

2013/41



Public and private hospitals,
congestion, and redistribution

Chiara Canta and Marie-Louise Leroux



CORE

DISCUSSION PAPER

Center for Operations Research
and Econometrics

Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve
Belgium

<http://www.uclouvain.be/core>

CORE DISCUSSION PAPER
2013/41

Public and private hospitals, congestion, and redistribution

Chiara CANTA¹ and Marie-Louise LEROUX²

July 2013

Abstract

This paper studies how congestion in the public health sector can be used as a redistributive tool. In our model, agents differ in income and they can obtain a health service either from a congested public hospital or from a non congested private one at a higher price. With pure in-kind redistribution, agents fail to internalize their impact on congestion, and the demand for the public hospital is higher than optimal. We show that under full information, the optimal redistribution and sorting across hospitals can be obtained using a lump-sum tax and a subsidy on the private hospital. If income is not observable but the social planner can assign agents across hospitals, the optimal congestion is higher than in the first best in order to relax incentive constraints. Finally, if agents can freely choose across hospitals, the optimal subsidy on the private hospital price may be negative or positive depending on the relative importance of redistribution and efficiency concerns.

Keywords: optimal taxation, mixed health care systems, waiting times, income redistribution

JEL Classification: H21, H23, H44, I11

¹ Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium; CESifo, Germany.
E-mail: chiara.canta@uclouvain.be

² Département des sciences économiques, ESG UQAM, Montréal, Canada; Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: leroux.marie-louise@uqam.ca

We would like to thank K. Cuff, P. De Donder, I. Irvin, R. Lindsey, P-C. Michaud, H. Mou, P. Pestieau, J-F. Wen and N.R. Ziebarth for their valuable comments on this paper.

1 Introduction

Public services often display congestion. In the public debate, this is perceived as a major drawback of public services, and there is a consensus that policy measures should be taken to reduce it. Furthermore, if private non congested facilities offering the same service exist, high-income individuals may opt for the private sector. While this may help to reduce congestion, it may also increase inequalities, because low-income agents are locked in low-quality public facilities. This paper studies how the coexistence of a congested public sector and a non congested private sector can be related to income redistribution. Its main result is that a congested public system can be beneficial for redistribution, both by relaxing incentive compatibility constraints and by redistributing in kind. At the same time, because congestion depends on the demand for the public sector, redistribution affects the congestion of the public sector. This undermines the extent of redistribution with respect to a case in which the quality of the public sector is under direct control of the social planner. To study this problem, we will focus on the market of health services, where public intervention is particularly strong and where the congestion of the public sector is often severe.¹

In countries where health services are provided by a tax funded universal health system (such as the UK, Italy, and Canada), patients can get treated at zero (or very low) price in public hospitals, but may face congestion, resulting in long waiting times, lower attention from practitioners and crowded facilities. Alternatively, they can patronize private hospital, where they face lower congestion, but are charged a price for the treatment. Also the market for health insurance presents similar characteristics. Public health systems often only covers basic services; private insurance, for a higher premium, offers a larger choice of providers, limited waiting times and better amenities. For instance, the German health system relies on compulsory basic insurance with premiums depending on income, but some individuals are allowed to opt out from the public system and to rely exclusively on private health insurance that ensures faster and better services.²

The fact that congestion in the public health sector is a very common phenomenon that affects individual choices and well-being is well documented in the literature.³ In the case of the UK, Besley et al. (1999) show that waiting times at public hospitals are the only significant variable explaining

¹While we concentrate on the case of health services, our model could be applied to any service suffering congestion, such as cultural and educational services (think about museums that charge different prices at different hours), environmental facilities (think about parks or a beaches), or transport networks (see for instance, Russo, 2013).

²Civil servants and high-income individuals can opt out from the public health insurance and join the private system. In Germany, as of 2004, 9.1% of the population had joined the private system.

³For a good review of the waiting time phenomenon in the case of elective surgery and policy measures to tackle it, see for instance Siciliani and Hurst (2005).

the demand for private health insurance. Similarly, using Spanish data, Jofre-Bonet (2000) shows that a higher differential between public and private waiting times increases the probability of purchasing private insurance. She also finds that waiting times in the private sector are not significantly different from zero. Acknowledging the importance of congestion in the public health sector and the potential link between income (re)distribution and the use of the public sector, the objective of the present paper is to design the optimal redistribution policy in the presence of both a congested public sector and a non congested private sector.

We study a model in which agents only differ in income and can be treated either in a public congested hospital or in a non congested private hospital that charges a higher price. The public hospital is financed through taxation and it has a fixed capacity; this latter assumption is reasonable, at least in the short run. Such capacity constraint leads to congestion, which decreases with the capacity of the public hospital and increases with the number of agents that use it. Depending on their income, agents have a different willingness to pay to avoid congestion so that, under pure in-kind redistribution (i.e. the public hospital is financed by a head tax), low-income agents use the public hospital while high-income agents use the private one. Thus, congestion is endogenous and it depends on income redistribution. We characterize the optimal income taxation policy and the optimal repartition of agents across hospitals, under, successively, complete and incomplete information. We also distinguish between the case where the social planner can directly assign patients to hospitals from the case where agents are free to choose.

The first best allocation is such that consumptions are equalized across agents, and each agent has the same probability of being treated in the public hospital. Moreover, fewer agents are treated in the public hospital in the first best than in the pure in-kind redistribution case. This is due to a simple externality problem: when choosing a hospital, agents do not internalize the effect of their choice on the congestion borne by all the other agents using the public hospital. If the social planner cannot directly assign agents to hospitals, the first best allocation can still be decentralized through individualized transfers and a subsidy (or a tax) on the price of the private hospital.

If the social planner cannot observe agents' income but can assign them across hospitals, congestion is distorted up, and the probability of being treated in the public hospital decreases in income. Such a mechanism enables income redistribution as it discourages mimicking from high-income agents, who prefer to pay a high tax than to bear congestion. Therefore, our results provide a redistributive rationale to systems like the German one, where only low-income individuals are locked in the public

system.

Finally, if the social planner cannot assign agents across hospital, the tax schedule can only be conditional on the individual hospital choice, and it consists of a uniform lump-sum tax and of a linear subsidy on the price of the private hospital treatment. The optimal subsidy can be either positive or negative depending on the relative importance of redistributive and efficiency (i.e. congestion) concerns. If redistribution concerns dominates, it is optimal to have a tax on the private hospital treatment in order to extract resources from high-income agents. If on the contrary, congestion concerns dominate, the optimal subsidy may be positive. In words, income redistribution through a tax on the use of the private hospital is limited by its effect on congestion in the public hospital.

To sum up our results, we find that congestion in the public sector is a double-edged tool in redistribution policy. On the one hand, it enables income redistribution: it relaxes incentive compatibility constraints if the social planner can assign agents, and allows redistribution in kind if agents are free to choose between the public and the private sector. On the other hand, with respect to the case in which the quality of the public sector does not depend itself on taxation, taxation entails a higher loss in efficiency because it enhances congestion.

Our paper overlaps different strands of the Public and Health Economics literature. First, it is close to the literature which studies optimal taxation models in the presence of externalities. This literature, which was initiated with Sandmo (1975), concentrates on mixed commodity-and-income taxation policies when one commodity generates an externality.⁴ Our paper differs from this literature in that, first, the decision to consume the externality-generating good is extensive: the agent either goes to the public system or not. Second, this decision, which is at the origin of the externality, only depends on individuals' income levels so that the size of the externality is directly linked to the population's income (re)distribution. Finally, one of our paper's objectives is to show how the congestion externality can foster redistribution through the relaxation of self-selection constraints.⁵

Second, our paper is close to the literature on the public provision of private goods as a mean to redistribute in-kind. Besley and Coate (1991) first showed that universal provision of private goods, financed by a head tax, enhances redistribution. Intuitively, the rich pay for the public provision but buy a service of better quality from private firms and let the poor use the public facility. Our paper

⁴See, for instance, Pirttila and Tuomala (1997), Cremer et al. (1998), Kopczuk (2003) and Sheshinski (2004).

⁵Using a different framework, Pirttila and Tuomala (1997) show that, if there is complementarity between leisure and environmental quality, the exacerbation of the externality can help relaxing self-selection constraints, in a two-income-type model. Also, Cremer et al. (1998) show that the externality affects income taxation when commodity taxation is linear; but this is due to the fact that agents with different incomes consume different quantities of the externality-generating good.

builds on the same idea except that, in our model, the quality of the public good is endogenous and suffers congestion, which can be only indirectly controlled by the planner. Clark and Kim (2007) also study an in-kind redistributive mechanism that enables the self-selection of types. Assuming only two types of agents (high and low ability), they show that a pricing-mechanism consisting in offering the target good at high money price and low time price or at a low money price but high time price achieves commodity-specific egalitarianism when types are not observable. The main differences with our paper reside in the redistributive objective and in the restriction to two types. Finally, Blackorby and Donaldson (1988), Blomquist and Christiansen (1995), Boadway and Marchand (1995) and Cremer and Gahvari (1997) also analyze the usefulness of the public provision of private goods as a redistribution instrument (together with different tax policies) under asymmetric information on agents' health status or productivity.⁶ Like our paper, these papers show how public provision of a private good can help relaxing self-selection constraints under asymmetric information. However, they assume that agents can consume different quantities of the public good for a fixed quality, while in our paper the quantity is fixed but the quality, i.e. the size of congestion, is endogenous and directly linked to the income redistribution policy.

Finally, our paper is also related to the literature that rationalizes the existence of waiting times. In that respect, our paper is close to Bucovetsky (1984) which shows that, in a model where labor is endogenous, the allocation of commodities by waiting is optimal. The idea is that when the planner cannot observe productivities, self-selection constraints create distortions. Yet, the willingness to wait is a signal of productivity that can be used to elicit information on agents' type. However, that model is different from ours as it considers only two productivity classes, a fixed waiting time and commodity taxation together with lump-sum taxes. In line with Bucovetsky (1984), Hoel and Sæther (2003) show, using a non linear taxation model, that it is optimal to provide the public good at a lower quality level so as to solve self-selection problems. They assume that the waiting time is directly set by the social planner and that agents have the same income but different costs associated to waiting.⁷ This is clearly different from our paper. Finally, Marchand and Schroyen (2005), in a linear income taxation model, use waiting times as a rationing device equating demand and supply in the public sector, whose size is controlled by the social planner. Our approach is different as we do not study the optimal size of the public sector but rather consider the public capacity as given and waiting times as the result of agents' choices.

⁶These papers only consider two types of agents while we allow for a continuum of types.

⁷To be fair, they mention that income and waiting costs are likely to be positively correlated.

All in all, the originality of our paper consists in considering that the quality of the public health system depends on the number of patients it treats. We also show that the taxation scheme affects the congestion in the public system. In turn, congestion affects the feasible taxation scheme. With non linear taxation, an upward distortion of congestion discourages mimicking of high-income agents and increases welfare. This is in line with the literature: distorting the quality of the public health system relaxes incentive compatibility constraints. However, in our model the distortion in quality is possible only through the distortion in the number of agents being treated in the public sector, which in turn depends on the taxation in place. Thus, the extent of redistribution is both a consequence and a cause of congestion.

The paper is organized as follows. In the following section, we provide the basic set up and derive the pure in-kind redistribution case. In Section 3, we characterize the first-best allocation and show how it can be decentralized. In the fourth section, we solve the problem under asymmetric information on agents' income. Section 5 concludes.

2 The model

2.1 Basic assumptions

We consider a one-period model in which there is a mass 1 of agents who differ only with respect to income, y . Income is a continuous variable with support $[\underline{y}, \bar{y}]$. In the following, we denote $f(y)$ and $F(y)$ as, respectively, the density and the cumulative functions of y . Agents obtain utility from the consumption of a normal good, c , and from the consumption of a health service which is available in two facilities, say hospitals. These hospitals differ in quality. The first one, is public and is characterized by congestion. The other is private and is not affected by congestion.⁸ The quality of the health treatment is equal to $(h - W)$, where h is a constant representing the benefit of being in good health. The variable W is the size of congestion. We assume that agents have identical preferences that are quasi linear in the net benefit from the treatment:

$$u(c) + (h - W),$$

where $u'(\cdot) > 0$ and $u''(\cdot) < 0$. We assume that h is high enough, to ensure that everybody chooses to be treated. For agents being treated in the public hospital, $W \geq 0$ while for agents being treated

⁸Without loss of generality, we could assume that congestion in the private hospital is positive, as long as it is lower than in the public hospital.

in the private hospital, $W = 0$. Thus, the public congested hospital yields a lower individual benefit than the private hospital. For instance, agents might experience delays or lower availability of health care professionals. Note also that, for simplicity, we assume that h is the same for every agent and is independent of income and of the hospital choice.⁹

In the following, w designates the size of congestion in the public hospital and takes the form

$$w = \begin{cases} \frac{\rho - \hat{x}}{\hat{x}} & \text{if } \rho \geq \hat{x} \\ 0 & \text{if } \rho < \hat{x} \end{cases} \quad (1)$$

where $\rho \in [0, 1]$ is the number of agents being treated in the public hospital. The capacity of the congested public hospital, $\hat{x} \leq 1$, is assumed to be fixed. This assumption is reasonable since adjusting the capacity might be difficult, at least in the short run.

The cost of a treatment in the public hospital is equal to k , while it is equal to p in the private hospital. In the following we will essentially concentrate on the case where $k \leq p$. A first possible justification for this assumption is that the private hospital has some idle capacity in order to ensure no congestion.¹⁰ However, we will also comment on the case where $k > p$.

In the remainder of the paper, we assume that health services are needed with probability one. Alternatively, we could have considered a framework where the service is needed with a probability $\pi \in (0, 1)$ and agents purchase insurance. Under the assumption of fair insurance contracts, the individual ex-ante utility would be $u(c) + \pi(h - W)$ and our qualitative results would remain unchanged. Thus, our model can easily be extended to health insurance markets where different contracts cover services of different quality in case of illness. For instance, public health insurance usually provides full coverage for basic treatments. Agents can opt for private insurance contracts to get higher quality or waiting-free services at a higher price.¹¹

Finally, in the following, we will call “public” the congested hospital but its ownership type is not crucial. What matters in our model is that it is financed by the government.

2.2 Pure in-kind redistribution

As a benchmark, we consider a situation where the government only intervenes through the public

⁹If h was agent-specific, we would have to deal with two dimensions of individual heterogeneity; under asymmetric information, this would significantly complicate the model. Such a framework is beyond the scope of the paper.

¹⁰For a full discussion see Hoel and Sæther (2003), p. 604.

¹¹Public insurance may be cheaper than private insurance also because of lower administrative costs, economies of scale or of smaller marketing costs. For instance, a 2011 study of the Kaiser Family Foundation estimates the administrative costs of Medicare to be approximately 2%, which are substantially lower than the ones of private insurers.

provision of health care. In this case, the public hospital is financed through a head tax, denoted by T . To satisfy the government resource constraint, the head tax must be equal to $\tilde{\rho}k$, where $\tilde{\rho}$ is the equilibrium demand for the public hospital. If agents prefer the private hospital, they pay the treatment out of pocket. In a perfectly competitive market, the price of the treatment in the private hospital equals the marginal cost, p .

Note that this case is equivalent to the one studied in Besley and Coate (1991) in that the publicly provided good is financed through a head tax. However, the main difference with our model is that they assume that the quality of the public sector is set directly by the government. Here, the quality of the public good is represented by congestion.

The agent's problem consists in choosing which hospital to patronize, taking T as given. Since the decision is discrete, we compare the agent's indirect utility in each case. If he chooses the public hospital, the agent's indirect utility is

$$u(y - T) + (h - w),$$

where w is defined by (1). On the contrary, if the agent chooses the private hospital, he has to pay p but he avoids congestion, so that his indirect utility is

$$u(y - T - p) + h.$$

Using the above indirect utility functions, an agent with income y chooses the public hospital if and only if

$$\varphi(y) = u(y - T) - u(y - T - p) - w \geq 0.$$

Under the assumption that $k \leq p$, $\varphi(y)$ is monotonically decreasing in y . This implies that the marginal willingness to pay for the private hospital treatment increases in y . Hence, all agents having an income below the threshold \tilde{y} , such that $\varphi(\tilde{y}) = 0$ choose the public hospital, and the demand for the public hospital is $\tilde{\rho} = F(\tilde{y})$. It is straightforward to see that in equilibrium, the size of congestion \tilde{w} is always positive. Thus, $\tilde{w} = F(\tilde{y})/\hat{x} - 1$. In equilibrium, the income threshold \tilde{y} is implicitly defined by the following equation:

$$u(\tilde{y} - F(\tilde{y})k) - u(\tilde{y} - F(\tilde{y})k - p) = \frac{F(\tilde{y})}{\hat{x}} - 1. \quad (2)$$

This is not surprising as it is less expensive than the private. Under the assumption that $1 - f(\tilde{y})k > 0$,

\tilde{y} is increasing in \hat{x} , in the price of treatment in the private hospital, p . Furthermore, it increases in the cost of the public hospital, k because a higher k implies a higher head tax and thus, lower willingness to pay for the private hospital.

3 First best allocation

In this section, we first derive the optimal allocation when the social planner has perfect information on agents' income, and can assign agents to the public and the private hospital. We then show how this solution can be decentralized when agents can freely choose between hospitals.

3.1 Optimal allocation when incomes are observable

We assume that the social planner seeks to maximize the sum of a concave transformation Φ of agents utility function. In our setting, the first best policy consists in the allocation of consumptions between agents and in the repartition of the population between hospitals. The problem of the social planner consists then in setting individualized lump-sum taxes T_y for each agent with income y .¹² Furthermore, the social planner sets the probability θ_y that an agent with income y is assigned to the public hospital.¹³

The first best program is the following:

$$\begin{aligned} \max_{\theta_y, T_y} \int_{\underline{y}}^{\tilde{y}} \Phi(u(y - T_y) - \theta_y w + h) f(y) dy \\ \text{s.to} \quad \int_{\underline{y}}^{\tilde{y}} [T_y - \theta_y k - (1 - \theta_y)p] f(y) dy \geq 0 \end{aligned}$$

$$\text{where} \quad w = \begin{cases} \frac{\int_{\underline{y}}^{\tilde{y}} \theta_y f(y) dy}{\hat{x}} - 1 & \text{if } \int_{\underline{y}}^{\tilde{y}} \theta_y f(y) dy \geq \hat{x} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

The first constraint is the standard resource constraint of the economy. Note that, in our framework, the social planner is inequality averse with respect to ex-ante utilities.

¹²We also solved the problem allowing the social planner to set transfers conditional on being treated in the public and private hospital, T_y^{PU} and T_y^{PR} , and found that, at the optimum, $T_y^{PU} = T_y^{PR}$.

¹³In order to be as general as possible, we do not restrict θ_y to be binary (i.e. $\theta_y \in \{0, 1\}$). We will indeed show that in the first best the optimal θ_y is the same for every agent, and that it is interior as long as w^{FB} is neither too high nor too low.

The first-order conditions with respect to T_y and θ_y are¹⁴

$$\frac{\partial \mathcal{L}}{\partial T_y} = -\Phi'(EU_y) u'(c_y) + \lambda = 0 \quad \forall y \quad (4)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta_y} &= -w\Phi'(EU_y) f(y) \\ &\quad - \frac{\partial w}{\partial \theta_y} \int_{\underline{y}}^{\bar{y}} \theta_y \Phi'(EU_y) f(y) dy + \lambda [p - k] f(y) = 0. \end{aligned} \quad (5)$$

where $c_y = y - T_y$ denotes consumption and EU_y is the ex-ante utility of an agent with income y :

$$EU_y = u(c_y) - \theta_y w + h.$$

For further use, we find the first best marginal rate of substitution between θ_y and T_y to be equal to

$$MRS(\theta_y, T_y) \equiv -\frac{w}{u'(y - T_y)} = -(p - k) + \frac{1}{\hat{x}} \int_{\underline{y}}^{\bar{y}} \frac{\theta_y}{u'(c_y)} f(y) dy. \quad (6)$$

Let us now study in more details the first order conditions and find the characteristics of the first best allocation. Equation (5) displays the marginal costs and benefits of θ_y . Marginal costs include the direct effect on the utility of agents y and the impact on congestion for all other agents assigned to the public hospital. The marginal benefit is the efficiency gain due to the fact that more agents use the less expensive facility. Note that it is always optimal to set $\int_{\underline{y}}^{\bar{y}} \theta_y f(y) dy \geq \hat{x}$, as

$$\left. \frac{\partial \mathcal{L}}{\partial \theta_y} \right|_{\int_{\underline{y}}^{\bar{y}} \theta_y f(y) dy < \hat{x}} = \lambda [p - k] f(y) > 0,$$

when $p - k \geq 0$. Since the public hospital is more efficient than its private counterpart, it is optimal to have congestion in the public hospital. This, in turn, implies that (5) holds with equality. Rearranging this expression, we obtain:

$$w\Phi'(EU_y) = \lambda [p - k] - \frac{1}{\hat{x}} \int_{\underline{y}}^{\bar{y}} \theta_y \Phi'(EU_y) f(y) dy. \quad (7)$$

Since the right-hand side of this equality is independent from y , $EU_y = EU \quad \forall y$. Replacing in (4), we conclude that $c_y = \bar{c}$, and thus, that higher-income agents pay higher taxes: if $y' > y$, then $T_{y'} > T_y$. In turn, the equalization of consumptions across agents implies that $\theta_y^{FB} = \theta^{FB}$ for all y , and condition

¹⁴We assume that the second order conditions are satisfied.

$\int_{\underline{y}}^{\bar{y}} \theta_y f(y) dy \geq \hat{x}$ reduces to $\theta^{FB} \geq \hat{x}$. Replacing $\Phi'(EU) u'(\bar{c}) = \lambda$ into (7) we get

$$\begin{aligned} w \frac{\lambda}{u'(\bar{c})} &= \lambda [p - k] - \frac{1}{\hat{x}} \int_{\underline{y}}^{\bar{y}} \theta \frac{\lambda}{u'(\bar{c})} f(y) dy \\ \iff \theta^{FB} &= \hat{x} \frac{u'(\bar{c}) [p - k] + 1}{2}. \end{aligned} \quad (8)$$

The optimal probability θ^{FB} is such that the marginal increase in congestion in the public hospital exactly equals the gains in efficiency. Note however that corner solutions are possible. For instance, if the public hospital has a large capacity, and/ or it is much more efficient than the private hospital, i.e., $\hat{x} (u'(\bar{c}) [p - k] + 1) \geq 2$, then $\theta^{FB} = 1$ and it is optimal to have all agents being treated in the public hospital. Conversely, if \hat{x} and $(p - k)$ are relatively small, i.e., $u'(\bar{c}) [p - k] < 1$, then $\theta^{FB} = \hat{x}$. Otherwise, the optimal θ is given by (8). To sum up, the optimal probability of being assigned to the public hospital is

$$\theta^{FB} = \min \left\{ 1, \max \left\{ \hat{x}, \hat{x} \frac{u'(\bar{c}) [p - k] + 1}{2} \right\} \right\}. \quad (9)$$

If the solution is interior,

$$w^{FB} = \frac{u'(\bar{c}) [p - k] - 1}{2}.$$

In the first best, the optimal number of patients treated in the public hospital only depends on the relative efficiency of the public and the private hospital, as well as on the size of \hat{x} . Note that the use of the public hospital is exclusively justified by its higher efficiency. Whenever the private hospital is less costly than its public counterpart ($p < k$), no agents should be assigned to the public hospital.

Let us stress that, as long as $\Phi(\cdot)$ is strictly concave, the social planner is averse to inequality in ex-ante utilities. Thus, an allocation where some agents have different θ s (for instance, some have $\theta_y = 1$, while others have $\theta_y = 0$) is not optimal in our framework. However, the social planner is not averse to inequalities in ex-post utilities, which implies that agents treated in the public hospital are not compensated for congestion. Ex post, they have a lower utility than patients treated in the private hospital: $u(\bar{c}) + (h - w^{FB}) \leq u(\bar{c}) + h$. If, on the opposite, $\Phi(\cdot)$ is linear, inequalities in ex-ante utilities do not matter, so that the optimal probabilities are undetermined and any set of probabilities such that $\int_{\underline{y}}^{\bar{y}} \theta_y f(y) dy / \hat{x} - 1 = w^{FB}$ is optimal.

We summarize our results in the following proposition:

Proposition 1 *Assume that the social planner observes the agents' income, y , and that he can assign agents to either the public or the private hospital. The first-best optimum can be decentralized as follows:*

- i) Taxes are increasing in income, but do not depend on the hospital in which agents are treated.*
- ii) The probability θ_y^{FB} to be assigned to the congested public hospital is the same across income levels and is given by (9).*

Let us finally compare the number of agents treated in the public hospital in the first best and when only redistribution in-kind is possible. It is likely to be different for two reasons. First, the distribution of income is different. Second, under pure in-kind redistribution, agents take ρ as given, not anticipating that by choosing the public hospital they increase the congestion borne by all other agents in that hospital. To make this second point explicit, consider the optimal allocation of patients across hospitals under pure in-kind redistribution. The problem is

$$\begin{aligned} \max_{\theta_y} \int_{\underline{y}}^{\bar{y}} \Phi(\theta_y u(y-T) + (1-\theta_y)u(y-T-p) - \theta_y w + h) f(y) dy. \\ \text{s.to } T = k \int_{\underline{y}}^{\bar{y}} \theta_y f(y) dy \end{aligned}$$

The first derivative with respect to θ_y is

$$\begin{aligned} & \Phi'(EU_y) [u(y-T) - u(y-T-p) - w] \\ & - k \int_{\underline{y}}^{\bar{y}} \Phi'(EU_y) (\theta_y u'(y-T) + (1-\theta_y)u'(y-T-p)) f(y) dy \\ & - \frac{1}{f(y)} \frac{\partial w}{\partial \theta_y} \int_{\underline{y}}^{\bar{y}} \theta_y \Phi'(EU_y) f(y) dy. \end{aligned} \tag{10}$$

As it was established in Section 2.2, the first term is decreasing in y , is positive when $y < \tilde{y}$ (as defined in (2)) and negative when $y > \tilde{y}$. Moreover, as before, one can show that $w \geq 0$, so that $\partial w / \partial \theta_y = f(y) / \hat{x}$ and the last two terms are constant and negative. Thus, there exists a threshold $y^* \leq \tilde{y}$ such that (10) is equal to zero. For all agents with income below y^* , (10) is positive so that they should be treated in the public hospital ($\theta_y = 1$). On the opposite, for all agents with income higher than y^* , (10) is negative and they should be treated in the private hospital ($\theta_y = 0$). This is due to the fact that the marginal willingness to pay to avoid congestion is increasing in income.

Hence, under pure in kind-redistribution, congestion ends up being higher than optimal, because

$\tilde{y} > y^*$. Indeed, agents do not internalize the impact of their choice on government resources (second term in (10)) and their impact on congestion (third term in (10)). This is different from Besley and Coate (1991), who assume that the quality of the public sector is set exogenously by the social planner.

Let us now discuss how the social planner can decentralize the first best optimum if he cannot assign agents across hospitals.

3.2 Decentralized solution when the social planner cannot assign agents

In the previous section, the social planner could assign agents to different hospitals. However, assuming that the social planner has no direct control on agents' choice of hospitals seems more natural. In this case, more instruments than lump-sum taxes are needed to implement the first best solution. We show in this section that it is possible to decentralize it with a uniform subsidy on the price of the non congested private hospital, in addition to individualized lump-sum taxes.¹⁵ Without loss of generality, we assume that the public hospital is financed exclusively through taxation, so that the price agents face to be treated in this hospital is equal to zero.

Since in the first best allocation each agent obtains the same disposable income and the same probability θ^{FB} of being treated in the congested public hospital, the social planner has to resort to random taxation. He then proposes a contract (T_y, T'_y, θ) to agents with income y such that agents pay either T_y with probability θ or T'_y with probability $(1 - \theta)$. He also sets the level of the subsidy τ on the price of the private hospital.¹⁶ Given the level of lump-sum taxes they face (T_y or T'_y), agents then choose which hospital to visit and, if they choose the private hospital, they receive the subsidy.

In Appendix A we prove the following proposition:

Proposition 2 *The first best optimum can be decentralized through the following individualized tax and transfer scheme and a subsidy on the price of the non congested private hospital:*

1. Agents with income y pay either $T_y = y - \bar{c}$ with probability θ^{FB} , defined by (9), or $T'_y = y - (\bar{c} + (1 - \tau)p)$ with probability $1 - \theta^{FB}$.
2. Agents choosing to be treated in the private hospital receive a subsidy, $\tau \in [\underline{\tau}, \bar{\tau}]$, where $\underline{\tau}$ and $\bar{\tau}$

¹⁵Note that the social planner could equivalently regulate the price p . However, we assume here that the social planner has no direct control on p .

¹⁶Implicit in our model is the full subsidization of the public sector. Yet, it is easy to show that reducing the subsidy on the congested public hospital is equivalent to increasing the subsidy on the price of the private hospital since the total demand for health treatments is inelastic (see Hoel and Sæther, 2003).

are both smaller than one and are implicitly defined by

$$u(\bar{c}) - u(\bar{c} - (1 - \bar{\tau})p) = \left[\frac{\theta^{FB}}{\hat{x}} - 1 \right],$$

and

$$u(\bar{c} + (1 - \underline{\tau})p) - u(\bar{c}) = \left[\frac{\theta^{FB}}{\hat{x}} - 1 \right].$$

This ensures the optimal sorting of agents: a proportion θ^{FB} of agents is treated in the congested public hospital.

Two aspects of this result should be highlighted. First, the optimal subsidy might be negative. Consequently, either a tax or a subsidy work in decentralizing the first best. These two options are indifferent for the final allocation, since taxes and transfers are adjusted so as to ensure that every agent has the same net consumption \bar{c} irrespective of the hospital they visit. Receiving a high disposable income and paying a tax is equivalent to receiving a smaller disposable income together with a subsidy.

Second, we assume here that the taxes only depend on income. Once they have paid their taxes, agents are free to choose their most preferred hospital, given the relative prices set by the social planner. To say it differently, the subsidy τ is such that one obtains the optimal repartition of agents between hospitals. If the lump-sum taxes were also dependent on the hospital choice, it is clear that the subsidy would be a redundant instrument.

4 Asymmetric information

We now assume that the social planner cannot observe agents' income, y (and thus, their consumption) and, like in the previous section, we consider successively cases where the social planner can or cannot assign agents to different hospitals.

In the first case, the social planner can assign agents to hospitals, and he offers them a menu of contracts (θ_y, T_y) , specifying the probability to be assigned to the public hospital and the income tax to pay, taking into account possible mimicking behaviors.

In the second case, the social planner cannot assign agents across hospitals. The social planner can nonetheless observe the hospital choice made by each agent, so that transfers depend on this choice.

4.1 Unobservable incomes and possibility to assign agents across hospitals

Let us first study the case where the social planner cannot observe incomes, but can nonetheless assign agents. The problem of the social planner consists in designing a contract (θ_y, T_y) which depends on the agent's reported income:

$$\begin{aligned} & \max_{\theta_y, T_y} \int_{\underline{y}}^{\bar{y}} \Phi(u(y - T_y) - \theta_y w + h) f(y) dy \\ & \text{s.to } \int_{\underline{y}}^{\bar{y}} [T_y - \theta_y k - (1 - \theta_y)p] f(y) dy \geq 0 \\ & \text{s.to } u(y - T_y) - \theta_y w \geq u(y - T_{\tilde{y}}) - \theta_{\tilde{y}} w \quad \forall y, \tilde{y} \end{aligned}$$

where w is defined by (3). The first constraint is the resource constraint while the last constraints are the incentive constraints. Combining them yields

$$u(y - T_y) - u(y' - T_y) \geq u(y - T_{y'}) - u(y' - T_{y'}),$$

implying that T_y has to be non decreasing in y .¹⁷ Moreover, the Spence-Mirrlees condition holds, since $MRS_y(\theta_y, T_y) = -w/u'(y - T_y)$ is monotonically decreasing in y . This means that high-income people need to be compensated more than low-income agents for an increase in θ_y . The set of incentive constraints can thus be replaced by a unique local incentive constraint of the form

$$E\dot{U}_y = u'(y - T_y),$$

together with the monotonicity constraint, $\dot{T}_y > 0$. This, in turn, implies that $\dot{\theta}_y < 0$, to ensure that incentive constraints are satisfied.¹⁸ Such a contract design (progressive taxation together with a decreasing probability of being treated in the congested public hospital) ensures redistribution from high-toward low-income agents, while preventing mimicking behavior. Using the local incentive constraint,

¹⁷See Laffont and Tirole, 1993.

¹⁸From now on, we assume that the first order approach is valid, i.e. that $\dot{T}_y > 0$ and $\dot{\theta}_y < 0$. More precisely, we assume that this is the case in the unconstrained problem in which $\theta_y \in \mathfrak{R}$. We will characterize a necessary condition for this to hold. Yet, we may have bunching due the fact that θ_y must belong to the interval $[0, 1]$. We come back on these points later.

the problem consists in maximizing the following Lagrangian:

$$\begin{aligned}\mathcal{L} &= \int_{\underline{y}}^{\bar{y}} \{ \Phi(EU_y) f(y) \\ &\quad + \lambda [T_y - \theta_y k - (1 - \theta_y)p] f(y) \\ &\quad + \mu_y [u'(y - T_y)] + \dot{\mu}_y EU_y \\ &\quad + \alpha_y [u(y - T_y) - \theta_y w + h - EU_y] \} dy,\end{aligned}$$

where λ is the Lagrange multiplier associated to the resource constraint, μ_y is the co-state variable and α_y is the shadow value of the constraint $EU_y = u(y - T_y) - \theta_y w + h$.

The first order conditions are

$$\frac{\partial \mathcal{L}}{\partial EU_y} = \Phi'(EU_y) f(y) - \alpha_y + \dot{\mu}_y = 0 \quad (11)$$

$$\frac{\partial \mathcal{L}}{\partial T_y} = \lambda f(y) - \mu_y u''(y - T_y) - \alpha_y u'(y - T_y) = 0 \quad (12)$$

$$\frac{\partial \mathcal{L}}{\partial \theta_y} = \lambda (p - k) f(y) - \alpha_y w - \frac{\partial w}{\partial \theta_y} \int_{\underline{y}}^{\bar{y}} \alpha_y \theta_y dy = 0. \quad (13)$$

We show in the appendix that μ_y has the following expression:

$$\mu_y = \frac{\int_y^{\bar{y}} (\Phi'(EU_x) u'(y - T_x) - \lambda) f(x) dx}{u'(y - T_y)} < 0,$$

with transversality conditions $\mu_{\bar{y}} = \mu_{\underline{y}} = 0$. The term μ_y measures the social net marginal loss associated with a marginal decrease of the utility of every agents with income higher than y .¹⁹ We show in the appendix that μ_y is negative, which means that reducing EU_y for all agents with income greater than y leads to an increase in welfare. In other words, redistributing from high-income agents to low-income ones maximizes social welfare. Therefore, as high-income agents are willing to pay more to avoid congestion, it is optimal to set a tax schedule (T_y, θ_y) such that $\dot{T}_y \geq 0$ and $\dot{\theta}_y \leq 0$. The probability of facing congestion is used to relax incentive compatibility constraints.

Remark that, since $\partial EU_y / \partial T_y = -u'(y - T_y)$, $\mu_y u'(y - T_y)$ could be interpreted as the social net marginal loss associated with an increase in T_y for all agents with income greater than y . The gain from an increase in T_y is represented by λ , and it corresponds to the gain in resources. At the same time, an increase in T_y leads to a decrease in the utility of agents with income above y measured by

¹⁹See Atkinson and Stiglitz (1980), p.417.

$\Phi'(EU_y)u'(y - T_y)$.

Let us now study in more details the first order condition with respect to θ_y . It can be rewritten as

$$A - \frac{\alpha_y}{f(y)} = 0,$$

where $A \equiv \left[\lambda(p - k) - \frac{1}{\hat{x}} \int_{\underline{y}}^{\bar{y}} \alpha_y \theta_y dy \right] / w$ is constant across y . First, we show in the appendix that a necessary condition for $\dot{\theta}_y < 0$ within the interval $[0, 1]$ is that $\alpha_y/f(y)$ increases in y . In our analysis, we impose that this is always the case. However, because of the constraint on the support of θ_y , there may be bunching at the top and/or at the bottom of the interval $[\underline{y}, \bar{y}]$, even though $\alpha_y/f(y)$ is increasing in y . For instance, if $p - k$ is positive and large, the private hospital is much more costly than the public one and it may be optimal to assign some agents to the public hospital with probability one. These are the agents at the bottom of the income distribution. If $p - k$ is large enough, we may even have bunching on the whole interval and in such case, no redistribution is possible. In the same way, if $p - k$ is very small, it may be the case that, in the second best, $\theta_y < 1$ for some agents while for others (the high-income ones), $\theta_y = 0$. Note that these results are very different from the first best, in which the use of a congested public hospital was justified only if $p > k$. In the second best, even if $p < k$, θ_y may be positive for some agents (or even for each of them). This is due to the fact that, under asymmetric information, a congested hospital serves redistribution, through the relaxation of self selection constraints.

Assuming an interior solution for θ_y and replacing (12) into (13) we obtain the second best marginal rate of substitution between θ_y and T_y :

$$MRS(\theta_y, T_y) = -(p - k) + \frac{1}{\hat{x}} \int_{\underline{y}}^{\bar{y}} \frac{\theta_y}{u'(c_y)} f(y) dy - \frac{1}{\lambda} \left[\frac{wu''(c_y)}{f(y)u'(c_y)} \mu_y + \frac{1}{\hat{x}} \int_{\underline{y}}^{\bar{y}} \frac{\mu_y u''(c_y)}{u'(c_y)} \theta_y dy \right].$$

As $\mu_y < 0$, both terms in brackets are positive and $MRS(\theta_y, T_y)$ is distorted downward (upward in absolute terms) in comparison to its first best value, defined in (6). In other words, in the second best, an increase in T_y has to be compensated by a higher decrease in θ_y than in the first best to ensure incentive compatibility. The distortion consists of two terms. The first one is rather standard: increasing θ_y has a distributional effect captured by μ_y (described above), and a distortionary effect captured by $u''(c_y)/u'(c_y)$. The second term is less standard and accounts for the effect of an increase in θ_y on the congestion cost borne by all agents in the economy. Consequently, congestion constitutes an additional limit to the size of redistribution, by increasing the distortion created by taxation. This

would not be the case if the quality was directly set by the social planner.

Let us finally study the size of congestion in the second best. Like in the first best, when $p - k > 0$, it is always optimal to have congestion since the first order condition with respect to θ_y , (13), evaluated in $w = 0$ is positive. Thus, replacing (12) into (13), one obtains

$$w^{SB} = \frac{u'(c_y)}{(\lambda f(y) - \mu_y u''(c_y))} \left[\lambda(p - k) f(y) - \frac{f(y)}{\hat{x}} \int_{\underline{y}}^{\bar{y}} \frac{\lambda f(y) - \mu_y u''(c_y)}{u'(c_y)} \theta_y f(y) dy \right],$$

Comparing w^{SB} with its first best counterpart defined in (7), it is possible to show that the two expressions only differ by the term $\mu_y u''(c_y) / u'(c_y)$. Since $\mu_y < 0$, we find that, everything else being equal, congestion is higher in the second best than in the first best. Increasing congestion enables income redistribution by imposing a higher tax burden on high-income agents while avoiding mimicking under asymmetric information. If high-income agents mimic low-income ones, they will suffer higher congestion and that, with a higher probability (as $\dot{\theta}_y < 0$). Hence, setting higher expected congestion for low-income agents than for high-income ones is optimal.

Our results are summarized in the following proposition:

Proposition 3 *Assume that the social planner does not observe agents' income, y , but that he can assign them across hospitals. When $p > k$, the second-best optimum involves that:*

- i) θ_y decreases with y and T_y increases with y .*
- ii) The marginal rate of substitution between θ_y and T_y is distorted downward at the second best.*
- iii) As compared to the first best, congestion in the public hospital is higher.*

These results also apply to the insurance market whenever the government can induce agents to buy different insurance contracts. For instance, in Germany, low-income agents are obliged to get public insurance (in the notation of the model, this would mean that their θ equals 1) and high-income agents can opt out (if private insurance is attractive enough, all such agents do opt out, and $\theta = 0$). Our results imply that this system can be justified on redistributive grounds: such an arrangement helps to relax the incentive compatibility constraints and allows for a higher fiscal pressure on high-income agents.

4.2 Unobservable incomes and free individual choice of hospitals

Let us finally assume that the social planner cannot use the probability θ_y as a policy instrument.

This is a reasonable scenario, since in practice it is unlikely that the social planner has enough power to assign agents across hospitals according to an income-dependent lottery. First, such a system would require a high implementation cost. Second, it is incompatible with a certain view of fairness, and in particular with the principle of equal treatment of equals: ex-ante identical agents may obtain different ex-post utility because of randomization.²⁰ Hence, in this section, we assume that the social planner imposes a lump-sum tax, T to all agents, but that agents who decide to be treated in the private hospital receive a subsidy, τ , on the price p .²¹

Given the social planner's instruments, agents freely choose where to be treated. An agent with income y chooses the public hospital if and only if

$$u(y - T) - w - u(y - T - (1 - \tau)p) \geq 0.$$

The left-hand side of this expression being decreasing in y , there exists an income threshold, $\tilde{y}(T, \tau)$ such that every agent with income above (resp. below) this threshold strictly prefers the private (resp. public) hospital, when the social planner proposes the fiscal scheme (T, τ) . This income threshold solves:

$$u(\tilde{y}(T, \tau) - T) - w(\tilde{y}(T, \tau)) = u(\tilde{y}(T, \tau) - T - (1 - \tau)p), \quad (14)$$

where

$$w(\tilde{y}(T, \tau)) = \frac{F(\tilde{y}(T, \tau))}{\hat{x}} - 1.$$

Note that in our framework, the subsidy on p can be either positive or negative. If $\tau < 0$, agents face a tax and high-income agents (who choose the private hospital) face higher overall taxation. This would redistribute income from high- to low-income agents, even though incomes are unobservable. However, a tax on p results in bigger congestion. Thus, the social planner, when setting τ , will have to trade-off redistribution concerns and congestion containment.

The social planner's problem is now:

$$\begin{aligned} \max_{T, \tau} \quad & \int_{\underline{y}}^{\tilde{y}(T, \tau)} \Phi(u(y - T) + h - w(\tilde{y}(T, \tau))) f(y) dy + \int_{\tilde{y}(T, \tau)}^{\bar{y}} \Phi(u(y - T - (1 - \tau)p) + h) f(y) dy \\ \text{s.to} \quad & T \geq F(\tilde{y}(T, \tau))k + [1 - F(\tilde{y}(T, \tau))] \tau p. \end{aligned}$$

²⁰For a discussion, see Brito et al. (1995) and Fleurbaey (2008).

²¹Also Hoel and Sæther (2003) study the optimal subsidy on private, non congested treatments. However, in their model, there is no concern for income redistribution as agents all have the same income.

The first-order conditions with respect to τ and T are, respectively

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \tau} = & -\frac{\partial \tilde{y}}{\partial \tau} \left[\frac{\partial w}{\partial \tilde{y}} \int_{\mathbf{Y}}^{\tilde{y}} \Phi' (u(y-T) + h - w(\tilde{y})) f(y) dy + \lambda(k - \tau p) f(\tilde{y}) \right] \\ & + p \int_{\tilde{y}}^{\tilde{y}} \Phi' (u(y-T - (1-\tau)p) + h) u'(y-T - (1-\tau)p) f(y) dy - \lambda [1 - F(\tilde{y})] p = 0 \end{aligned} \quad (15)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial T} = & - \left[\int_{\mathbf{Y}}^{\tilde{y}} \Phi' (u(y-T) + h - w(\tilde{y})) u'(y-T) f(y) dy \right. \\ & \left. + \int_{\tilde{y}}^{\tilde{y}} \Phi' (u(y-T - (1-\tau)p) + h) u'(y-T - (1-\tau)p) f(y) dy \right] \\ & - \frac{\partial \tilde{y}}{\partial T} \left[\frac{\partial w}{\partial \tilde{y}} \int_{\mathbf{Y}}^{\tilde{y}} \Phi' (u(y-T) + h - w(\tilde{y})) f(y) dy + \lambda(k - \tau p) f(\tilde{y}) \right] + \lambda = 0, \end{aligned} \quad (16)$$

where, for ease of notation, we set $\tilde{y} \equiv \tilde{y}(T, \tau)$.

Our results are derived in the appendix and summarized in the following proposition:

Proposition 4 *Assume that the social planner does not observe income, y , and that agents are free to choose between the public and the private hospital. The optimal tax scheme consists of a lump-sum tax T and a linear subsidy τ on the price of the private hospital, equal to:*

$$\tau^* = \frac{-\text{cov}(\Phi' u', \mathbb{k}) + \frac{\partial \tilde{y}^c}{\partial \tau} \frac{1}{p} \left[\frac{1}{\tilde{x}} \int_{\mathbf{Y}}^{\tilde{y}} \Phi' (u(y-T) + h - w(\tilde{y})) f(y) dy + \lambda k \right] f(\tilde{y})}{\lambda f(\tilde{y}) \partial \tilde{y}^c / \partial \tau}, \quad (17)$$

where $\mathbb{k} = \{0, 1\}$ takes value 0 (resp. 1) if the agent gets treated in the public (resp. private) hospital. The term $\text{cov}(\Phi' u', \mathbb{k})$ is the covariance between the social marginal utility of consumption and the purchase of a treatment in the private hospital.

Defining the compensated derivative of \tilde{y} with respect to τ as

$$\frac{\partial \tilde{y}^c}{\partial \tau} = \frac{\partial \tilde{y}}{\partial \tau} + \frac{\partial \tilde{y}}{\partial T} \frac{\partial T}{\partial \tau}, \quad (18)$$

it can be proven that it is negative, by fully differentiating (14) together with the resource constraint of the government. The denominator in (17), which represents the efficiency term, is thus negative. Increasing the subsidy on p distorts individual choices, since agents become more likely to buy a treatment in the private hospital. This has to be weighted by the number $f(\tilde{y})$ of agents at the threshold income and by the cost of public funds, λ .

The first term in the numerator of (17) is the equity term and is positive since $\text{cov}(\Phi' u', \mathbb{k}) < 0$

(i.e. choosing the private hospital is negatively correlated with the marginal utility of consumption as this corresponds to higher income levels). It calls for a tax on the private hospital treatment as this enables to increase the tax paid by high-income agents and thus to foster redistribution.

The second term in the numerator of (17) is the effect of \tilde{y}^c on the congestion, $F(\tilde{y})/\hat{x} - 1$, and on the resource constraint. It is negative since the term in brackets is positive and $\partial\tilde{y}^c/\partial\tau < 0$. Indeed, when τ increases, more agents choose the private hospital. The benefits are twofold. First, an increase in the subsidy decreases congestion in the public hospital, which increases the utility of all agents using it. Second, it relaxes the resource constraint by reducing the total cost from running the public hospital. Summing up, this second term in the numerator is related to efficiency and pushes toward subsidization of the price of the treatment in the private hospital so as to decrease the number of agents using the public hospital.

Depending on the relative importance of redistribution and efficiency concerns, τ^* could be either positive or negative. If the efficiency effect dominates the redistribution effect, it is optimal to subsidize the treatment in the private hospital. As an example, let us consider the case where $u(c)$ is linear and the social planner is utilitarian ($\Phi(\cdot)$ is linear) so that there is no redistributive concern ($cov(\Phi'u', \mathbb{k}) = 0$). The subsidy on p simply corrects for the fact that agents do not perfectly internalize congestion and the true cost of the public hospital. On the contrary, if redistributive concerns are very large, a tax on the price of the treatment in the private hospital ($\tau^* < 0$) is optimal.

To sum up, we showed that, when the social planner cannot assign agents, income redistribution is limited by the fact that the quality of the congested hospital depends on the number of patients treated by this hospital, which increases in T and decreases in τ .²² Our result differs from Hoel and Sæther (2003): in their framework, the absence of redistribution motives always calls for a positive subsidy on the treatment in the private hospital. In our model, we could as well have a subsidy or a tax.

5 Conclusion

This paper studies the optimal income taxation scheme in the presence of both a public congested hospital and a private non congested one. In our model, agents have different incomes and have to

²²Similarly, Marchand and Schroyen (2005) show that linear taxation is limited by its impact on waiting times. In their model, labor is endogenous so that an increase in the income tax decreases the opportunity cost of waiting. While they consider linear taxation of income, we consider a subsidy on the non congested private hospital.

choose whether to be treated in the public or the private hospital. The public hospital is free of charge but exhibits congestion whose size depends on the number of patients. Differently from the previous literature, we do not focus on the optimal size of the public health system, but rather on the relation between endogenous congestion and income redistribution under asymmetric information.

We first find that, under pure in-kind redistribution, the number of agents being treated in the public hospital is too high compared to the first-best: agents do not internalize the effect of their choice on congestion. Under symmetric information, even if agents can freely choose the hospital, it is possible to implement the first best through progressive taxation and a subsidy on the use of the private hospital.

We then turn to the asymmetric information problem. First, we analyze the second-best allocation when the social planner can assign agents across hospitals. Agents are proposed a menu including a probability of being assigned to the public hospital and an income tax. We find that the probability of being treated in the public hospital decreases in income and that congestion is higher than in the first best. Such a taxation scheme prevents mimicking by high-income agents, and allows for higher marginal tax rates than otherwise possible. Thus, a system where the rich can opt out or are left out from the public health system could have some appeal in terms of income redistribution. However, we also show that the presence of congestion makes redistribution more distortive; as such, it limits the extent of redistribution with respect to the case where the quality of the public sector is directly set by the social planner.

Second, we study a problem where agents are free to choose which hospital to patronize. In this case, the only instruments available are a lump-sum tax and a subsidy on the price of the private hospital. We find that the optimal subsidy is either positive or negative depending on the relative importance of redistribution and efficiency concerns. Interestingly, if the public sector can be congested, its redistributive role is undermined. If congestion is not too much of a problem, high-income agents face a tax on the use of the private health sector so as to redistribute resources toward low income-agents. Yet, it is lower than if the quality of the public sector had been directly controlled by the planner. If congestion is a strong concern, it may thus even be optimal to have a subsidy, in order to encourage people to patronize the private system and to reduce congestion.

In concluding, it is important to mention that public or private ownership does not really matter in our model: if the congested hospital was privately owned, our results would still be valid as long as the government acts as a third-party payer.

Our model relies on some important assumptions. The first one is that the public capacity is fixed. As we already mentioned, we believe that it is a reasonable assumption in the short run. Moreover, our research question does not concern the optimal size of the public system, but rather the optimal redistribution schemes when a public system already exists. This is certainly relevant for many countries in which the presence of a universal health system can be considered as given, due to political constraints.

Another assumption concerns the modeling of individual utility functions. We use a utility function that is separable in the utility of consumption and in the benefit from the treatment (i.e. on congestion). Allowing for non separability would have been at the expense of increased complexity.

Finally, in this paper, we assume that all agents suffer from the same illness and that the treatment can be obtained both in the public and in the private system. This may not always be the case. Some public hospitals may specialize in treating illnesses that require specific and expensive technologies which are not profitable for private hospitals. It would be interesting to see how our results would be modified when taking into account differences in hospital specialization together with different treatments needs (which may eventually be correlated with incomes). This is in our research agenda.

References

- [1] Besley T. and Coate S., 1991, Public provision of private goods and the redistribution of income, *The American Economic Review*, 18(4), 979-984.
- [2] Besley T., Hall J. and Preston J., 1999, The demand for private health Insurance: do waiting lists matter?, *Journal of Public Economics*, 72(2), 155-181.
- [3] Blackorby C. and Donaldson D., 1988, Cash versus kind, self-selection, and efficient transfers, *The American Economic Review*, 78(4), 691-700.
- [4] Blomquist S. and Christiansen V., 1995, Public provision of private goods as a redistributive device in an optimum income tax model, *Scandinavian Journal of Economics*, 97(4), 547-567.
- [5] Boadway R. and Marchand M., 1995, The use of public expenditure for redistributive policy, *Oxford Economic Papers*, 47, 45-59.
- [6] Brito D.L., Hamilton J.H., Slutsky S.M. and Stiglitz J.E., 1995, Randomization in optimal income tax schedules, *Journal of Public Economics*, 56, 189-223.
- [7] Bucovetski S., 1984, On the use of distributional waits, *Canadian Journal of Economics*, 17(4), 699-717.
- [8] Clark J. and Kim B. , 2007, Paying vs. waiting in the pursuit of specific egalitarianism, *Oxford Economic Papers*, 59, 486-512.
- [9] Cremer H. and Gahvari F., 1997, In-kind transfers, self-selection and optimal tax policy, *European Economic Review*, 41, 97-114.
- [10] Cremer H., F. Gahvari and Ladoux N., 1998, Externalities and optimal taxation, *Journal of Public Economics*, 70, 343-364.
- [11] Fleurbaey, M., 2008. *Fairness, responsibility and welfare*. Oxford University Press.
- [12] Hoel M. and Sæther E.M., 2003, Public health care with waiting time: the role of supplementary private health care, *Journal of Health Economics*, 22, 599-616.
- [13] Iversen T., 1997, The effect of a private sector on the waiting times in a national health service, *Journal of Health Economics*, 16, 381-396.

- [14] Jofre-Bonet M., 2000, Public health and private insurance demand: the waiting time as a link, *Health Care Management Science*, 3, 51-71.
- [15] Kaiser Family Foundation and Health Research Education Trust, 2011, A primer on medicare financing, available at <<http://www.kff.org/medicare/upload/7731-03.pdf>>.
- [16] Kopczuk W., 2003, A note on optimal taxation in the presence of externalities, *Economics Letters*, 80, 81-86.
- [17] Laffont J-J. and Tirole J., 1993, *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press.
- [18] Marchand M. and Schroyen F., 2005, Can a mixed health care system be desirable on equity grounds?, *Scandinavian Journal of Economics*, 107(1), 1-23.
- [19] OECD, 2004, *Private Health Insurance in OECD Countries - Policy Brief*, available at <<http://www.oecd.org/dataoecd/42/6/33820355.pdf>>
- [20] Pirttila J. and Tuomala M., 1997, Income tax, commodity tax and environmental policy, *International Tax and Public Finance*, 4, 379-393.
- [21] Russo A., 2013, Pricing of transport networks, redistribution and optimal taxation, forthcoming *Journal of Public Economic Theory*.
- [22] Sandmo A., 1975, Optimal taxation in the presence of externalities, *The Swedish Journal of Economics*, 77(1), 86-98.
- [23] Sheshinski E., 2004, On atmosphere externality and corrective taxes, *Journal of Public Economics*, 88, 727-734.
- [24] Siciliani L. and Hurst J., 2005, Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries, *Health Policy*, 72, 201-215.

APPENDIX

A Proof of Proposition 2

Let us first define the disposable income, i.e. the income y net of taxes, as

$$D_y \equiv y - T_y$$

In the following, we show how to choose the appropriate levels of (T_y, T'_y, τ) in order to implement the first-best allocation (\bar{c}, θ^{FB}) . To do so, an agent with income y will be attributed D_y with probability θ^{FB} , and D'_y with probability $(1 - \theta^{FB})$. To be optimal, this allocation (D_y, D'_y, τ) should satisfy three conditions. The first two conditions relate to the choice of optimal consumption levels while the last one relates to the optimal partition between hospitals.

1. Consumptions should be equal to the first best level, namely $D_y = \bar{c}$ and $D'_y - (1 - \tau)p = \bar{c}$.
2. This allocation (D_y, D'_y, τ) should also satisfy the government budget constraint, evaluated at the optimum,

$$E[y] - \theta^{FB}(D_y + k) - (1 - \theta^{FB})(D'_y + \tau p) \geq 0.$$

3. Under this allocation (D_y, D'_y, τ) , agents with income y should choose the public hospital, that is

$$[u(D_y) - u(D_y - (1 - \tau)p)] \geq \left[\frac{\theta^{FB}}{\hat{x}} - 1 \right],$$

and agents with income y' should choose the private hospital, that is

$$[u(D'_y) - u(D'_y - (1 - \tau)p)] \leq \left[\frac{\theta^{FB}}{\hat{x}} - 1 \right].$$

Setting, as in condition 1,

$$\begin{aligned} D_y &= \bar{c} \\ D'_y &= \bar{c} + (1 - \tau)p \end{aligned}$$

trivially satisfies condition 2. This implies differentiated lump-sum taxes / transfers equal to

$$\begin{aligned} T_y &= y - \bar{c} \leq 0 \\ T'_y &= y' - (\bar{c} + (1 - \tau)p) \leq 0 \end{aligned}$$

for agents visiting the public and the private hospital, respectively.

Replacing for the expressions of D_y , D'_y and θ^{FB} in the inequalities of condition 3, we obtain

$$A \equiv u(\bar{c}) - u(\bar{c} - (1 - \tau)p) \geq \left[\frac{\theta^{FB}}{\hat{x}} - 1 \right] \quad (19)$$

and

$$B \equiv u(\bar{c} + (1 - \tau)p) - u(\bar{c}) \leq \left[\frac{\theta^{FB}}{\hat{x}} - 1 \right]. \quad (20)$$

Let us first mention that the above inequalities imply that necessarily $\tau \leq 1$ so that there can be at best a complete reimbursement of the price of the treatment at the private hospital. Since u is concave, $A \geq B$. Furthermore, both A and B increase when τ decreases. Thus, (19) is satisfied for any $\tau \leq \bar{\tau}$, where $\bar{\tau}$ is implicitly defined by

$$u(\bar{c}) - u(\bar{c} - (1 - \bar{\tau})p) = \left[\frac{\theta^{FB}}{\hat{x}} - 1 \right].$$

In a similar way, (20) is satisfied by any $\tau \geq \underline{\tau}$, where $\underline{\tau}$ is implicitly defined by

$$u(\bar{c} + (1 - \underline{\tau})p) - u(\bar{c}) = \left[\frac{\theta^{FB}}{\hat{x}} - 1 \right].$$

By concavity of $u(\cdot)$, $\bar{\tau} > \underline{\tau}$. Consequently, there exists a set of subsidies $[\underline{\tau}, \bar{\tau}]$ satisfying inequalities (19) and (20).

B Asymmetric information and possibility to assign agents to the public system

Replacing (11) into (12) and rearranging terms yield:

$$\begin{aligned} & \mu_y u''(y - T_y) + u'(y - T_y) \dot{\mu}_y = \lambda f(y) - \Phi'(EU_y) u'(y - T_y) f(y) \\ \Leftrightarrow & \frac{\partial \mu_y u'(y - T_y)}{\partial y} = \lambda f(y) - \Phi'(EU_y) u'(y - T_y) f(y) \end{aligned}$$

Using the transversality conditions, $\mu_{\underline{y}} = \mu_{\bar{y}} = 0$, we obtain

$$\mu_y = \frac{\int_y^{\bar{y}} (\Phi'(EU_x) u'(y - T_x) - \lambda) f(x) dx}{u'(y - T_y)}$$

The sign of μ_y depends on the sign of the numerator. Since $\Phi'(EU_y) u'(y - T_y)$ is decreasing, the numerator is first decreasing and then increasing. This, combined with the transversality conditions, implies that the numerator is always negative and $\mu_y < 0$. Overall, given the transversality conditions, one has that $\mu_y \leq 0$ for every $y \in [\underline{y}, \bar{y}]$.

Using $\alpha_y/f(y) = \lambda/u'(y - T_y) - \mu_y u''(y - T_y)/(u'(y - T_y)f(y))$ and the expression for μ_y , we compute the derivative of $\alpha_y/f(y)$ with respect to θ_y :

$$\begin{aligned} \frac{\partial \alpha_y/f(y)}{\partial \theta_y} &= -\frac{u''(y - T_y)}{f(y)u'(y - T_y)} \frac{\partial \mu_y}{\partial \theta_y} \\ &= \frac{u''(y - T_y)}{f(y)u'(y - T_y)} \\ &\times \left\{ w\Phi''(EU_y) u'(y - T_y) f(y) + \frac{\int_y^{\bar{y}} \frac{\partial w}{\partial \theta_y} \theta_x \left(\Phi''(EU_x) u'(y - T_x) \right) f(x) dx}{u'(y - T_y)} \right\} \end{aligned}$$

which is always positive. Thus the second-order condition with respect to θ_y holds so that, using the implicit function theorem, we get:

$$\text{sign} \left(\frac{d\theta_y}{dy} \right) = -\text{sign} \left(\frac{\partial \alpha_y/f(y)}{\partial y} \right).$$

Consequently, θ_y is decreasing in y only if $\alpha_y/f(y)$ is increasing in y .

C Proof of Proposition 4

Using the first-order conditions (15) and (16), we obtain

$$\frac{\partial \mathcal{L}^c}{\partial \tau} = \frac{\partial \mathcal{L}}{\partial \tau} + \frac{\partial \mathcal{L}}{\partial T} \frac{\partial T}{\partial \tau} \quad (21)$$

where \mathcal{L}^c denotes the compensated Lagrangian and where $\partial T/\partial \tau = [1 - F(\tilde{y}(T, \tau))]p$ is obtained from the resource constraint of the government. In the following we use the compensated derivative of \tilde{y} with respect to τ , $\partial \tilde{y}^c/\partial \tau$, defined by (18). Replacing for (15), (16), and (18), (21) can be rewritten as

$$\begin{aligned} \frac{\partial \mathcal{L}^c}{\partial \tau} &= -\frac{\partial \tilde{y}}{\partial \tau} \left[\frac{\partial w}{\partial \tilde{y}} \int_{\mathbf{Y}}^{\tilde{y}} \Phi'(u(y-T) + h - w(\tilde{y})) f(y) dy - \lambda(k - \tau p) f(\tilde{y}) \right] \\ &\quad + p \int_{\tilde{y}}^{\tilde{y}} \Phi'(u(y-T - (1-\tau)p) + h) u'(y-T - (1-\tau)p) f(y) dy - \lambda [1 - F(\tilde{y})] p \\ &\quad + \frac{\partial T}{\partial \tau} \left[\begin{aligned} &-\int_{\mathbf{Y}}^{\tilde{y}} \Phi'(u(y-T) + h - w(\tilde{y})) u'(y-T) f(y) dy \\ &-\int_{\tilde{y}}^{\tilde{y}} \Phi'(u(y-T - (1-\tau)p) + h) u'(y-T - (1-\tau)p) f(y) dy \\ &-\frac{\partial \tilde{y}}{\partial T} \left[\frac{\partial w}{\partial \tilde{y}} \int_{\mathbf{Y}}^{\tilde{y}} \Phi'(u(y-T) + h - w(\tilde{y})) f(y) dy - \lambda(k - \tau p) f(\tilde{y}) \right] + \lambda \end{aligned} \right] \\ &= -\frac{\partial \tilde{y}^c}{\partial \tau} \left[\frac{\partial w}{\partial \tilde{y}} \int_{\mathbf{Y}}^{\tilde{y}} \Phi'(u(y-T) + h - w(\tilde{y})) f(y) dy + \lambda(k - \tau p) f(\tilde{y}) \right] \\ &\quad + p \left[\begin{aligned} &\int_{\tilde{y}}^{\tilde{y}} \Phi'(u(y-T - (1-\tau)p) + h) u'(y-T - (1-\tau)p) f(y) dy \\ &-[1 - F(\tilde{y})] \int_{\mathbf{Y}}^{\tilde{y}} \Phi'(u(y-T) + h - w(\tilde{y})) u'(y-T) f(y) dy \\ &-[1 - F(\tilde{y})] \int_{\tilde{y}}^{\tilde{y}} \Phi'(u(y-T - (1-\tau)p) + h) u'(y-T - (1-\tau)p) f(y) dy \end{aligned} \right] \end{aligned}$$

The last term in brackets is the covariance between the marginal utility of consumption and the choice of the private hospital and is defined by

$$\begin{aligned} cov(\Phi' u', \mathbb{k}) &= \int_{\tilde{y}}^{\tilde{y}} 1 \times \Phi'(U^{PR}) u'(y-T - (1-\tau)p) f(y) dy + \int_{\mathbf{Y}}^{\tilde{y}} 0 \times \Phi'(U^{PU}) u'(y-T) f(y) dy \\ &\quad - [0 \times F(\tilde{y}) + 1 \times (1 - F(\tilde{y}))] \\ &\quad \times \left(\int_{\mathbf{Y}}^{\tilde{y}} \Phi'(U^{PU}) u'(y-T) f(y) dy + \int_{\tilde{y}}^{\tilde{y}} \Phi'(U^{PR}) u'(y-T - (1-\tau)p) f(y) dy \right) \end{aligned}$$

where $U^{PR} = u(y-T - (1-\tau)p) + h$ and $U^{PU} = u(y-T) - w(\tilde{y}) + h$. The indicator $\mathbb{k} = \{0, 1\}$ takes value 0 (resp. 1) if the agent chooses the public (resp. private) hospital. At the optimum, rearranging $\partial \mathcal{L}^c/\partial \tau = 0$ yields Proposition 4.

Recent titles

CORE Discussion Papers

- 2013/1 Pierre PESTIEAU and Maria RACIONERO. Harsh occupations, health status and social security.
- 2013/2 Thierry BRECHET and Henry TULKENS. Climate policies: a burden or a gain?
- 2013/3 Per J. AGRELL, Mehdi FARSI, Massimo FILIPPINI and Martin KOLLER. Unobserved heterogeneous effects in the cost efficiency analysis of electricity distribution systems.
- 2013/4 Adel HATAMI-MARBINI, Per J. AGRELL and Nazila AGHAYI. Imprecise data envelopment analysis for the two-stage process.
- 2013/5 Farhad HOSSEINZADEH LOTFI, Adel HATAMI-MARBINI, Per J. AGRELL, Kobra GHOLAMI and Zahra GHELEJ BEIGI. Centralized resource reduction and target setting under DEA control.
- 2013/6 Per J. AGRELL and Peter BOGETOFT. A three-stage supply chain investment model under asymmetric information.
- 2013/7 Per J. AGRELL and Pooria NIKNAZAR. Robustness, outliers and Mavericks in network regulation.
- 2013/8 Per J. AGRELL and Peter BOGETOFT. Benchmarking and regulation.
- 2013/9 Jacques H. DREZE. When Borch's Theorem does not apply: some key implications of market incompleteness, with policy relevance today.
- 2013/10 Jacques H. DREZE. Existence and multiplicity of temporary equilibria under nominal price rigidities.
- 2013/11 Jean HINDRIKS, Susana PERALTA and Shlomo WEBER. Local taxation of global corporation: a simple solution.
- 2013/12 Pierre DEHEZ and Sophie POUKENS. The Shapley value as a guide to FRAND licensing agreements.
- 2013/13 Jacques H. DREZE and Alain DURRE. Fiscal integration and growth stimulation in Europe.
- 2013/14 Luc BAUWENS and Edoardo OTRANTO. Modeling the dependence of conditional correlations on volatility.
- 2013/15 Jens L. HOUGAARD, Juan D. MORENO-TERNERO and Lars P. OSTERDAL. Assigning agents to a line.
- 2013/16 Olivier DEVOLDER, François GLINEUR and Yu. NESTEROV. First-order methods with inexact oracle: the strongly convex case.
- 2013/17 Olivier DEVOLDER, François GLINEUR and Yu. NESTEROV. Intermediate gradient methods for smooth convex problems with inexact oracle.
- 2013/18 Diane PIERRET. The systemic risk of energy markets.
- 2013/19 Pascal MOSSAY and Pierre M. PICARD. Spatial segregation and urban structure.
- 2013/20 Philippe DE DONDER and Marie-Louise LEROUX. Behavioral biases and long term care insurance: a political economy approach.
- 2013/21 Dominik DORSCH, Hubertus Th. JONGEN, Jan.-J. RÜCKMANN and Vladimir SHIKHMAN. On implicit functions in nonsmooth analysis.
- 2013/22 Christian M. HAFNER and Oliver LINTON. An almost closed form estimator for the EGARCH model.
- 2013/23 Johanna M. GOERTZ and François MANIQUET. Large elections with multiple alternatives: a Condorcet Jury Theorem and inefficient equilibria.
- 2013/24 Axel GAUTIER and Jean-Christophe POUDOU. Reforming the postal universal service.
- 2013/25 Fabian Y.R.P. BOCART and Christian M. HAFNER. Fair re-valuation of wine as an investment.
- 2013/26 Yu. NESTEROV. Universal gradient methods for convex optimization problems.
- 2013/27 Gérard CORNUEJOLS, Laurence WOLSEY and Sercan YILDIZ. Sufficiency of cut-generating functions.
- 2013/28 Manuel FORSTER, Michel GRABISCH and Agnieszka RUSINOWSKA. Anonymous social influence.

Recent titles

CORE Discussion Papers - continued

- 2013/29 Kent WANG, Shin-Huei WANG and Zheyao PAN. Can federal reserve policy deviation explain response patterns of financial markets over time?
- 2013/30 Nguyen Thang DAO and Julio DAVILA. Can geography lock a society in stagnation?
- 2013/31 Ana MAULEON, Jose SEMPERE-MONERRIS and Vincent VANNETELBOSCH. Contractually stable alliances.
- 2013/32 Jean-François CAULIER, Ana MAULEON and Vincent VANNETELBOSCH. Allocation rules for coalitional network games.
- 2013/33 Georg KIRCHSTEIGER, Marco MANTOVANI, Ana MAULEON and Vincent VANNETELBOSCH. Limited farsightedness in network formation.
- 2013/34 Ana MAULEON and Vincent VANNETELBOSCH. Relative concerns and delays in bargaining with private information.
- 2013/35 Kristof BOSMANS, Koen DECANCO and Erwin OOGHE. What do normative indices of multidimensional inequality really measure?
- 2013/36 Alain PHOLO BALA, Dominique PEETERS and Isabelle THOMAS. Spatial issues on a hedonic estimation of rents in Brussels.
- 2013/37 Lionel ARTIGE, Antoine DEDRY and Pierre PESTIEAU. Social security and economic integration.
- 2013/38 Nicolas BOUCKAERT and Erik SCHOKKAERT. Differing types of medical prevention appeal to different individuals.
- 2013/39 Pierre M. PICARD. Trade, economic geography and the choice of product quality.
- 2013/40 Tanja B. MLINAR and Philippe CHEVALIER. Pooling in manufacturing: do opposites attract?
- 2013/41 Chiara CANTA and Marie-Louise LEROUX. Public and private hospitals, congestion, and redistribution.

Books

- V. GINSBURGH and S. WEBER (2011), *How many languages make sense? The economics of linguistic diversity*. Princeton University Press.
- I. THOMAS, D. VANNESTE and X. QUERRIAU (2011), *Atlas de Belgique – Tome 4 Habitat*. Academia Press.
- W. GAERTNER and E. SCHOKKAERT (2012), *Empirical social choice*. Cambridge University Press.
- L. BAUWENS, Ch. HAFNER and S. LAURENT (2012), *Handbook of volatility models and their applications*. Wiley.
- J-C. PRAGER and J. THISSE (2012), *Economic geography and the unequal development of regions*. Routledge.
- M. FLEURBAEY and F. MANIQUET (2012), *Equality of opportunity: the economics of responsibility*. World Scientific.
- J. HINDRIKS (2012), *Gestion publique*. De Boeck.
- M. FUJITA and J.F. THISSE (2013), *Economics of agglomeration: cities, industrial location, and globalization*. (2nd edition). Cambridge University Press.
- J. HINDRIKS and G.D. MYLES (2013). *Intermediate public economics*. (2nd edition). MIT Press.
- J. HINDRIKS, G.D. MYLES and N. HASHIMZADE (2013). *Solutions manual to accompany intermediate public economics*. (2nd edition). MIT Press.

CORE Lecture Series

- R. AMIR (2002), Supermodularity and complementarity in economics.
- R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.
- A. SHAPIRO (2010), Stochastic programming: modeling and theory.