a single hospital whose number of patients is given. These patients differ in their severity of illness, denoted

⁴With limited ability to monitor agents' performance, contracts would fail to remove such incentives, a phenomenon known as 'double marginalization' (Melumad, Mookherjee and Reichelstein (1995)).

by s . The distribution of the patients' illness severities in a hospital, that is its casemix of patients, is assumed to be fully specified by its mean, denoted by θ . More precisely, illness severity s is uniformly distributed on the interval $[\theta - d, \theta + d]$ with the length of the interval representing the given number of patients to be treated. The number of patients with a given s is assumed to be unity, so the total number of patients in a hospital is $2d$. We assume that four types of hospitals can be distinguished according to their level of θ . This is the result of two successive random draws, as shown in Figure 1.

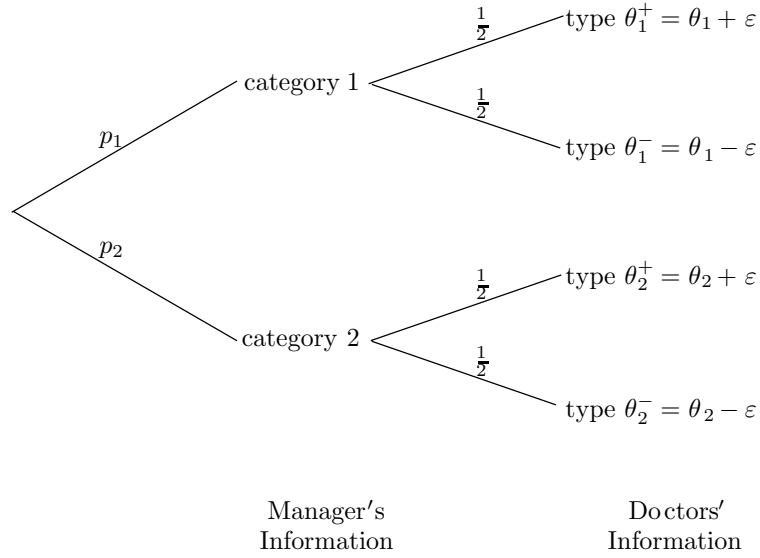


Figure 1: The Ex-ante Information Structure

The first of these draws divides hospitals into two broad categories, namely the low- and high-severity ones, which we refer to as categories 1 and 2 respectively. A proportion p_1 of hospitals are known to be of category 1; similarly, a proportion $p_2 = 1 - p_1$ are of category 2. In turn, as a result of the second draw, each of these broad categories of hospitals is divided with equal probabilities⁵ into two subcategories (or types) that differ in their patients' average illness severity θ . We call them types θ_1^+ and θ_1^- for category 1 and types θ_2^+ and θ_2^- for category 2. To fix ideas, let us posit that $\theta_i^- = \theta_i - \varepsilon$

⁵For the sake of simplicity we have adopted equal probabilities.

and $\theta_i^+ = \theta_i + \varepsilon$ ($i = 1, 2$) with $\theta_2 > \theta_1$ (category 2 is characterized by more severe illnesses than category 1). Therefore, θ_i (labeled without superscript) is also the average of the θ 's of category- i hospitals. It is worth stressing that rather than being fixed for ever the type of a hospital can be assumed to fluctuate from period to period. To be consistent with the above probabilities, a hospital of category i would then, on average, be of type θ_i^+ during half of the periods and of type θ_i^- during the remaining ones. In the following we shall sometimes resort to this interpretation of our setting.

As stated in the introduction, doctors and managers of a given hospital category have different information about their hospital type. Whereas doctors are assumed to know the true type of their hospital (θ_i^+ or θ_i^-),⁶ the manager has only imperfect information about it when designing the menu of contracts he will offer to doctors. He only knows the category i to which his hospital belongs. Of course, the manager of a category- i hospital is aware of the probability distribution of its type (either θ_i^+ or θ_i^- with equal probabilities). Moving up the hierarchy, the government is assumed to have no information on the category or type of the hospital it faces when designing its contract with the hospital's manager. It is only aware of the double-draw process through which a value for θ will be assigned to a hospital. Therefore, according to this nested information structure, the government knows that the information available to the manager is better than its own and that the doctors' information is still better than that of the manager. Thus, at each stage of the hierarchy, the principal is constrained by being less well-informed than the agent. This reflects differences in professional skills as well as the closeness of contact with the patients being served. On both accounts, the doctors are in a better position than [the manager] to estimate the average severity of the patients they treat. In turn, the government is one step further removed from the hospital setting. At each hierarchical level, the principal designs contracts – a prospective payment and a fee-for-service – so as to extract the private information possessed by the agent at the time the contract is signed.

⁶It may seem unrealistic to assume that doctors know the exact casemix of their patients at the time they contract with the manager. An alternative assumption is that the casemix that doctors in a type- θ hospital will face after having signed their contract is characterized by a mean illness severity given by $\theta + \delta$ where δ is a random white noise. In other words the illness severity of patients in a type- θ hospital is then distributed uniformly on support $[\theta + \delta - d, \theta + \delta + d]$ where δ is a random drawing from a white-noise distribution. The model can be modified accordingly, though at the expense of some intricacies in the presentation. For the sake of expositional simplicity, this feature is not included in the present paper. The more complicated version of the model can be requested from the authors.

The contracts are set sequentially, first between the government and the hospital manager (Stage 1), and then between the manager and the doctors (Stage 2). This timing is justified on the following grounds. As we shall see in the following, the contract in Stage 1 specifies the size of the high-tech equipment that the government allows the hospital to install. Since this equipment is a fixed input, Stage 1 involves long-run decisions. On the other hand, the contract between the manager and the doctors of a hospital in Stage 2 states the capacity for treating patients with the high-tech therapy that the manager will make available to doctors. This capacity results from combining non-medical resources with the high-tech equipment whose size is specified in the contract signed between the government and the hospital manager in Stage 1. Contrary to the fixed equipment, these medical resources are variable inputs that can be adjusted in the short or medium run. It is worth stressing that the above timing fits well the situation where the category of a hospital is fixed in the long run whereas its type within this category can fluctuate from period to period. In this case, a new Stage-2 contract would be signed between the manager and the doctors of a hospital every period.

Let us now turn to the treatment of patients. For ease of analysis, this is assumed to take a simple form. As mentioned in the Introduction, any given patient can be given either a high-tech treatment H or a low-tech one L . Let $q_H(s)$ and $q_L(s)$ be the benefit in terms of improved health status when a patient of severity s is given treatment H and L respectively. These functions are assumed to take on the following simple forms:

$$q_H(s) = a \quad \text{and} \quad q_L(s) = a - bs$$

where $a, b > 0$. Thus, the high-tech treatment H improves a patient's health status to the same degree irrespective of the severity of his illness s , while the improvement in health status when providing treatment L is declining as s becomes higher.⁷ Therefore, we have $q_H(s) - q_L(s) = bs$. This implies that a patient with a higher s should be treated with H , while he should be treated with L if his illness is not too severe. Doctors who by assumption know the illness severity of patients will assign them to treatments H or L . As shown in Figure 2, we can find a unique cutoff level denoted by $\hat{s}$ such that a patient with $s > \hat{s}$ is assigned to treatment H , while the remainder with less severe illnesses will be treated with L . Since the number of patients with a given s is taken to be unity and the distribution is uniform, the number of patients given

⁷Any other specification of $q_H(s)$ and $q_L(s)$ fits our setting as long as their difference is increasing and linear in s .

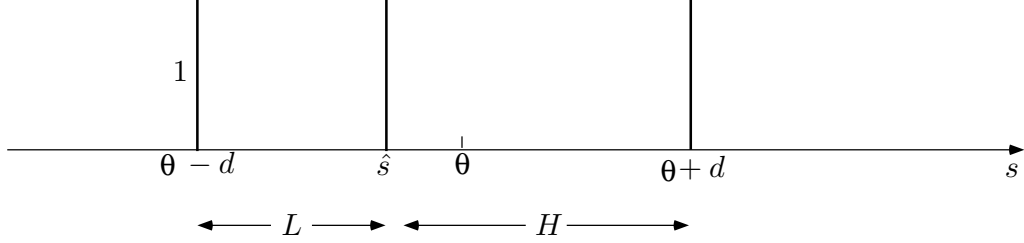


Figure 2: Severity of Illness and Choice of Treatment

treatment H is defined by $n \equiv \theta + d - \hat{s}$. Given n , the aggregate improvement in health status is given by:

$$\begin{aligned} Q(n, \theta) &= \int_{\theta-d}^{\theta+d-n} q_L(s) ds + \int_{\theta+d-n}^{\theta+d} q_H(s) ds \\ &= 2(a - b\theta)d + (\theta + d)bn - b\frac{n^2}{2}. \end{aligned} \quad (1)$$

Differentiation yields (recalling that the total number of patients in a hospital is $2d$):

$$Q_n = (\theta + d - n)b > 0 \quad \text{and} \quad Q_\theta = (n - 2d)b < 0. \quad (2)$$

The number of patients that can be given treatment H without further doctor effort is limited by a capacity constraint given by $C = \sqrt{2NK}$, where K is the size of the equipment and N is the amount of non-medical resources used to operate this equipment (e.g., nurses and technicians). This capacity can readily handle up to C patients. But for $n > C$, the doctors must exert increasing effort, as specified shortly. This implies that the doctors have some discretion over the number of patients assigned to treatment H . For the sake of simplicity, we ignore any further variable costs that might be incurred when applying the high-tech treatment. No such capacity constraint applies for the low-tech treatment. Again for the sake of simplicity, the cost of the latter is taken to be nil.

3 The nature of contracts

As mentioned, contracts are set in two stages – first between the government and the hospital manager, and then between the manager and the doctors. At

each stage a menu of contracts is offered by the principal, and one of these is chosen by the agent. The contracts offered in the first stage are based on the ex ante information possessed by the government and the manager, but anticipating the nature of contracts in the second stage and how they are affected by the first-stage outcome.

Each of the contracts that are proposed by the government to the hospital manager in the first stage specifies three components – a lump-sum or prospective payment T , a size of high-tech equipment K , and a fee-for-service g directly paid to doctors per patient treated with the high-tech therapy. The menu of contracts is designed to reveal the ex ante private information possessed by the manager, that is, his hospital category. As there are two hospital categories, this menu is made of two alternative contracts, (T_1, g_1, K_1) and (T_2, g_2, K_2) . These contracts cannot be based on the doctors' information θ since it has not yet been revealed to the manager. They are formulated such that the manager chooses the first contract if the hospital is of category 1, and the second one if it is of category 2. Obviously, some incentive constraints as well as participation constraints must be satisfied for this to be the case. Also, the effect that this contract has on the one that will be settled between the manager and the doctors in the second stage must be taken into account. Notice that although the fee-for-service g is directly paid to doctors, it is part of the contract chosen by the hospital manager in the menu of contracts offered by the government. In this sense, there is no direct contracting between the government and the doctors.

The terms of the government-manager contract in Stage 1 have not been made contingent on the manager-doctor contract implemented in Stage 2. The justification for doing so is that those terms cannot realistically be adjusted to the fluctuations of the hospital type. The size of the high-tech equipment (K) remains fixed in the short run while adjusting the fee-for-service (g) to these fluctuations would be difficult to implement by the government for administrative reasons.

Given the hospital category $i (= 1, 2)$, the menu of contracts offered by the manager to doctors in the second stage also consists of two alternative bundles, (P_i^+, h_i^+, C_i^+) and (P_i^-, h_i^-, C_i^-) , where the first element of each contract is a lump-sum remuneration paid to doctors, the second is a fee-for-service paid by the manager to doctors per patient treated with the high-tech therapy, and the third as we have seen is the capacity for high-tech treatment

provided by the manager. The values of these three components are set so as to induce the doctors to select the contract that reveals their information about θ . These values take into account both the contract that has already been chosen in the first stage and the behaviour of the doctors in assigning patients to the high-tech treatment. Given a hospital of category i , the first contract is intended for type $\theta_i^+ = \theta_i + \varepsilon$, and the second one is for $\theta_i^- = \theta_i - \varepsilon$.

Once the doctors have settled their contract with the hospital manager, they decide upon the number of patients, n , to be treated with the high-tech therapy. This is done in Stage 3.

To summarize, the order of decision making in our model is as follows:

- Stage 1: The government offers contracts (T_1, g_1, K_1) and (T_2, g_2, K_2) , and the manager selects one depending on the hospital category.
- Stage 2: Given (T_i, g_i, K_i) chosen by the manager, the manager offers contracts (P_i^+, h_i^+, C_i^+) and (P_i^-, h_i^-, C_i^-) . The doctors select one on the basis of their private information about θ .
- Stage 3: The doctors choose n .

We characterize the sub-game perfect Nash equilibrium by solving the above three stages by backward induction, starting with Stage 3.

4 Stage 3: the doctors assign patients to treatments

At this stage, the doctors will have signed a contract offering a remuneration P , a total fee-for-service $f(= h + g)$ and a size of capacity C . To simplify notation we omit the sub- and superscripts of these contract components even though the contract chosen by doctors is contingent on their information on θ . By some diagnosis, assumed to be costless for simplicity, doctors learn s for every patient and must decide how many patients n to treat on the high-tech facility. For $n \leq C$, giving patients treatment H is routine and takes no more effort than giving them treatment L . But, for $n > C$, the doctors must exert additional effort, which induces a disutility of $\beta(n - C)^2/2$, where $\beta > 0$. The doctors' utility function is then:

$$\begin{aligned} & Q(n, \theta) + (P + fn) && \text{if } n \leq C, \\ & Q(n, \theta) + (P + fn) - \beta(n - C)^2/2 && \text{if } n > C \end{aligned}$$

where $Q(n, \theta)$ is given by (1). The first term is the satisfaction that doctors obtain from improving the health of their patients, reflecting altruism, professional obligation, desire to enhance reputation, etc. The second term involving $P + fn$ is their financial compensation.⁸ As we have already mentioned, for any n doctors will maximize $Q(\cdot)$ by assigning those patients with the highest s to facility H .

The doctors' problem is then to maximize the above utility function with respect to n . Provided that $Q_n + f > 0$, n will exceed C and will satisfy the following first-order condition:

$$Q_n + f = \beta(n - C). \quad (3)$$

This is intuitive: since there is a benefit to doctors of increasing n , they will do so until it is offset by the disutility of effort from exceeding the capacity of facility H . For inequality $Q_n + f > 0$ to be satisfied for any type of hospital, the lowest value of θ (i.e. θ_1^-) must be large enough.⁹ Throughout the paper this condition is assumed to be verified. Using (2), the solution for n becomes:

$$n(C, f; \theta) = \frac{(\theta + d)b + \beta C + f}{b + \beta}. \quad (4)$$

Differentiating (4) yields:

$$\frac{\partial n}{\partial C} = \frac{\beta}{b + \beta} > 0; \quad \frac{\partial n}{\partial \theta} = \frac{b}{b + \beta} > 0; \quad \frac{\partial n}{\partial f} = \frac{1}{b + \beta} > 0 \quad \text{and} \quad \frac{\partial n}{\partial P} = 0. \quad (5)$$

It is worth noting that these derivatives are all independent of n . This property, which will simplify the formulation of the contract considerably, comes from the specification of the doctors' utility function.

The solution to the doctors' problem yields an indirect utility function that we denote as $V(C, f; \theta) + P$, where:

$$V(C, f; \theta) \equiv Q(n(C, f; \theta), \theta) + fn(C, f; \theta) - \frac{\beta(n(C, f; \theta) - C)^2}{2}. \quad (6)$$

Given the hospital type θ which is known to the doctors, (6) shows how the doctors' utility depends on C and f . Applying the envelope theorem yields:

$$\frac{\partial V}{\partial C} = \beta(n - C) > 0; \quad \frac{\partial V}{\partial f} = n > 0; \quad \frac{\partial V}{\partial \theta} = Q_\theta(n, \theta) < 0. \quad (7)$$

⁸Apart from the additional costs that must be incurred by the doctors to treat more than C patients on treatment H , there is a cost of treatment which is the same for all patients regardless of treatment. Since this is a constant, it is suppressed.

⁹See footnote 19 for a sufficient condition for $Q_n + f > 0$ to be satisfied.

This yields the following characterization of the doctors' marginal rate of substitution between C and P whose absolute value is the doctors' marginal willingness to pay for C :

$$\left. \frac{dP}{dC} \right|_{V+P} = -\beta(n - C) < 0; \quad \left. \frac{d}{dC} \frac{dP}{dC} \right|_{V+P} = \frac{b\beta}{b + \beta} > 0 \quad (8)$$

where we have used (5) to obtain the second result. This indicates that the doctors' indifference curves in (C, P) -space are downward-sloping and strictly convex, i.e. the doctor's marginal willingness to pay for C falls with C at a decreasing rate. They are also vertically parallel (with P on the vertical axis) because of the quasilinearity in P of the doctors' utility function. Differentiating their slope with respect to θ yields:

$$\left. \frac{d}{d\theta} \frac{dP}{dC} \right|_{V+P} = -\frac{b\beta}{b + \beta} < 0 \quad \text{for any } (C, P) \quad (9)$$

implying that the single-crossing property is satisfied. The larger is θ the higher is the marginal willingness to pay for C .

5 Stage 2: the manager offers contracts and the doctors choose

At this second stage, the hospital manager takes T_i, g_i and K_i as given since they were determined in the first stage in response to the manager's private information on his hospital category i . This information is known to the doctors who are also aware of the hospital's average casemix, that is, whether their hospital is of type θ_i^+ or θ_i^- .

The cost to the hospital of the non-medical staff that is needed to secure a given capacity for the high-technology facility can be inferred from the production function $C = \sqrt{2NK}$ introduced earlier. With w denoting the unit cost of the non-medical staff, this cost amounts to $w(2K)^{-1}C^2$. The two contracts proposed to doctors by the hospital manager specify values for C , P and h . One of these contract is intended for type θ_i^+ , the other for type θ_i^- .

The hospital's manager is assumed to care only about its *expected* financial surplus (or profit)¹⁰ that we denote by Π_i :

$$\Pi_i = T_i - S_i \quad (10)$$

¹⁰At the cost of some complication, the manager's objective function can be made a weighted sum of the expected financial surplus and the benefit to patients.

where S_i stands for the manager's expected overall spending (or cost):

$$S_i \equiv E_{\theta|i}[P] + \frac{w}{2K_i} E_{\theta|i}[C^2] + E_{\theta|i}[hn], \quad (11)$$

which is made of the doctors' lump-sum remuneration, the cost of the non-medical staff, and the fee-for-service paid by the manager to doctors. In the above expression, the expectation is taken with respect to θ for given i , where θ is either θ_i^+ or θ_i^- with equal probability.¹¹ To simplify the notation we omit subscript i in the rest of this section.

The manager's isocost curves in the (C, P) -space can be inferred from (11). Differentiating this expression yields the slope of these iso-cost curves whose absolute value is the marginal cost of C .

$$\left. \frac{dP^+}{dC^+} \right|_S = -\frac{w}{K} C^+ < 0; \quad \left. \frac{d}{dC^+} \frac{dP^+}{dC^+} \right|_S = -\frac{w}{K} < 0 \quad (12)$$

for $\theta = \theta^+$, and likewise for $\theta = \theta^-$. Thus, iso-cost curves are downward sloping and strictly concave to the origin, which means that the marginal cost of C rises in C at an increasing rate. Furthermore, they are vertically parallel.

We can now characterize the menu of contracts proposed to doctors by a manager of a given hospital category. However before doing so it will be helpful as a preliminary step to characterize the so-called 'full information' contracts. These would be chosen by the hospital manager if he were aware of the hospital type (either θ^+ or θ^-). In this case, for a given θ , the manager chooses (P, h, C) to minimize spending $P + wC^2/(2K) + hn$ subject to the doctors' participation constraint (for simplicity the doctors' reservation utility is taken equal to 0): $V(C, g; \theta) + P \geq 0$. Since the manager aims to minimize his payment to doctors, the participation constraint is always binding.

Assuming for now that $h^+ = h^- = 0$, the optimal solution to the manager's problem can easily be depicted graphically. In Figure 3, the doctors' indifference curves corresponding to $V(C, g; \theta^k) + P = 0, k \in \{+, -\}$ – that is the binding participation constraints – are drawn so as to fulfil the properties derived at the end of the previous section. In particular, at any value of C , the indifference curve related to θ^+ is higher and steeper than the one related to θ^- . On the other hand, as mentioned above the iso-cost curves of managers are vertically parallel and independent of θ .

¹¹For instance, $E_{\theta|i}[P] = (P_i^+ + P_i^-)/2$ with P_i^+ and P_i^- being the doctors' remuneration conditional on $\theta = \theta_i^+$ and $\theta = \theta_i^-$ respectively.

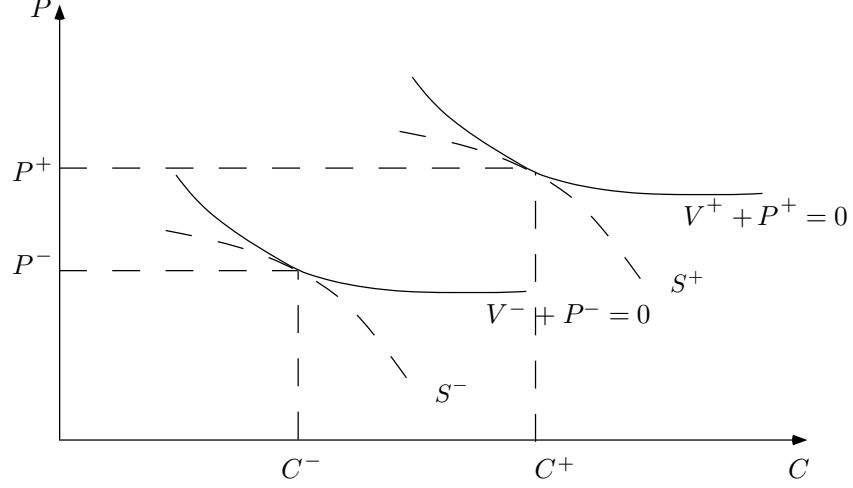


Figure 3: The Manager-Doctor Contract with Full Information

Accordingly, the optimal contracts are – as shown in Figure 3 – at tangency points between doctor indifference and manager iso-cost curves. It is easily seen from (8) and (12) that the values of the C 's appearing in these contracts satisfy:

$$\beta \left(n(C^k, f^k; \theta) - C^k \right) = \frac{w}{K} C^k, \quad k \in \{+, -\}, \quad (13)$$

which means that C is increased the level where the doctors' marginal willingness to pay for C is equated to its marginal cost. Starting from these optimal contracts with $h^+ = h^- = 0$, it can also easily be shown¹² that making these fees-for-service different from 0 will cause the managers' surplus to fall. In the 'full-information' case this would indeed entail some deadweight loss from the manager's perspective.

In the absence of full information, the solution just obtained and depicted in Figure 3 cannot be implemented. Doctors in a type- θ^- (less severe casemix) hospital would claim that their hospital is of type θ^+ (more severe casemix) in order to benefit from the more generous bundle (C^+, P^+) intended for this hospital type. In other words, the full-information contracts are not incentive

¹²To show this, substitute P from the doctor's participation constraint into the manager's objective. This yields: $-V(C, g+h; \theta) + wC^2/(2K) + hn$. Using (7), the first-order condition with respect to h yields at the optimum: $h = 0$.

compatible.

Let us therefore look for incentive-compatible contracts.¹³ To induce doctors to accept a contract and to reveal their information about the hospital type, two sorts of constraints must be satisfied. As above in the full-information case, there are first the participation – or individual rationality – constraints:

$$(IR) \quad V(C^k, h^k + g; \theta^k) + P^k \geq 0, \quad k \in \{+, -\}. \quad (14)$$

Notice that these IR constraints assume that doctors reveal their hospital type truthfully, which will be the case in equilibrium. In addition, truthful reporting requires that some incentive-compatibility constraints be satisfied. As suggested by Figure 4, the relevant one here is

$$(IC) \quad V(C^-, h^- + g; \theta^-) + P^- \geq V(C^+, h^+ + g; \theta^-) + P^+$$

where the right-hand side is the utility when doctors in a type- θ^- hospital choose the contract intended for type- θ^+ hospitals. One wants to avoid a low-severity hospital claiming that it is a high-severity one. This IC constraint may be rewritten as:

$$V(C^-, h^- + g; \theta^-) + P^- \geq V(C^+, h^+ + g; \theta^+) + P^+ + R(C^+, h^+ + g) \quad (15)$$

where

$$R(C^+, h^+ + g) \equiv V(C^+, h^+ + g; \theta^-) - V(C^+, h^+ + g; \theta^+) > 0, \quad (16)$$

the inequality coming from the last relation in (7) with $\theta^+ > \theta^-$.

Since it is always desirable for the manager to make the payments to doctors as low as possible, the participation constraint for type θ^+ and the

¹³Note that in our setting, the manager is limited to offering a linear contract consisting of a prospective payment P and a fee-for-service h that can be differentiated between different values of θ . In the present context, the doctors communicate directly with the manager regarding the realized value of θ . In a more general context, we can consider the situation where n serves as a signal used to condition a non-linear payment to the doctors. That is, the payment to the doctors can be $P(n)$, thus encompassing the fee-for-service h . Then our problem resembles a non-linear pricing model in which the manager effectively constrains the doctors' choice of n subject to an incentive constraint so as to induce truthful revelation. Needless to say, this non-linear form of contract dominates our linear one. We choose not to adopt this non-linear contract partly for analytical simplicity and partly to reflect hospital practice.

above IC constraint will be binding at the optimum:

$$(IR) \quad V(C^+, h^+ + g; \theta^+) + P^+ = 0, \quad (17)$$

$$(IC) \quad V(C^-, h^- + g; \theta^-) + P^- = R(C^+, h^+ + g). \quad (18)$$

Accordingly, $R(C^+, h^+ + g)$ can be interpreted as the *informational rent* accruing to doctors in a hospital of type θ^- . Since this rent is positive, the doctors' participation constraint in this type of hospital will not bind at the optimum. Also, given the single-crossing property in (9), the IC constraint for type- θ^+ will be slack at the optimum: $V(C^+, h^+ + g; \theta^+) > V(C^-, h^- + g; \theta^+)$.

In what follows, knowing how the manager's choice of C^+ and h^+ affects the information rent left to the doctors of a type- θ^- hospital will be crucial. Differentiating (16) and using (7) and (4) yields:

$$\frac{\partial R}{\partial C^+} = -\frac{\beta b}{\beta + b}(\theta^+ - \theta^-) < 0, \quad (19)$$

$$\frac{\partial R}{\partial h^+} = -\frac{b}{\beta + b}(\theta^+ - \theta^-) < 0. \quad (20)$$

The fact that the informational rent is decreasing in both variables will have important implications for the menu of contracts offered by the manager, which we now investigate.

The manager aims to minimize his expected expenditures (10) under the IR and IC constraints (17) and (18). By substituting P^+ and P^- from these constraints into (11) and recalling [that $\text{Prob}[\theta = \theta^+] = \text{Prob}[\theta = \theta^-] = 1/2$,] the manager's problem takes the following unconstrained form:

$$\begin{aligned} \min_{C^k, h^k} S \quad &\equiv \quad \frac{1}{2} \sum_{k \in \{+, -\}} \left[-V(C^k, h^k + g; \theta^k) + \frac{w}{2K}(C^k)^2 \right. \\ &\quad \left. + h^k n(C^k, h^k + g; \theta^k) \right] + \frac{1}{2} R(C^+, h^+ + g). \end{aligned} \quad (21)$$

The necessary conditions for a minimum of this expected expenditure function can be written as follows¹⁴:

$$\beta (n(C^-, g; \theta^-) - C^-) = \frac{w}{K} C^-, \quad (22)$$

¹⁴To derive these first-order conditions we make use of (5), (7), (19) and (20).

$$\beta (n(C^+, h^+ + g; \theta^+) - C^+) - \frac{\partial R}{\partial C^+} = \frac{w}{K}C^+ + h^+ \frac{\beta}{b + \beta}, \quad (23)$$

$$h^+ = b(\theta^+ - \theta^-) \quad \text{and} \quad h^- = 0, \quad (24)$$

Comparing (22) with (13) indicates that the manager's choice of C^- is made in the same way as in the full-information case (given g chosen by the government). It is also optimal for the manager to set $h^- = 0$. Of course, the doctors must be given a lump-sum remuneration P^- which includes the informational rent needed to induce them to choose the contract intended for type θ^- . Therefore, this remuneration is larger than if full information prevailed within the hospital.

The contract intended for hospitals of type θ^+ also differs from the one in the full information case. The optimal fee-for-service, h^+ , is now positive in order to reduce the information rent earned by hospitals of type θ^- . However, when h^+ has been set according to (24) the optimal rule for the choice of C^+ looks very similar to that used for the choice of C^- . This follows from substituting h^+ by its optimal value in (24) and using (19):

$$\beta (n(C^+, h^+ + g; \theta^+) - C^+) = \frac{w}{K}C^+. \quad (25)$$

This is a striking result: when choosing C^+ the manager can ignore the effect of C^+ on the information rent. This can be explained as follows. As shown by (19) and (20), C^+ and h^+ can both serve to reduce the information rent to doctors, and they are perfect substitutes for this purpose. However, while h^+ has no other role, manipulating C^+ has direct benefit and cost of its own (in (23), these are reflected in the first terms on the left- and right-hand sides respectively). This implies that the manager prefers to use only h^+ for lowering the information rent. Using C^+ for this purpose would entail in (25) some distortion from the standard rule that direct benefit be, at the margin, equated to direct cost. To correctly appraise this result it should however be kept in mind that the positive value of h^+ increases the number of patients treated in the high-tech facility and therefore the value of C^+ that satisfies (25). This is made clear by (26) below.

Accordingly, as depicted in Figure 4, optimal contracts are at tangency points between doctor indifference and manager iso-cost curves. And, to satisfy the IC constraint they are located on the same indifference curve belonging to doctors in hospitals of type θ^- .

The results are summarized in the following proposition.

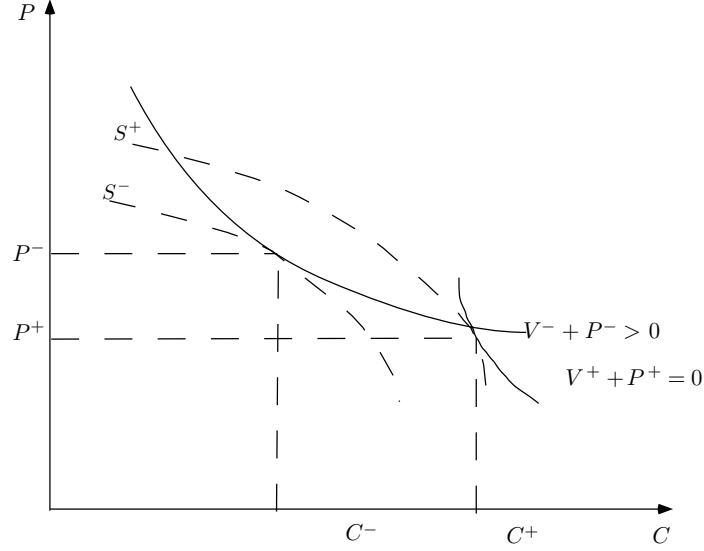


Figure 4: The Manager-Doctor Separating Contracts

Proposition 1 *In the contracts offered to doctors by the manager of a given hospital category, the capacities made available for providing the high-tech therapy are such that their marginal cost is equal to their marginal benefit. These contracts differ from the full-information ones only through the fee-for service paid by the manager to doctors: it is positive in the contract intended for the high-severity hospital and zero in the one intended for the other hospital type. This is motivated by the minimization of the informational rent left to doctors in the low-severity hospital.*

These contracts are obviously influenced by the government's instruments, K_i and g_i , that are targeted to hospitals of category i and are chosen in Stage 1 (see the next section). While the fees-for-service within the hospital, $h_i^- = 0$ and $h_i^+ = b(\theta_i^+ - \theta_i^-)$, are independent of K_i and g_i , the C 's depend upon them. Using (4) it is possible to derive from (22) and (25) explicit expressions for the C 's:

$$C_i^+ = \frac{\beta}{\beta b + (w/K_i)(b + \beta)}((\theta_i^+ + d)b + g_i + b(\theta_i^+ - \theta_i^-)) \quad (26)$$

and

$$C_i^- = \frac{\beta}{\beta b + (w/K_i)(b + \beta)}((\theta_i^- + d)b + g_i). \quad (27)$$

Thus, both C_i^+ and C_i^- are increasing in K_i and g_i .

Using these results, we can also derive the *expected* profit function of the manager:

$$\Pi(T_i, K_i, g_i; i) = T_i - S(K_i, g_i; i) \quad (28)$$

where $S(K_i, g_i; i)$ is here the expected spending given in (11) evaluated at the manager's optimal policy. Applying the envelope theorem yields:

$$\frac{\partial S_i}{\partial K_i} = -\frac{w}{2K_i^2} E_{\theta|i}[C_i^2]. \quad (29)$$

For use in the following section it will be helpful to investigate the properties of the manager's iso-profit curves in the (K, T) space. Differentiating (28) and using relations (29) yields:

$$\left. \frac{dT_i}{dK_i} \right|_{\Pi} = -\frac{w}{2K_i^2} E_{\theta|i}[C_i^2] < 0 \quad (30)$$

whose absolute value is the hospital manager's marginal willingness to pay for K in a category- i hospital. It is straightforward to show that¹⁵:

$$\left. \frac{d}{dK_i} \frac{dT_i}{dK_i} \right|_{\Pi} < 0 \quad \text{and} \quad \left. \frac{d}{d\theta_i} \frac{dT_i}{dK_i} \right|_{\Pi} < 0. \quad (31)$$

From the above inequalities we conclude that the iso-profit curves in (K, T) space are decreasing and satisfy the single-crossing property. The absolute value of their slope is rising as θ_i increases. The manager of a hospital with more severe casemixes has a higher willingness to pay for K than one with less severe casemixes. Furthermore, the iso-profit curves are vertically parallel.

6 Stage 1: the government offers contracts and the manager chooses one

The government aims at maximizing the expected social surplus generated by the hospital sector. This is defined as the social benefits of treatment to patients minus the social costs of the hospital sector. Let r be the per-unit cost of equipment K (assumed to be directly financed by public revenues) and λ the marginal cost of public funds ($\lambda > 1$). The government's objective is:

¹⁵To prove the first inequality we use (26) and (27), and to prove the second one we notice again from (26) and (27) that C_i^+ and C_i^- increases with θ_i^+ and θ_i^- , and so with θ_i (since $\theta_i^+ = \theta_i + \varepsilon$ and $\theta_i^- = \theta_i - \varepsilon$).

$$W = \sum_{i=1,2} p_i \left\{ E_{\theta/i} \left[(1 + \mu)Q(n, \theta) - \frac{\beta}{2} E_{\alpha/\theta} (n - C)^2 - \frac{w}{2K_i} C^2 \right] - rK_i \right\} - (\lambda - 1) \sum_{i=1,2} p_i (rK_i + T_i + g_i E_{\theta/i} n) \quad (32)$$

where, to recall, p_i is the probability that a hospital is of category i ($i = 1, 2$). The expression in the braces accounts for the social benefits coming from the patients' improved health status net of the private costs of resources needed to secure these benefits (including the doctors' effort). We assume that the government attaches a higher monetary value to the patients' benefits than doctors: in the above expression Q is weighted by $(1 + \mu)$ with $\mu > 0$. The second term appearing in W is the difference between the social and private costs of resources financed by the government. It is equal to the government's expenditures in the hospital sector, weighted by $(\lambda - 1)$, that is, the dead-weight loss per unit of tax revenue, due to the allocative inefficiencies caused by the tax system.

The government sets a menu of contracts, (T_1, g_1, K_1) and (T_2, g_2, K_2) , that are targeted to hospitals of categories 1 and 2 respectively. As with the contracts between the manager and the doctors (stage 2), the government policy that would be optimal in case of full information cannot be implemented when the government is not aware of the hospital's category. This is because a hospital with the less severe expected casemixes (category 1) would claim to be a hospital with the more severe ones (category 2). Therefore the government must design its menu of contracts so as to induce the manager to reveal the category of his hospital. Formally this involves choosing the menu of contracts to maximize the government's above objective function subject to the following participation and incentive-compatibility constraints applying to the manager:¹⁶

$$(IR) \quad T_2 - S(K_2, g_2; 2) \geq 0,$$

$$(IC) \quad T_1 - S(K_1, g_1; 1) \geq T_2 - S(K_2, g_2; 1),$$

¹⁶One could object that a mimicking hospital can be detected by observing the number of patients treated with the high-tech therapy (n). Once the casemix of a given hospital type is ex ante uncertain (as in the version of the model referred to in footnote 6), it is the ex post revealed casemix of the hospital that determines the choice of n by doctors. Accordingly neither the type nor the category of an hospital can be inferred from the observation of n . In particular it is not feasible to punish a mimicking hospital on the basis of this ex post information.

which can be shown by the single crossing property (31) to be the sole *IR* and *IC* constraints that are binding by analogy to the earlier manager-doctors problem in Stage 2. Once the above *IC* constraint is satisfied, the participation constraint will only be binding for hospitals of category 2. Note that in this category of hospitals, managers obtain zero expected surplus, implying that ex post surplus can be negative. This is an implication of managers being risk-neutral, the presumption being that over time negative surpluses cancel with positive ones.

As the above constraints are both binding at the optimum, they can be rewritten as:

$$T_2 = S(K_2, g_2; 2), \quad (33)$$

$$T_1 = S(K_1, g_1; 1) + \psi(K_2, g_2), \quad (34)$$

where

$$\psi(K_2, g_2) = S(K_2, g_2; 2) - S(K_2, g_2; 1)$$

can be interpreted as a (positive) informational rent, accruing to managers of category-1 hospitals. Substituting these expressions for T_1 and T_2 in the government's objective function (32) and using relations obtained earlier¹⁷ yields the following unconstrained problem:

$$\begin{aligned} \max_{g_i, K_i} W = & \sum_{i=1,2} p_i \left[\mu E_{\theta/i} Q(n(C, f, \theta), \theta) + \frac{1}{2} R(C_i^+, h_i^+ + g_i) \right. \\ & \left. - \lambda(rK_i + S(K_i, g_i, i) + g_i E_{\theta/i} n(C, f, \theta)) \right] - p_1(\lambda - 1)\psi(K_2, g_2). \end{aligned} \quad (35)$$

Together with the T 's that can be inferred from the *IR* and *IC* constraints (33) and (34) once the $[g$'s] and K 's are known, the solution to this problem provides the menu of contracts offered by the government to hospital managers: $\{(T_i, K_i, g_i), i = 1, 2\}$. When designing this menu of contracts, the government has to keep in mind that its policy instruments, K_i and g_i ($i = 1, 2$), affect the choice by managers of capacities C_i^+ and C_i^- through (26) and (27).

¹⁷We proceed as follows. We first substitute $E_{\theta/i}[Q(n, \theta) - \beta(n - c)^2/2]$ by its expression drawn from (6): $W = \sum_{i=1,2} p_i \left\{ E_{\theta/i}(\mu Q(n, \theta) - fn) + E_{\theta/i} \left(V - \frac{w}{2K_i} C^2 \right) \right\} - \lambda \sum_{i=1,2} p_i rK_i - (\lambda - 1) \sum_i p_i (E_{\theta/i} S + g_i E_{\theta/i} n) - p_1(\lambda - 1)\psi$. The V 's in this expression can then be eliminated by using *IR* and *IC* constraints (17) and (18). After using (11) to eliminate the P 's in the expression so obtained, we obtain (35).

6.1 Optimal contract for the type-1 hospital

Let us first focus on the optimal contract intended for hospitals of category 1. In the appendix we show that it must satisfy the following necessary conditions:

$$\frac{1}{2} \left(\mu Q_n(n, \theta_1^+) - h^+ - \lambda g_1 \right) + \frac{1}{2} \left(\mu Q_n(n, \theta_1^-) - \lambda g_1 \right) = 0 \quad (36)$$

and

$$\begin{aligned} \lambda \left[\frac{w}{2K_1^2} E_{\theta/1} C^2 - r \right] = & - \left[\frac{1}{2} \left(\mu Q_n(n, \theta_1^+) - h^+ - \lambda g_1 \right) \frac{dn}{dC} \frac{dC_1^+}{dK_1} \right. \\ & \left. + \frac{1}{2} \left(\mu Q_n(n, \theta_1^-) - \lambda g_1 \right) \frac{dn}{dC} \frac{dC_1^-}{dK_1} \right]. \end{aligned} \quad (37)$$

In the second of these first-order conditions, the effect of the high-tech capacity C on the number of patients treated with the expensive treatment is accounted for through dn/dC , appearing on the right-hand side. As shown in (5), this effect is in our setting constant and independent of the hospital type (θ) and the fee-for-service ($dn/dC = \beta(\beta + b)^{-1}$).

To interpret the above conditions as well as those on g_2 and K_2 later on, some preliminary considerations are in order. The first consideration is that the doctors and the manager do not share the objective of the government. This implies that they make decisions that are inefficient from the government's viewpoint, so the government will choose its policy instruments so as to reduce those inefficiencies. Second, given the hierarchical nature of contracts, the government cannot sign a contract with the doctors directly: the manager is the only agent with which the government can deal. This implies that while the government can extract the private information possessed by the manager on the hospital category by offering separating contracts, it cannot access the doctors' information on θ . Therefore, if the government wants to influence the doctors' choice of n by its policy instruments, it can only do so in a blind way since it cannot distinguish hospitals of types θ_i^+ and θ_i^- within category i .

These considerations explain the choice of g_1 in (36). This can be written as:

$$g_1 = \frac{1}{\lambda} \left[\mu E_{\theta/1} Q_n(n, \theta) - \frac{h^+}{2} \right], \quad (38)$$

which shows that g_1 is set to correct for two inefficiencies in the behaviour of managers and doctors. On the one hand, the government attaches a higher weight to improvement in health status than the doctors (this weight is higher by $\mu > 0$), which calls for making g_1 positive. However, this consideration is mitigated by the fact that in order to reduce the informational rent left to doctors in type- θ_1^- hospitals, the manager encourages the use of the high-tech therapy in type- θ_1^+ hospitals by setting h_1^+ positive. This is inefficient from the government's viewpoint, which calls for making g_1 negative. Therefore, the sign of g_1 is ambiguous. The higher is μ the more likely is g_1 to be positive. But g_1 might be negative.^{18,19}

The optimal choice of K_1 must satisfy condition (37). According to this condition, the standard rule that an input be used up to the level where its marginal productivity equates its cost (here, $w(2K_1^2)^{-1}E_{\theta/1}C^2 = r$) will not generally be fulfilled. As just explained for g_1 , this discrepancy comes from the government's desire to reduce the inefficiencies characterizing the decisions taken by hospital managers. In this respect, g_1 and K_1 are not perfect substitutes. While one infers from (5) that a change in g_1 affects n uniformly and independently of the hospital type, a change in K_1 does not. This is because a rise of K_1 makes hospitals managers increase C_1^+ more than C_1^- : from (26) and (27), $dC_1^+/dK_1 > dC_1^-/dK_1$, which in turn affects n differently according to whether the hospital is of type θ_1^+ and θ_1^- . We know from (36) that the average of the multiplicative factors of dC_1^+/dK_1 and dC_1^-/dK_1 in (37) is zero. Therefore, we conclude the following:

$$\frac{w}{2K_1^2}E_{\theta/1}C^2 > r \quad \text{iff} \quad \mu[Q_n(n, \theta_1^+) - Q_n(n, \theta_1^-)] < h^+. \quad (39)$$

In the appendix we show that the second inequality relationship is equivalent to:

$$2\mu\beta < (1 + \mu)[b\beta K_1/w + (b + \beta)], \quad (40)$$

¹⁸It may look unrealistic to have a negative fee-for-service. The reader should however keep in mind that for the sake of simplicity we have ignored some operating costs of hospitals. If some additional costs proportional to n (e.g. medical material) are incurred these should be included in g_1 . Then, having $g_1 < 0$ means that only part of these additional costs ought to be reimbursed.

¹⁹In Section 4, we have assumed that the lowest value of θ (i.e. θ_1^-) is large enough for $n > C$ to hold for any θ . Here we provide a sufficient condition for this to be satisfied. For a type- θ_1^- hospital, we have from (3) and (38): $\beta(n - C_1^-) = Q_n + (\mu/\lambda)E_{\theta/1}Q_n - h^+/2\lambda > (1 + (\mu/\lambda))Q_n - h^+/2\lambda$. The inequality relation comes from (2): the lowest value of Q_n amongst hospitals of category 1 is achieved for type- θ_1^- hospital. Substituting Q_n from (2) and h^+ from (24), we obtain: $\beta(n - C_1^-) > (1 + (\mu/\lambda))(\theta_1^- - d)b - b(\theta^+ - \theta^-)/2\lambda$. Therefore, a sufficient condition for $n > C_1^-$ is $\theta_1^- > d + (\theta^+ - \theta^-)/2(\lambda + \mu)$.

which holds if μ is not too high.

It is worth comparing this result with the one that would prevail if the manager and the doctors had the same information about the hospital type. In this case, h^+ is deleted from the right-hand side of (38). As shown in the appendix, $\mu Q_n(n, \theta_1^+) - \lambda g_1 > 0$ and so $\mu Q_n(n, \theta_1^-) - \lambda g_1 < 0$. Therefore, one infers from (38) with h^+ deleted that K_1 is always overprovided by the government with full information within the hospital. Since high-tech treatments are underprovided in type- θ_1^+ hospitals (i.e. $\mu Q_n(n, \theta_1^+) > \lambda g_1$) and, consequently, overprovided in type- θ_1^- hospitals (i.e. $\mu Q_n(n, \theta_1^-) < \lambda g_1$), this result is easy to interpret: an expansion of K_1 improves the underprovision more than it worsens the overprovision. Thus, the government would find it beneficial to raise the provision of K_1 to a level greater than what the standard rule would impose.

The above results on the choice of g_1 and K_1 enable us to state the following proposition.

Proposition 2 *In the contract intended for the less severe hospital category if μ is low enough, the government sets the fee-for-service for the high-tech treatment negative and underinvests in the high-tech equipment relative to the standard rule that the marginal productivity of an input be equated to its market price. The opposite holds if μ is high.*

Therefore the standard no-distortion-at-the-top rule does not hold. This is because the government wants to correct for the inefficiencies described above.

6.2 Optimal contract for the type-2 hospital

Let us now turn to the contract intended for the more severe hospital category. To begin with, it is worth mentioning that if the government knew which hospitals were of categories 1 and 2 (so there was no asymmetry of information between hospital managers and the government), first-order conditions similar to (36) and (37) would also prevail for hospitals of category 2. With information asymmetry, there are however additional terms in these conditions, that reflect the government's desire to extract information from hospital managers. To show this, we now turn to the government's choice of g_2 and K_2 appearing in the contract intended for hospitals of category 2.

The first-order conditions that g_2 and K_2 ought to satisfy are similar to the ones for g_1 and K_1 except that they account for the impact of g_2 and K_2 on the informational rent reaped by hospitals of category 1. It is shown in the appendix that these conditions can be written as follows:

$$g_2 = \frac{1}{\lambda} \left[\mu E_{\theta/2} Q_n(n, \theta) - \frac{h^+}{2} \right] - \frac{1}{\lambda} (b + \beta) \left[1 + \beta \frac{dC_2}{dg_2} \right]^{-1} \frac{p_1(\lambda - 1)}{p_2} \frac{d\psi}{dg_2} \quad (41)$$

where

$$\frac{d\psi}{dg_2} = - \left(E_{\theta/2}[n] - E_{\theta/1}[\hat{n}] \right) < 0, \quad (42)$$

and

$$\begin{aligned} \frac{w}{2K_2^2} E_{\theta/2} C^2 - r = & -\frac{1}{\lambda} \left[\frac{1}{2} \left(\mu Q_n(n, \theta_2^+) - h^+ - \lambda g_2 \right) \frac{dn}{dC} \frac{dC_2^+}{dK_2} \right. \\ & \left. + \frac{1}{2} \left(\mu Q_n(n, \theta_2^-) - \lambda g_2 \right) \frac{dn}{dC} \frac{dC_2^-}{dK_2} \right] + \frac{1}{\lambda} \frac{p_1(\lambda - 1)}{p_2} \frac{d\psi}{dK_2} \end{aligned} \quad (43)$$

where

$$\frac{d\psi}{dK_2} = -\frac{w}{2K_2^2} (E_{\theta/2}(C^2) - E_{\theta/1}(\hat{C}^2)) < 0. \quad (44)$$

In condition (41), the last term on the right-hand side reflects the impact of g_2 on the information rent, $\psi(K_2, g_2)$, and is positive. That impact is given by (42), in which $E_{\theta/1}[\hat{n}]$ stands for the expected number of patients who would be treated by the high-tech therapy in a category-1 hospital whose manager signs the contract intended for a category-2 hospital. Therefore, this mimicking hospital only differs from a category-2 one by its less severe casemixes of patients. The following inequality then holds: $E_{\theta/2}[n] > E_{\theta/1}[\hat{n}]$. This is because a heavier casemix increases the number of patients $n(C; f; \theta)$ both directly and indirectly through the larger capacity of the high-tech facility provided by managers. The above inequality implies that an increase in g_2 benefits a category-2 hospital more than a category-1 one that mimics a category-2 hospital. Therefore, increasing g_2 serves to relax the incentive compatibility constraint and so reduce the informational rent, $\psi(K_2, g_2)$.

In Appendix 3 we show that condition (41) can also be written as:²⁰

$$\mu E_{\theta/2} Q_n - \frac{h^+}{2} - \lambda g_2 + B = 0. \quad (45)$$

or

$$g_2 = \frac{1}{\lambda} \left[\mu E_{\theta/2} Q_n - \frac{h^+}{2} + B \right] \quad (46)$$

where $B \equiv (p_1/p_2)(\lambda-1)b\Delta\theta > 0$ with $\Delta\theta = \theta_2 - \theta_1 = \theta_2^k - \theta_1^k > 0, k \in (+, -)$. This condition makes clear that in choosing g_2 , the government trades off the considerations influencing the choice of g_1 in (36) and the minimization of the informational rent.

Likewise, the condition on K_2 in (43) differs from the one on K_1 in (37) by the impact of K_2 on the informational rent, $\psi(K_2, g_2)$. This impact is given by (44), in which $\hat{C}$ is equal to either $\hat{C}_1^+$ or $\hat{C}_1^-$ with equal probabilities and stands for the capacity that would be chosen by the manager of a category-1 hospital if he accepted the contract intended for a category-2 hospitals. For the same reasons as above, the following inequality holds: $E_{\theta/2}[C^2] > E_{\theta/1}[\hat{C}^2]$, which implies that the last term on the right-hand side of (43) pushes K_2 upwards. Since the sign of the first term in (43) is ambiguous, it is difficult without further analysis to infer from this condition whether K_2 is under- or overprovided with respect to the standard rule. However in Appendix 3 we show that this condition can be expressed as follows:

$$\begin{aligned} \frac{w}{2K_2^2} E_{\theta/2} C^2 - r = & -\frac{1}{\lambda} \left\{ \frac{1}{2} \left(\mu Q_n(n, \theta_2^+) - h^+ - \lambda g_2 + B \right) \frac{dn}{dC} \frac{dC_2^+}{dK_2} \right. \\ & \left. + \frac{1}{2} \left(\mu Q_n(n, \theta_2^-) - \lambda g_2 + B \right) \frac{dn}{dC} \frac{dC_2^-}{dK_2} \right\} \\ & + \frac{1}{\lambda} B \frac{w}{2K_2^2} \beta^2 b \left(\beta b + \frac{w}{K_2} (b + \beta) \right)^{-2} \Delta\theta \end{aligned} \quad (47)$$

Therefore, in the same way as in (38) for K_1 , we infer from (45) that the multiplicative factors of dC_2^+/dK_2 and dC_2^-/dK_2 in condition (47) average to zero, and we show in Appendix 3 that the expression in curly brackets in this condition is negative if and only if

$$2\mu\beta < (1 + \mu)[\beta b K_2/w + (b + \beta)]. \quad (48)$$

²⁰Note also that according to (26) and (27), the impact of g_2 on C_2^k is independent of $k \in (+, -)$ and positive: $dC_2^+/dg_2 = dC_2^-/dg_2 \equiv dC_2/dg_2 > 0$.

The following conclusion then follows: if inequality (48) holds, the government ought to underprovide K_2 relative to the standard rule; otherwise, the sign of the right-hand side of (47) is ambiguous. It would also be ambiguous if there were full information within the hospital. In this case, h^+ has to be deleted from (47), and the expression in curly brackets is always positive. We summarize our results in the following proposition.

Proposition 3 *In the contract intended for the more-severe hospital, the asymmetry of information between the government and the manager makes the government increase both the fee-for-service for the high-tech treatment and the investment in the high-tech equipment, which is motivated by the minimization of the informational rent left to the less-severe hospital. Otherwise, the design of this contract is similar to that intended to the other category of hospital.*

7 Conclusions

In this paper, we have investigated how to finance hospitals in a model involving a hierarchy of three decision-makers: the government, the hospital manager, and the doctors. Principal-agent interactions arise at each level of the hierarchical structure, because the three decision-makers do not share the same objective and the same information. The nested information structure assumed in this paper reflects differences in professional skills and experiences as well as closeness of contacts with patients. The doctors located at the bottom of the hierarchical structure are in a better position than the hospital manager to know the illness severity of patients, and the government located at the top of the structure is further removed from this information. The nested information structure (see Figure 1) is such that at the start, the government has no information on the casemix of the hospital, the manager knows its severity category ($i = 1, 2$) and the doctors are further aware of its casemix (within category i , $\theta = \theta_i^+, \theta_i^-$).

When offering contracts to doctors, the manager is motivated to maximize his own expected payoff by inducing the doctors to reveal the private information they have at the contract stage. As is usual in an adverse-selection setting, some rent needs to be left to doctors facing the least severe casemix in order that they do not pretend to face the most severe one. Making this rent as low as possible explains why it is optimal for managers to pay some fee-for-service for the high-tech treatments provided by doctors facing the most severe casemix. This is the only departure from the optimal contracts that

would prevail if managers had the same information as doctors on the hospital casemixes. In particular, whatever the casemix the high-tech capacity made available to doctors by the manager satisfies the standard rule that its marginal benefit be equated to its marginal cost.

Turning to the design of the contracts offered to the hospitals by the government, two inefficiencies occurring at the hospital level need to be corrected. First, if the government is more concerned than doctors by the benefits of treatment to patients, it will increase the size of the high-tech equipment invested in the hospital and the fee-for-service paid to doctors for the high-tech treatment above the levels that would be optimal otherwise. Second, the fee-for-service paid by the manager to doctors facing the most severe casemix is seen by the government as unduly encouraging the utilisation of the high-tech therapy. This induces the government to reduce its fee-for-service for this therapy. These considerations would be present even if the government and the manager shared the same information about the severity category of the hospital. However, because of the asymmetric information the contracts offered by the government must also be designed to make the manager reveal his hospital's category. For the same reason as earlier some rent must be left to the hospital with the most severe casemixes. In order to reduce this rent, the contract intended for the hospital category with the most severe casemixes specifies a high-tech equipment size larger than required by the standard rule as well as a fee-for-service per high-tech therapy higher than if there were no asymmetry of information between the government and the manager.

These results hold in the setting that has been adopted in this paper. It is certainly desirable to extend the model to other institutional settings. In particular, the within-hospital contractual framework could be modified by replacing the principal-agent relationship by a bargaining process between the manager and the doctors. It is indeed often the case that doctors bargain with managers about their remuneration structure and the resources they are provided within their medical services. For the sake of simplicity, several assumptions were made in this paper. Some could be relaxed without modifying our qualitative results. It is not however the case for the absence of competition for patients amongst hospitals. As our analysis in the present paper has shown it is for treating the less severe cases that hospitals collect some informational rent. This should induce hospitals to be more aggressive in attracting such cases, which amounts to some sort of risk selection.

References

- Baron, D.P. Design of regulatory mechanisms and institutions. Chapter 24 in Schmalensee, R. and R.D. Willig (eds.), *Handbook of Industrial Organization*, North-Holland, 1989.
- Chalkley, M. and J. M. Malcomson. Contracting for health services when patient demand does not reflect quality. *Journal of Health Economics*, 17, 1–19, 1998.
- Custer, W.S., Moser, J.W., Musacchio, R.A., and R.J. Willke. The production of health care services and changing hospital reimbursement. The role of hospital-medical staff relationships. *Journal of Health Economics*, 9, 167–192, 1990.
- Ellis, R.P. and T.G. McGuire. Provider behavior under prospective reimbursement. *Journal of Health Economics*, 5, 129–151, 1986.
- Laffont, J.J. and D. Martimort. Collusion and delegation. *Rand Journal of Economics*, 29, 280–305, 1998.
- Macho-Stadler, I. and D. Pérez-Castrillo. *An introduction to the theory of information: incentives and contracts*. Oxford, Oxford University Press, 1997.
- McClellan, M. Hospital reimbursement incentives: An empirical analysis. *Journal of Economics and Management Strategy*, 6(1), 91–128, 1997.
- Melumad, N.D. and D. Mookherjee. Delegation as commitment: the case of income tax audit. *Rand Journal of Economics*, 20(2), 139–163, 1989.
- Melumad, N.D., D. Mookherjee and S. Reichelstein. Hierarchical decentralization of incentive contracts. *Rand Journal of Economics*, 26(4), 654–672, 1995.
- Myerson, R.B. Optimal coordination mechanisms in generalized principal-agent problems. *Journal of Mathematical Economics*, 10, 67–81, 1982.

Appendix 1: Government optimization

The optimal contracts offered by the government to hospitals must satisfy the first-order conditions for a maximum of (35).

A. To obtain (36), we first differentiate (35)

$$\begin{aligned} \frac{1}{p_1} \frac{dW}{dg_1} &= \mu E_{\theta/1} E_{\alpha/\theta} Q_n \left(\frac{dn}{dg} + \frac{dn}{dC} \frac{dC_1}{dg_1} \right) \\ &+ \frac{1}{2} \left(\frac{dR_1}{df^+} + \frac{dR_1}{dC_1^+} \frac{dC_1}{dg_1} \right) - \lambda g_1 \left(\frac{dn}{dg} + \frac{dn}{dC} \frac{dC_1}{dg_1} \right) = 0 \end{aligned} \quad (\text{A.1})$$

where we have used (29) to eliminate some terms. In this condition, $dn/dg = (b + \beta)^{-1}$ and $dn/dC = \beta(b + \beta)^{-1}$ are constant; according to (26) and (27), $dC_1^+/dg_1 = dC_1^-/dg_1 \equiv dC_1/dg_1$ does not depend upon $k \in (+, -)$; and from (19), (20) and (24) we have $dR_1/df^+ = -(b + \beta)^{-1}h^+$ and $dR_1/dC_1^+ = -\beta(b + \beta)^{-1}h^+$. Substituting these expressions into (A.1) yields:

$$(b + \beta)^{-1} \left(1 + \beta \frac{dC_1}{dg_1} \right) (\mu E_{\theta/1} Q_n - \frac{h^+}{2} - \lambda g_1) = 0, \quad (\text{A.2})$$

from which (36) is obtained.

B. To obtain (37), we start from

$$\begin{aligned} \frac{1}{p_1} \frac{dW}{dK_1} &= \frac{\mu}{2} \frac{dn}{dC} \left[Q_n(n, \theta_1^+) \frac{dC_1^+}{dK_1} + Q_n(n, \theta_1^-) \frac{dC_1^-}{dK_1} \right] \\ &+ \frac{dR_1}{dC_1^+} \frac{dC_1^+}{dK_1} - \lambda \left[r + \frac{dS_1}{dK_1} + \frac{g_1}{2} \frac{dn}{dC} \left(\frac{dC_1^+}{dK_1} + \frac{dC_1^-}{dK_1} \right) \right] = 0. \end{aligned} \quad (\text{A.3})$$

We then substitute dS_1/dK_1 from (29) and use $dR_1/dC_1^+ = -h^+ dn/dC$, which yields condition (37) after some straightforward manipulations.

C. To obtain (41), we have for $(1/p_2)dW/dg_2$ the same expression as in (A.1) with subscript 1 replaced by subscript 2 and with the following additional term: $-(p_1/p_2)(\lambda - 1)d\psi/dg_2$, where $d\psi/dg_2$ is given by (42). Adding this term to the left-hand side of (A.2) directly yields (41).

D. To obtain (43), we proceed in the same way as for (41). As compared to $(1/p_1)dW/dK_1$, the expression of $(1/p_2)dW/dK_2$ contains an additional term: $-(p_1/p_2)(\lambda - 1)d\psi/dK_2$, with $d\psi/dK_2$ given by (44).

Appendix 2: Derivation of (40)

We have to prove that inequality

$$\mu[Q_n(n, \theta_1^+) - Q_n(n, \theta_1^-)] - h^+ < 0 \quad (\text{A.4})$$

holds if and only if inequality (40) is satisfied. To do so, let us first transform the following relations obtained from (3), $Q_n(n, \theta_1^k) = \beta(n - C_1^k) - f_1^k$, $k \in (+, -)$ by means of (22) and (23). This yields $Q_n(n, \theta_1^+) = (w/K_1)C_1^+ - h^+ - g_1$ and $Q_n(n, \theta_1^-) = (w/K_1)C_1^- - g_1$. Therefore, inequality (A.6) is equivalent to

$$\mu \frac{w}{K_1} (C_1^+ - C_1^-) - (1 + \mu)h^+ < 0. \quad (\text{A.5})$$

From (26) and (27) we have:

$$C_1^+ - C_1^- = \frac{K_1}{w} \frac{2\beta}{b\beta(K_1/w) + (b + \beta)} h^+,$$

which we substitute into (A.7) to yield the result.

Appendix 3: Derivation of (45), (46) and (47)

We first simplify the notation by defining

$$\begin{aligned} \Delta n_2 &\equiv E_{\theta/2}[n] - E_{\theta/1}[\hat{n}], \\ \Delta C_2 &\equiv E_{\theta/2}[C] - E_{\theta/1}[\hat{C}] = \frac{1}{2}(C_2^+ - \hat{C}_1^+) + \frac{1}{2}(C_2^- - \hat{C}_1^-). \end{aligned}$$

A. From (4) we obtain:

$$\Delta n_2 = (b + \beta)^{-1}(\beta \Delta C_2 + b \Delta \theta) \quad (\text{A.6})$$

and from (26) and (27):

$$C_2^+ - \hat{C}_1^+ = C_2^- - \hat{C}_1^- = \beta b [\beta b + \frac{w}{K_2}(b + \beta)]^{-1} \Delta \theta. \quad (\text{A.7})$$

Using the definition of ΔC_2 and (A.9), relation (A.8) can be written as:

$$\Delta n_2 = b \left(\beta + \frac{w}{K_2} \right) [\beta b + \frac{w}{K_2}(b + \beta)]^{-1} \Delta \theta. \quad (\text{A.8})$$

We also have from (26) and (27):

$$1 + \beta \frac{dC_2}{dg_2} = (b + \beta) \left(\beta + \frac{w}{K_2} \right) \left[\beta b + \frac{w}{K_2} (b + \beta) \right]^{-1} \Delta \theta. \quad (\text{A.9})$$

By means of (A.10) and (A.11), first-order condition (41) can then be written as condition (46).

B. Let us now turn to first-order condition (43) on K_2 . In this condition, we have in the last term on the right-hand side:

$$\begin{aligned} E_{\theta/2}[C^2] - E_{\theta/1}[\hat{C}^2] &= \frac{1}{2} \sum_k (C_2^k - \hat{C}_1^k)(C_2^k + \hat{C}_1^k) \\ &= \beta b \left[\beta b + \frac{w}{K_2} (b + \beta) \right]^{-1} \Delta \theta \left[\frac{1}{2} (C_2^+ + \hat{C}_1^+) + \frac{1}{2} (C_2^- + \hat{C}_1^-) \right], \end{aligned} \quad (\text{A.10})$$

where we use (A.9) to obtain the last expression. From (26) and (27) we obtain:

$$C_2^k + \hat{C}_1^k = 2C_2^k - \beta b \left[\beta b + \frac{w}{K_2} (b + \beta) \right]^{-1} \Delta \theta, \quad k \in (+, -),$$

which we substitute in (A.12):

$$\begin{aligned} E_{\theta/2}[C^2] - E_{\theta/1}[\hat{C}^2] &= \beta b \left[\beta b + \frac{w}{K_2} (b\beta) \right]^{-1} \Delta \theta (C_2^+ + C_2^-) \\ &\quad - \beta^2 b^2 \left[\beta b + \frac{w}{K_2} (b + \beta) \right]^{-2} (\Delta \theta)^2, \end{aligned} \quad (\text{A.11})$$

which we can substitute in (43). Then to obtain (45), we make use of the following relationship:

$$\frac{dn}{dC} \frac{dC_2^k}{dK_2} = \beta \frac{w}{K_2^2} \left[\beta b + \frac{w}{K_2} (b + \beta) \right]^{-1} C^k, \quad k \in (+, -)$$

which we infer from (5), (26) and (27).

C. To show that the expression in curly brackets in (45) is negative if and only if (47) holds, we have to prove that inequality (47) is equivalent to

$$\mu[Q_n(n, \theta_2^+) - Q_n(n, \theta_2^-)] < h^+,$$

which we do as in Appendix 2.