TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

ECON
RESEARCH GROUP ECONOMICS
Institute of Mathematical Methods in Economics

ECON WPS

September 2013

# Saving the public from the private? Incentives and outcomes in dual practice

by
Michael Kuhn
Robert Nuscheler

Working Paper 02/2013

This paper can be downloaded without charge from http://www.econ.tuwien.ac.at/wps/econ_wp_2013_02.pdf

# Saving the public from the private? Incentives and outcomes in dual practice[*]

Michael Kuhn[†] Robert Nuscheler[‡]

September 2013

**Abstract**

We consider a setting of dual practice, where a physician offers free public treatment and, if allowed, a private treatment for which patients have to pay out of pocket. Private treatment is superior in terms of health outcomes but more costly and time intensive. For the latter reason it generates waiting costs. As patients differ in their propensity to benefit from private treatment and in their costs of waiting for treatment, we study the physician's incentives to supply private care and to allocate waiting time to public and private sectors and contrast it with the first-best allocation. The physician shifts waiting costs to public patients in order to increase the willingness-to-pay for private treatment. While this waiting time allocation turns out to be socially optimal, the resulting positive network effect leads to an over-provision of private care if and only if waiting costs are sufficiently high. A second-best allocation arises when the health authority selects physician reimbursement in the public segment but has no control over private provision. Depending on the welfare weight the health authority attaches to physician profits a ban of dual practice may improve on the second-best allocation. Due to patient heterogeneity, such a ban would affect patients differently.

*Keywords:* Dual practice, health outcomes, health care financing, provider contract, waiting times.

*JEL classification numbers:* I11, I18, H51, L33, L51.

[*] We thank Mathias Kifmann for helpful comments on an earlier version. The paper also benefitted from discussions during the economics workshop in Ottobeuren 2012 and the annual conference of the German Economic Association in Göttingen 2012.

[†] Vienna Institute for Demography / Wittgenstein Centre, Wohllebengasse 12-14, 1040 Vienna, Austria, Email: michael.kuhn@oeaw.ac.at.

[‡] Corresponding auhtor. University of Augsburg, Department of Economics, Universitätsstr. 16, 86159 Augsburg, Germany, Email: robert.nuscheler@wiwi.uni-augsburg.de, Phone: +49-821-598-4202, Fax: +49-821-598-4232.

ECON WPS

# 1 Introduction

The provision of health services of both a public and private nature by one and the same provider, typically labeled dual practice, is a common yet strongly contested arrangement within many health care systems.[1] Generally, dual practice is an arrangement that admits the inflow of additional private resources into a public sector that is subject to more or less severe resource constraints. In a number of countries with a national health service, such as the UK or Southern European countries, the provision of treatments (e.g. surgery) within private practice is allowed as a means to reduce public sector waiting times (e.g. Iversen 1997, Barros and Olivella 2005, Gonzalez 2005). The idea is that the profit incentive from private practice may stimulate additional treatment effort so that total waiting time is reduced to the benefit of even those patients who are not prioritized and remain on the public waiting list. However, such an arrangement may be plagued by cream-skimming (Barros and Olivella 2005, Gonzalez 2005) and the incentive to manipulate upwards public waiting times in order to stimulate private demand (Iversen 1997). A second argument as to why dual practice may contribute to relieving funding pressures within a public health sector is provided by Bir and Eggleston (2003) and Eggleston and Bir (2006) who argue that the additional income (and perhaps professional satisfaction) from private practice helps to attract highly skilled personnel into the public health sector even at comparatively low levels of public reimbursement.

A third argument relates to dual practice as a vehicle to allow for the provision of additional and/or higher quality services which are not included in public health care plans but may nevertheless be valued by some patients (e.g. Brekke and Sørgard 2007, Biglaiser and Ma 2007). Indeed, funding pressures within the public health care system force more and more funding agents (governments or health insurers) to restrict services to a ‘basic bundle’. Additional services, which may be valuable to some individuals but do not prove cost-effective for the whole population, may then be purchased on the private market. In Germany, for instance, such arrangements are already in place for dentistry, where many intensive treatments have been excluded from statutory health insurance and deferred to a private market (possibly private insurance). Similar arrangements apply to the services of ambulatory physicians in Austria, where certain treatments are exempt from

[1] Garcia-Prado and Gonzalez (2011) provide an excellent summary on both the prevalence of dual practice within developed and developing countries and the debate on its benefits and disadvantages. See Garcia-Prado and Gonzalez (2007) and Socha and Bech (2011) for additional surveys on the theme.

ECON WPS

statutory health insurance or subject to considerable co-payments with patients having an option to purchase private top-up insurance. Finally, within secondary care, in many countries such as Austria and Germany public patients do not typically have a statutory right to be treated by senior consultants, but such an option may be purchased privately. In the light of increasing pressures on public health care budgets triggered, e.g. by the ageing of populations, we would expect a rather general and far-reaching shift in health care finance towards a mix of public funding for basic services together with top-up private funding.

The focus of our analysis lies exactly on this third rationale for dual practice. We consider the provision of a basic treatment to which patients have access at no cost within a public health plan, while, within dual practice, a more effective but more costly treatment may be offered at an out-of-pocket fee. Even though the more intensive treatment is not cost-effective for the whole population, dual practice may be valuable in that it allows access to the more intensive treatment for those patients who stand to benefit from it, while at the same time containing public health expenditure. The more intensive treatment, however, is not only more costly financially but often more time consuming (this is immediate if treatment intensity is related to the duration of the consultation). If this is the case, private provision generates (additional) waiting times whenever the provider is time constrained. It is not clear whether waiting times are allocated to private or public patients; however, in reality, it will typically be public patients who are subject to longer waiting - or alternatively to further reductions in treatment intensity (e.g. consultation times). With patients placing a negative value on waiting times, from the physician's perspective the generation of public waiting times is attractive as it stimulates demand for private treatment. From a social perspective the generation of waiting times constitutes an external cost placed on public sector patients. This suggests the over-provision of private treatments. However, as long as the physician is acting monopolistically within the private segment, the presence of market power implies an offsetting tendency towards under-provision. Thus, it is not clear a priori, as to whether or not dual practice will generate an excessive or insufficient supply of private care; and even more fundamentally, it is not clear as to whether dual practice is socially valuable or whether it should be disallowed.

More specifically, we consider a set-up where private treatment is offered by a monopolistic physician. Those patients who do not purchase privately the more intensive treatment remain in the public segment. Patients differ in their propensity to benefit from

ECON WPS

treatment (i.e. in their severity) and, therefore, in their valuation of the extra benefit from private treatment. At the same time they suffer a disutility from waiting, which is assumed to increase in severity. Patients are fully informed about their severity and purchase whichever option generates the higher net benefit. By comparing with the social optimum the profit-maximizing supply of private (and public) treatments and allocation of waiting times we are able to characterize situations of under-supply and over-supply of private care depending on the financial and time cost. In addition, we show that the physician's private incentive to impose the waiting costs entirely on public patients is in line with the social incentive, albeit for very different reasons. Distortions arise with respect to the size of the private practice: Low waiting costs tend to imply an under-supply while the reverse is true for high waiting costs, the latter implying a substantial level of negative externalities. As such, this result may be expected, but more surprisingly, it turns out that while the problem of over-supply can be resolved by offering the physician a sufficient level of reimbursement within the public segment, the reverse is not true with regard to the problem of under-supply. Assuming that the physician cannot be forced to provide services within the public segment at a loss, the best a health authority can do is to reduce the physician's income from public sector services to zero. However, as it turns out, correcting the under-supply of private treatment would require that public remuneration is set at a level that induces a negative income from public services, an arrangement that is infeasible.

We then expand the analysis to a second-best context, where the health authority can only indirectly affect the provision of private treatments by altering the public sector reimbursement and possibly issuing a ban on dual practice. In so doing we allow for varying weights of physician profit within the social welfare function, ranging from a weight of one, i.e. a utilitarian welfare function, to a weight of zero, i.e. a patient centered welfare function. For all but strictly utilitarian welfare functions the containment of the physician's profit turns up as an additional influence on decision-making. Specifically, non-utilitarian health authorities have an incentive to allow an excessive supply of private care and, thus, excessive external costs in order to contain the physician's income. Thus, for all but strictly utilitarian health authorities private treatment is, indeed, over-supplied for high levels of waiting costs. However, as our analysis shows this is not because such an allocation is unavoidable but rather because it turns up in a trade-off against excessive rents.

Finally, we consider whether the health authority can improve second-best welfare by

ECON WPS

banning dual practice altogether. Indeed, given the high levels of public reimbursement that are needed to reduce private practice at high levels of waiting costs, it turns out that non-utilitarian health-authorities prefer, in this instance, to ban dual practice. More surprisingly, however, for (relatively) patient centered health authorities a ban of dual practice may even be warranted in a situation where there is an under-supply of private practice to begin with. This apparently paradoxical reduction (to zero) of private provision from a level that is already too low from a first-best perspective, follows as patient-centered health authorities consider not so much the cost-effective private-public mix but rather the mix that maximizes patient surplus net of private fees. In such an instance, dual practice may be shut down even at modest levels of the external cost. In addition we can show that while a ban of dual practice obviously benefits public patients by relieving them from waiting, it also benefits private patients with a relatively low propensity to benefit from the more intensive treatment. This is because their willingness to pay for avoiding public sector waiting induces them to pay a private price in excess of the additional treatment benefit. Nevertheless, high severity patients are always rendered worse off by a ban of private practice.

The current paper is most related to Iversen (1997) and Brekke and Sørgard (2007).[2] The idea that public sector waiting time affects the tradeoff between public and private health care was first formalized by Iversen (1997). Like in our framework, longer public waits have a positive impact on the demand for private health care. Apart from this mechanism, there are several important differences. In Iversen's unconstrained case, where all patients can potentially be added to the public waiting list, public waiting is not even an instrument that the physician chooses – it is the health authority that sets waiting time optimally. Only in the constrained case, where patients need to exceed some severity threshold in order to be added to the public waiting list can the physician decide on waiting time. But even in this case the structure of our paper is markedly different. In our framework the regulator chooses public reimbursement before the physician decides on

[2] Other related papers include, Gonzalez (2005) and Barros and Olivella (2005) who highlight the problem of cream skimming in dual practice environments. Gonzalez (2004) concentrates on the importance of public sector prestige for private practice profits, while Biglaiser and Ma (2007) investigate the impact of physician heterogeneity (dedicated physicians versus moonlighters) on the optimal contract. Gonzalez and Macho-Stadler (2013) build a very specific framework to analyze a number of policies on how to regulate dual practice, including a ban and softer limitations of it. Finally, there are two more papers that investigate the role of waiting/rationing for supplementary private health care (Hoel and Sæther 2003) and parallel private health insurance (Cuff et al. 2012), respectively.

ECON WPS

waiting time, while in Iversen the hospital moves first. It acts as a Stackelberg leader so as to maximize its profit by selecting the most favorable public waiting time / public capacity combination from the health authority's reaction function. Finally, the welfare analysis in Iversen is somewhat incomplete as he primarily concentrates on how the existence of a private sector affects public waiting times and how this relationship is mediated by the way the public waiting list is organized. Moreover, Iversen remains silent about whether or not dual practice should be allowed.

In terms of structure, our paper closely mirrors Brekke and Sørgard (2007). Both papers consider the same 5-stage sequential setting but there are several important departures in the modeling of dual practice that clearly differentiate our paper from theirs. Brekke and Sørgard adopt the representative consumer approach. While this framework certainly has its merits there is an important drawback – patient heterogeneity is not captured but buried in the utility function of the representative consumer. By contrast, we allow for patient heterogeneity by suggesting a Hotelling (1929) framework. This allows for different treatment decisions of patients (which depend on severity) and enables us to analyze the differential impact of market regulations on health outcomes and individual utility. When access to health care is an important objective, which we believe it is, accounting for individual heterogeneity seems rather important. Also, our framework builds on the empirical fact that patients typically have to make a discrete choice, that is, they consume public care or private care but not both. In Brekke and Sørgard the representative patient consumes both types of care and it is not entirely clear how the substitutability / complementarity between public and private care emerges in the aggregate. An additional difference between Brekke and Sørgard and our framework is that waiting time plays a key role in our model while labor supply is at the heart of the Brekke and Sørgard model. Finally, they consider an oligopoly where physicians compete in a Cournot fashion; we consider a monopolistic physician. Given these differences it is perhaps not surprising that the results of the two papers and their interpretation differ tremendously.

The remainder of the paper is organized as follows: The following Section 2 introduces the model; Section 3 derives the profit-maximizing allocation of private care and waiting, while Section 4 contrasts the laissez-faire outcome with the first-best allocation. In Section 5 we turn to a second-best analysis. We investigate the properties of the second-best allocation and derive conditions under which a ban of dual practice is socially desirable. Some model extensions are discussed in Section 6. Section 7 concludes.

ECON WPS

## 2 The model

We consider a population of patients that all suffer from a disease. A patient's propensity to benefit from treatment (or her severity) is denoted by $h$ and is uniformly distributed on the unit interval, $[0,1]$. The characteristic $h$ is private information of the patient, the physician only knows the distribution of $h$.[3]

The gross utility of a patient when cured is $(\theta - t)\,h > 0$, where $\theta$ and $t$ denote treatment intensity and a disutility from waiting and/or loss in treatment intensity, respectively. Two treatment options are available, a public and a private one. While the private treatment is more intensive than the public one, $\overline{\theta} > \underline{\theta}$, it is also more costly. Specifically, it is associated both with a financial cost, $c > 0$, and with a 'time' cost $\delta \geq 0$.[4] The administration of the private treatment takes longer, and if the physician is time constrained this implies either waiting times or reductions in treatment intensity (e.g. consultation time). Consequently, $\delta = 0$ describes a situation in which the physician's time constraint is not binding. The treatment technology is summarized as follows:[5]

| treatment | public | private |
|---|---|---|
| intensity | $\underline{\theta}$ | $\overline{\theta}$ |
| financial cost | 0 | $c > 0$ |
| time/waiting cost | 0 | $\delta \geq 0$ |

In the following, let us denote by $\Delta := \overline{\theta} - \underline{\theta} \geq 0$ the 'extra' intensity from the private treatment. Throughout our analysis we make the following assumptions

$$c \in \left(\frac{3\Delta}{5}, \Delta\right), \tag{A1}$$

$$\delta \in [0, 2(\Delta - c)]\,. \tag{A2}$$

Disregarding the waiting cost $t$ for the moment, $c < \Delta$ implies that it is always efficient to treat the most severe type with the intensive (private) technology. Furthermore, $\frac{\Delta}{2} <$

[3] To keep the model simple we refrain from incorporating costly diagnosing. Otherwise the agency problem (diagnosing effort and truthful reporting) would become dominant. The main argument we will be making here would remain, though.

[4] We discuss further on below how $\delta$ translates into the patient's disutility $t$.

[5] That the public treatment is costless both in terms of money and in terms of time is just a convenient normalization and none of the results depend on these assumptions.

ECON WPS

$\frac{3\Delta}{5} < c$ implies that it is not cost effective to introduce the intensive technology within the public domain, provided that no public patient can be denied access to this technology.[6],[7] We comment on the assumption (A2) further on below.

The physician works in the public system and receives an exogenous payment $w \geq 0$ per public treatment (or patient).[8] For the reason of cost-effectiveness outlined above only the less-intensive treatment is included in the public health care plan and offered within public practice. If allowed, the physician may also engage in private practice. Here, we assume that only the intensive treatment is offered privately. Although this assumption seems to be rather strong, it is well in line with existing health care systems. In Germany, for example, it is legally forbidden for panel doctors to charge public patients for services covered by statutory health insurance. Canada is another example, where public physicians may offer public services privately but must not charge a price higher than they would get reimbursed in the public system. There is thus no incentive to offer these services privately so that our specialization, public = non-intensive and private = intensive, results. As usual we assume that public treatment is free for patients while they have to pay the price for private care, $p$, out of pocket. Since health outcomes are better with the private treatment there will generally be a positive willingness to pay for private care.

In public health care systems patients often have to wait for treatment while there is no waiting time in the private system.[9] For the purpose of our analysis, we will endogenize the allocation of waiting time. Denoting by $x \in [0, 1]$ the share of private patients, we obtain a total time cost $\delta x$. We assume that a share $\psi \in [0, 1]$ of this time cost is allocated to public patients. Thus, $\psi = 1$ would imply full prioritization of private patients, whereas $\psi = 0.5$ would imply an equal sharing of the total time cost between the public and

[6] To see this, note that the average health benefits of the public and private treatments across all patients are given by $\underline{\theta} \int_0^1 h dh = \underline{\theta}/2$ and $\overline{\theta}/2$, respectively. The incremental average benefit from the private technology then amounts to $\Delta/2$ and, according to (A1), falls short of the average cost, $c$. Thus, introduction of the intensive technology into the public health plan would not be cost-effective.

[7] Note that for the purpose of this argument it is sufficient to assume $c \in \left(\frac{\Delta}{2}, \Delta\right)$. It will become evident further on below why we require the slightly more restrictive assumption (A1).

[8] We assume that each patient receives a single treatment (episode) so that we do not distinguish between treatments and patients.

[9] This is a standard assumption in the literature on public private health care financing (see, e.g., Gonzalez 2005) and is also supported empirically (for references see Garcia-Prado and Gonzalez 2007, 2011; and Socha and Bech 2011).

ECON WPS

private segment. Furthermore, we assume that within each segment the waiting cost is shared equally between patients. Thus, we obtain $t^{pub} = \frac{\psi\delta x}{1-x}$ and $t^{priv} = (1-\psi)\,\delta$ as the (expected) waiting cost a patient has to bear in the public and private practice, respectively. Notably, the public waiting cost is convex in the level of private demand. Besides raising the total time cost $\delta x$ a larger share of private patients also reduces the number of public patients across whom the waiting cost is spread. In contrast, within private practice time costs are born one-to-one by patients.[10]

The net benefit of patient $h$ from public and private care can then be written as

$$u^{pub} = \left(\underline{\theta} - t^{pub}\right) h = \left(\underline{\theta} - \frac{\psi\delta x}{1-x}\right) h, \tag{1}$$

$$u^{priv} = \left(\overline{\theta} - t^{priv}\right) h - p = \left[\overline{\theta} - (1-\psi)\,\delta\right] h - p. \tag{2}$$

As we wish to focus on the allocation in which all patients receive care, we need to guarantee that $u^{pub} \geq 0$ and $u^{priv} \geq 0$. The former condition implies

$$x \leq \frac{\underline{\theta}}{\underline{\theta} + \psi\delta} =: \overline{x}\,(\psi)\,, \tag{3}$$

a requirement we need to verify in the course of analysis. The latter condition, $u^{priv} \geq 0$, is always satisfied as $\overline{\theta} \geq \Delta > 2\,(\Delta - c) \geq \delta \geq (1-\psi)\,\delta$, where the second and third inequality follow from (A1) and (A2), respectively.

We analyze the impact and incentives of physician dual practice and its social desirability in the following sequential setting:

1. The health authority decides between a pure public health care system and a system with public-private health care financing (physician dual practice).

2. The health authority sets public sector reimbursement $w \geq 0$.

3. The physician offers the patient a choice between public and private treatment at price $p$.

4. The patient decides whether he wants to be treated in the public or the private practice.

[10] We should reiterate at this point that, although in the following we will continue to refer to 'waiting costs', in fact to patients the time cost $t$ may also reflect a reduction in treatment intensity (e.g.. shorter consultation hours).

5. Treatment is performed and payoffs realize.

Of course, should the health authority impose a ban on dual practice, stages 3 and 4 are inactive. As usual, the game is solved by backward induction leading to a subgame perfect Nash equilibrium.

## 3 Physician dual practice

We begin by studying the decentral allocation, in which the physician selects the price $p$ for the private treatment and allocates the waiting time according to $\psi$ and patients then decide on the type of treatment to receive. Thus, consider first a patient's treatment choice at stage 4. Offsetting against each other the benefits from private and public care, (2) and (1), respectively, we obtain the willingness to pay for private care

$$wtp = \left[\Delta + \left(t^{pub} - t^{priv}\right)\right] h = \left[\Delta + \frac{\delta\,(x - 1 + \psi)}{1 - x}\right] h. \tag{4}$$

Whenever there are some private treatments ($x > 0$), then the willingness-to-pay for private care increases in the fraction of waiting costs attributed to public patients, $\psi$. Interestingly, for $\psi > 0$, the willingness-to-pay for private care includes a positive network effect. The more patients are treated privately, $x$, the longer the public waiting list and, in turn, the higher the willingness to pay for private treatment. A monopolistic physician will, of course, use this network effect strategically in order to maximize profit. Whenever the price for private treatment $p$ is smaller than an individual's willingness to pay, the patient will opt for private care. Suppose that at a given price $p$ the individual with severity $\widehat{h}$ is indifferent between public and private care, that is,

$$p = \left[\Delta + \frac{\delta\,(x - 1 + \psi)}{1 - x}\right] \widehat{h}. \tag{5}$$

Then all individuals with a lower severity $h < \widehat{h}$ will strictly prefer public care, implying a demand $\widehat{h} = 1 - x$ for public care. Substituting into (4) and canceling terms we get the inverse demand function for private care

$$p(x, \psi) = (1 - x)\Delta + (x - 1 + \psi)\,\delta = \Delta - (1 - \psi)\,\delta - (\Delta - \delta)x. \tag{6}$$

Inverse demand is downward sloping if $\Delta > \delta$, which is satisfied by (A1) and (A2).[11] Notably, although the waiting costs within the public segment are convex in the level of

[11] Specifically, we have $\delta \leq 2\,(\Delta - c) < \Delta$, where the first and second inequality follow from (A2) and (A1), respectively.

ECON WPS

private demand, this does not translate into a convex inverse demand. Although high levels of private demand, $x$, may cause very high individual waiting costs within a small public segment, the patient types who remain within the public segment have a very low propensity to benefit from the treatment and, thus, to suffer from long waiting times (and/or reductions in treatment intensity).

The price for private treatment and the waiting times in the public and private sector announced at stage 2 are assumed to be consistent with one another in the sense that $p = p(x, \psi)$ according to (6). As the inverse-demand function is strictly monotone in $x$, this is without loss of generality even if the physician is only able to announce $p$ and $\psi$. The profit of the physician is given by

$$\Pi = (p(x, \psi) - c)\, x + w(1 - x) \tag{7}$$

and maximization with respect to $\psi$ and $x$ yields the optimal allocation of waiting times and share of private patients

$$\psi^* = 1 \tag{8}$$

$$x^* = \frac{\Delta - (c + w)}{2(\Delta - \delta)}. \tag{9}$$

The second order conditions are readily verified, given that $\Delta > \delta$. Furthermore, we need to ensure that $x^* \in [0, \overline{x}(1)]$, where $\overline{x}(\psi)$, as defined in (3), corresponds to the maximum size of private practice at which public patients still receive a non-negative benefit while being exposed to the full waiting cost ($\psi = 1$). Under (A1) and (A2) we obtain $x^* \in [0, \overline{x}(1)]$ if and only if $w \in [0, \Delta - c]$ and $\underline{\theta} > \frac{[\Delta - (c+w)]\delta}{\Delta - 2\delta + c + w}$.[12] The profit maximizing price is given by

$$p^* = \frac{1}{2}(\Delta + c + w). \tag{10}$$

Concerning the allocation of waiting, we see immediately that the physician has a strategic interest in prioritizing private patients. Not only does this directly raise the willingness-to-pay for private treatment by reducing $t^{priv}$ by an amount $\delta$; but also indirectly by increasing public waiting time $t^{pub}$ by an amount $\frac{\delta x}{1-x}$. Indeed, allocating waiting time

[12] Obviously, $w \leq \Delta - c$ implies $x^* \geq 0$, where $\Delta - c > 0$ is guaranteed by (A1). Furthermore, we obtain $x^* \leq \overline{x}(1)$, with $\overline{x}(1)$ as from (3), if and only if $c + w > 2\delta - \Delta$ and $\underline{\theta} > \frac{[\Delta - (c+w)]\delta}{\Delta - 2\delta + c + w}$. Noting that $c + w \geq c > 3\Delta - 4c \geq 2\delta - \Delta$, where the second and third inequality result from (A1) and (A2), respectively, it follows that $w \in [0, \Delta - c]$ and $\underline{\theta} > \frac{[\Delta - (c+w)]\delta}{\Delta - 2\delta + c + w}$ are, indeed, necessary and sufficient for $x^* \in [0, \overline{x}(1)]$.

ECON WPS

to the public segment is the more attractive the larger is the share of patients already attending private practice. Given full prioritization of private patients, the physician then selects the level of private supply that equates marginal revenue to marginal cost, including the opportunity cost of foregone public sector income, $w$. Indeed, this opportunity cost then shows up in the private sector fee, implying that more generous public reimbursement also triggers an increase in private fees. Interestingly, waiting cost, as measured by $\delta$, bears on the optimal private supply only through its effect on the slope of the inverse demand function (rendering it flatter) and, therefore, does not have a direct bearing on the private fee.

We conclude this discussion by briefly considering the comparative static properties of the decentral supply of private treatments

$$\begin{aligned} x_c^* = x_w^* &= \frac{-1}{2(\Delta - \delta)} < 0, \\ x_\delta^* &= \frac{\Delta - (c + w)}{2(\Delta - \delta)^2} > 0, \\ x_\Delta^* &= \frac{c + w - \delta}{2(\Delta - \delta)^2} \gtreqless 0. \end{aligned}$$

The negative impact of marginal cost and public reimbursement is immediately intuitive. Note at this stage that the response $x_w^*$ can be used by the policy-maker to control private provision. Similarly, it is unsurprising that greater time costs tend to increase the demand for private practice, given that it is fully prioritized. But surprisingly perhaps, the impact of (additional) private treatment intensity on private demand is ambiguous. Indeed, it is positive if and only if the time cost is not too large, i.e. if and only if $\delta < c+w$. In this case, as one would expect, the extra benefit from a greater private treatment intensity increases both private demand and the private fee the physician is able to charge. However, it cannot be ruled out that $\delta > c + w$.[13] In case of high waiting costs, a greater private treatment benefit leads to a smaller private practice. To understand the intuition for this counterintuitive finding consider the first-order condition $\frac{d\Pi}{dx} = p(x^*) - (c + w) + p'(x^*)x^* = 0$. While an increase in the intensity gap, $\Delta$, raises the equilibrium price, $\frac{dp(x^*)}{d\Delta} = 1 - x^* > 0$, it also increases the steepness of the (inverse) demand function, $\frac{d[p'(x^*)x^*]}{d\Delta} = -x^*$. As is readily verified, the net effect of $\Delta$ on marginal revenue is negative if $x^* \geq \frac{1}{2}$. From (9) we see that this is precisely true when $\delta > c + w$. In a setting where waiting costs within

[13] To see this, consider the case, where $w = 0$. Our assumptions (A2) and (A1) require that $\delta \leq 2(\Delta - c) \leq \frac{4}{3}c$. But then, for $\Delta \to \frac{5}{3}c$, a level of $\delta \in \left[c, \frac{4}{3}c\right]$ is feasible.

ECON WPS

the public sector stimulate high levels of private demand to begin with, an increase in the intensity gap may, therefore, trigger a reduction in demand. This notwithstanding, it can be readily shown that the physician's profit

$$\Pi\left(x^{*}\right) = w + \frac{\left[\Delta - (c+w)\right]^{2}}{4\left(\Delta - \delta\right)}$$

increases in both the intensity gap $\Delta$ and the time cost $\delta$. With regard to the intensity gap this implies that in a situation where high waiting costs generate a high level of private demand to begin with, the physician has an incentive to extract patient surplus from treatment intensity to an extent that it may even lower the level of demand.

Since waiting time is a strategic choice of the physician one wonders whether dual practice will ever be allowed in such a situation. The following section investigates the social desirability of dual practice and how it should be optimally structured.

## 4 The first-best allocation

The social objective function is the weighted sum of economic rents, with $\lambda \in [0,1]$ the utility weight of profit:

$$\begin{aligned} W &= \int_{0}^{1-x} \left[\left(\underline{\theta} - t^{pub}\right) h - w\right] dh + \int_{1-x}^{1} \left[\left(\overline{\theta} - t^{priv}\right) h - p\right] dh + \lambda \Pi\left(x\right) \\ &= \left(\underline{\theta} - t^{pub}\right) \frac{(1-x)^{2}}{2} + \left[\left(\overline{\theta} - t^{priv}\right) \frac{(2-x)}{2} - c\right] x - (1-\lambda)\,\Pi\left(x\right) \\ &= \frac{\underline{\theta}\,(1-x)^{2}}{2} + \left[\frac{\overline{\theta}\,(2-x)}{2} - c\right] x - \frac{\delta x}{2}\left(2 - x - \psi\right) - (1-\lambda)\,\Pi\left(x\right), \qquad (11) \end{aligned}$$

where the first term gives the gross benefit of the public treatment, the second term gives the net benefit of the private treatment (excluding waiting costs), the third term gives aggregate waiting costs and the fourth term weighted profit. Evidently, social welfare is (weakly) decreasing in profit, with only a full utilitarian ($\lambda = 1$) health authority being indifferent about the distribution of rents.

As a benchmark we consider the first-best optimum where the health authority can freely set the prices $p$ and $w$, quantity $x$ and the allocation of waiting time $\psi$. When normalizing the physicians's outside utility to zero we can then describe the health authority's problem by

$$\max_{w,p,\psi\in[0,1],x\in[0,1]} W \quad s.t. \quad w \geq 0, \quad p \geq c$$

ECON WPS

where $w \geq 0$ and $p \geq c$ guarantee the physician's participation in the public and private sector, respectively, and by implication $\Pi \geq 0$. As is readily checked from (11), a first-best allocation entails $w^{fb} = 0$, $p^{fb} = c$ and, thus, $\Pi = 0$. Unsurprisingly, the health authority has a (weakly) dominant incentive to extract all profit.

Furthermore, we obtain $W_\psi = \frac{\delta x}{2} \geq 0$ and, therefore, $\psi^{fb} = 1$. Hence, it is optimal to fully prioritize private patients. The reason is that in the set-up we are considering it is the most severe cases who select (or who are selected) into the private segment. At the same time it is the most severe cases who stand the most to gain from low waiting costs. As it is efficient to reduce waiting times for the most severe types to the minimum, it follows that it is private waiting which is cut to a minimum. Also note that the consequences to public patients are limited to a maximal loss of $\delta x$. Hence, while the physician strategically prioritizes private patients, for our particular set-up, this incentive coincides, albeit for different reasons, with the social one.

Finally, we obtain[14]

$$x^{fb} = \frac{\Delta - c - \delta/2}{\Delta - \delta}. \tag{12}$$

Again, it is readily verified that (A1) and (A2) ensure that the second-order condition holds, and, together with $\underline{\theta} > \frac{(\Delta - c - \delta/2)\delta}{c - \delta/2}$, guarantee that $x^{fb} \in [0, \overline{x}(1)]$.[15] The comparative statics

$$x_c^{fb} = \frac{-1}{\Delta - \delta} < 0,$$
$$x_\delta^{fb} = \frac{\Delta/2 - c}{(\Delta - \delta)^2} < 0,$$
$$x_\Delta^{fb} = \frac{c - \delta/2}{(\Delta - \delta)^2} > 0.$$

are intuitive.[16] The first-best supply of private care decreases in both the financial and the time cost of provision and it increases in the extra intensity afforded by the private treatment. Comparing the impact of the time cost, $\delta$, on the first-best as opposed to

[14] If, for whatever reason, $\psi$ is outside the planner's control, we have $x^{fb} = \frac{\Delta - c - (1+\psi)\delta/2}{\Delta - \delta}$, which illustrates that the first-best market share should be reduced with the extent to which private patients are subject to waiting costs.

[15] Here, $x^{fb} \geq 0$ follows immediately from (A2). Furthermore, $x^* \leq \overline{x}(1)$, with $\overline{x}(1)$ as from (3), if and only if $c > \frac{\delta}{2}$ and $\underline{\theta} > \frac{(\Delta - c - \delta/2)\delta}{c - \delta/2}$. Noting that $c > \Delta - c \geq \frac{\delta}{2}$, where the first and second inequality result from (A1) and (A2), respectively, it follows that $\underline{\theta} > \frac{(\Delta - c - \delta/2)\delta}{c - \delta/2}$ is, indeed, necessary and sufficient for $x^{fb} \in [0, \overline{x}(1)]$.

[16] Recall that $c > \frac{\Delta}{2} > \frac{\delta}{2}$ is implied by (A1) and (A2).

ECON WPS

the decentral supply of private care reveals the opposing incentives: While a higher time cost benefits the private physician and stimulates private practice, the health-authority would rather curtail private supply in order to contain the waiting costs imposed on public patients. However, the external waiting costs imposed on public patients only amount to a part of the gap between private and social incentives. In order to obtain a more complete picture consider the impact on social welfare of a marginal increase in the share of private patients evaluated at the private optimum $x = x^*$ when $\psi^* = \psi^{fb} = 1:$

$$\begin{aligned} \frac{dW}{dx}\,|_{x=x^*} &= \Delta\left(1-x^*\right) - c - \frac{\delta}{2}\left(1-2x^*\right) - \left(1-\lambda\right)\underbrace{\frac{d\Pi\left(x^*\right)}{dx}}_{=0} \\ &= p\left(x^*\right) - c - \frac{\delta}{2} = \frac{1}{2}\left(\Delta + w - c - \delta\right), \end{aligned} \tag{13}$$

where the second and third equalities follow when inserting, in turn, the inverse demand function (6) and the equilibrium price (10). Inspection of the second equality shows that the decentral allocation entails two offsetting distortions: On the one hand, market power, $p\left(x^*\right) - c > 0$, tends to restrain private supply below the social optimum; on the other hand, the physician ignores the external waiting cost, $\frac{\delta}{2}$, imposed on public patients. We then find $\frac{dW}{dx}\,|_{x=x^*} \geq 0 \Leftrightarrow \Delta + w - c \geq \delta$, implying for $w = 0$ that private supply falls short of the social optimum as long as the time cost does not exceed the benefit of private treatment net of its monetary cost, $\Delta - c$. Reimbursement in the public sector, $w > 0$, increases the tendency towards an under-supply of private practice in that it inflates the private-sector mark-up $p\left(x^*\right) - c$. For sufficiently high levels of the time cost, a situation of excessive private practice arises. This is easily checked for the limiting case, where for $\delta = 2\left(\Delta - c\right)$ it would be socially optimal to shut down private practice, whereas for $w \leq \Delta - c$ the physician will engage in it.[17]

For the purpose of further analysis it is convenient to contrast graphically the first-best supply $x^{fb}\left(\delta\right)$ and decentral supply $x^*\left(\delta, w\right)$ as depicted in Figure 1.[18]

[Insert Figure 1 about here]

When comparing the "laissez-faire" supply[19] $x^*\left(\delta, 0\right)$ to the social optimum $x^{fb}\left(\delta\right)$

[17] Intuitively, for $\frac{\delta}{2} \geq \Delta - c$ the (minimal) expected waiting cost per public patient exceeds the net benefit of private treatment, implying that it is never efficient to admit private practice.

[18] The slopes of the functions follow immediately from the comparative static results.

[19] "Laissez-faire" in the sense that the health authority provides no incentives through public reimbursement, i.e. $w = 0$.

ECON WPS

we see that there is under-provision of private practice for low levels of waiting cost, $\delta \in [0, \Delta - c)$, and over-provision for high levels, $\delta \in (\Delta - c, 2(\Delta - c)]$. Note that the tendency towards excessive provision for increasing levels of external costs $\delta$ is magnified by the increasing demand for private practice. Starting from the laissez-faire, the health-authority could in principle use the public-sector reimbursement, $w$, to align the private with the social incentive. Setting $\frac{dW}{dx}|_{x=x^*} = 0$ and solving from (13) for the corresponding reimbursement, we obtain $\widehat{w}(\delta) = -(\Delta - c - \delta)$. Thus, the public reimbursement is set equal to the *negative* net value of a private treatment including the full time cost. Obviously, this would imply a negative reimbursement $\widehat{w}(\delta) < 0$ for positive net values of the private treatment $\Delta - c - \delta > 0$. In order to provide an incentive to the physician to expand private practice beyond the laissez-faire level, public provision would need to be penalized. Such a policy, however, is infeasible, as it would lead to the shut-down of the public practice. If time cost is such that $\Delta - c - \delta = 0$, the health-policy maker could establish the first-best allocation at no cost, $\widehat{w}(\Delta - c) = 0$. Nevertheless, it should be noted that social welfare falls short of the first-best whenever $\lambda < 1$ since private treatments are priced above marginal cost. Finally, for a negative value of the private treatment $\Delta - c - \delta < 0$, the health-authority could, in principle, establish the first-best allocation of patients by setting a public reimbursement, $\widehat{w}(\delta) = -(\Delta - c - \delta) > 0$, which induces the physician to reduce private practice to its first-best level, $x^*(\delta, \widehat{w}(\delta)) = x^{fb}(\delta)$. One such example is depicted in figure 1 for some $\delta' \in [\Delta - c, 2(\Delta - c)]$. We can summarize as follows.

**Proposition 1** *(i) For $\delta \in [0, \Delta - c)$, there is an under-supply of private practice which cannot be mitigated. (ii) For $\delta \in (\Delta - c, 2(\Delta - c)]$, there is an over-supply of private practice at the"laissez-faire" level of public reimbursement $w = 0$, while the first-best can be induced by setting an appropriate public reimbursement $\widehat{w}(\delta) > 0$.*

This said, it is not clear a priori whether in a second-best context the health-authority has an interest in inducing the first-best level of private practice even if this is feasible. To study this question we turn now to the second-best analysis.

# 5 Second-best analysis

## 5.1 Second-best with dual practice

In many real world cases, the health authority has no means of selecting patients into private practice, neither directly by choice of $x$ nor by determining the price for private treatments, $p$. Nevertheless, they can exert some indirect influence on private provision by varying the reimbursement, $w$, within the public domain. Formally, the health authority faces the problem

$$\max_w W \quad s.t. \quad w \geq 0, \quad x = x^*(\delta, w), \quad p = p(x), \quad \psi = \psi^* = \psi^{fb} = 1$$

where the share of private provision is determined by the physician's best-response (9) and the price for private treatments follows according to the inverse demand function (6). The allocation of waiting times is also determined by the physician, but as we have argued, this poses no problem as it corresponds to the first-best. The first-order condition for this problem reads[20]

$$\begin{aligned} \frac{dW}{dw} &= \frac{dW}{dx}\Big|_{x=x^*(\delta,w)} x_w^* - (1-\lambda)\frac{d\Pi(x^*)}{dw} \\ &= \frac{1}{2}(\Delta + w - c - \delta)\, x_w^* - (1-\lambda)(1-x^*) \\ &= \frac{-(\Delta + w - c - \delta)}{4(\Delta-\delta)} - \frac{(1-\lambda)(\Delta - 2\delta + c + w)}{2(\Delta-\delta)} = 0. \end{aligned}$$

The last equality implies

$$w = \frac{-[\Delta - c - \delta + 2(1-\lambda)(\Delta - 2\delta + c)]}{3-2\lambda} =: \widehat{\widehat{w}}(\delta, \lambda). \qquad (14)$$

It is readily verified that $\widehat{\widehat{w}}_\lambda = \frac{2(2c-\delta)}{(3-2\lambda)^2} > 0$ and $\widehat{\widehat{w}}(\delta, 1) = \widehat{w}(\delta)$. Hence, we obtain the following result.

**Lemma 1** *$\widehat{\widehat{w}}(\delta,\lambda) \leq \widehat{\widehat{w}}(\delta,1) = \widehat{w}(\delta)$ and $x^*\left(\delta, \widehat{\widehat{w}}(\delta,\lambda)\right) \geq x^*\left(\delta, \widehat{\widehat{w}}(\delta,1)\right) = x^*(\delta, \widehat{w}(\delta)) = x^{fb}(\delta)$, with strict equalities for $\lambda = 1$ and strict inequalities otherwise.*

Any non-utilitarian health authority has an incentive to depress the public reimbursement in order to minimize the rents paid to the physician. In so doing they are prepared to

[20] The second-order condition is satisfied by $\frac{d^2W}{dw^2} = \frac{-(3-2\lambda)}{4(\Delta-\delta)} < 0$.

ECON WPS

accept a level of private practice that is strictly greater than the first-best. Only a strictly utilitarian health authority will be prepared to implement the first-best level of private practice wherever this is feasible. This said, it is evident that the fee $\widehat{\widehat{w}}\left(\delta,\lambda\right)$ is not always implementable. Indeed, as is readily verified, $\widehat{\widehat{w}}_{\delta} > 0$ and $\widehat{\widehat{w}}\left(\Delta - c, \lambda\right) \leq \widehat{w}\left(\Delta - c\right) = 0$. Hence, there exists some boundary $\widehat{\delta}\left(\lambda\right) = \frac{\Delta - c + 2(1-\lambda)(\Delta + c)}{5 - 4\lambda}$ with $\widehat{\delta}\left(\lambda\right) \geq \Delta - c$; $\widehat{\delta}\left(\lambda\right) \leq 2\left(\Delta - c\right) \Leftrightarrow \lambda \geq \frac{11c - 7\Delta}{2(5c - 3\Delta)}$; and $\widehat{\delta}_{\lambda} < 0$.[21] The following is then true for the second-best allocation.

**Proposition 2** *If* $\lambda \in \left[\frac{11c-7\Delta}{2(5c-3\Delta)}, 1\right]$, *the second-best allocation is given as follows: (i)* $w^{sb} = 0$ *and* $x^{sb} = x^{*}\left(\delta, 0\right) < x^{fb}\left(\delta\right)$ *for* $\delta \in \left[0, \Delta - c\right)$; *(ii)* $w^{sb} = 0$ *and* $x^{sb} = x^{*}\left(\delta, 0\right) > x^{fb}\left(\delta\right)$ *.for* $\delta \in \left(\Delta - c, \widehat{\delta}\left(\lambda\right)\right]$; *and (iii) by* $w^{sb} = \widehat{\widehat{w}}\left(\delta, \lambda\right) > 0$ *and* $x^{sb} = x^{*}\left(\delta, \widehat{\widehat{w}}\left(\delta, \lambda\right)\right) > x^{fb}\left(\delta\right)$ *.for* $\delta \in \left(\widehat{\delta}\left(\lambda\right), 2\left(\Delta - c\right)\right]$. *(iv) If* $\lambda < \frac{11c-7\Delta}{2(5c-3\Delta)}$, *the second-best allocation is given by* $w^{sb} = 0$ *and* $x^{sb} = x^{*}\left(\delta, 0\right)$ *for all* $\delta \in \left[0, 1\right]$, *with* $x^{sb} < x^{fb}\left(\delta\right) \Leftrightarrow \delta \in \left[0, \Delta - c\right)$ *and* $x^{sb} > x^{fb}\left(\delta\right) \Leftrightarrow \delta \in \left(\Delta - c, 1\right]$.

Consider first the case, where $\lambda \in \left(\frac{11c-7\Delta}{2(5c-3\Delta)}, 1\right]$ implies the existence of three 'regimes' as described in (i)-(iii) of the Proposition. Figure 2 illustrates the following argument by depicting the supply function $x^{*}\left(\delta, 0\right)$ as well as the first- and second-best levels of private practice, $x^{fb}\left(\delta\right)$ and $x^{sb}$, respectively.[22]

[Insert Figure 2 about here]

For low levels of the time cost, $\delta < \Delta - c$, the health authority would like to set a negative public sector reimbursement for reason of both, inducing additional provision of private practice and extracting surplus from the physician. As such a policy is infeasible, the best the health authority can do is to set $w^{sb} = 0$ and, thereby, accept the laissez-faire level of private provision although it is falling short of the first-best. For intermediate levels of the time cost, $\delta \in \left(\Delta - c, \widehat{\delta}\left(\lambda\right)\right]$,a non-utilitarian health-authority ($\lambda < 1$) continues to set $w^{sb} = 0$ even if by now the laissez-faire supply $x^{*}\left(\delta, 0\right)$ exceeds the first-best level of private practice. For high levels of the time cost $\delta \in \left(\widehat{\delta}\left(\lambda\right), 2\left(\Delta - c\right)\right]$ even a

[21] Note that $\frac{11c-7\Delta}{2(5c-3\Delta)} \leq 0$ if $c \in \left[\frac{3}{5}\Delta, \frac{7}{11}\Delta\right]$ and $\frac{11c-7\Delta}{2(5c-3\Delta)} \leq 1$ for all $c \leq \Delta$.

[22] Recall that $x^{sb} = x^{*}\left(\delta, 0\right)$ for $\delta < \widehat{\delta}\left(\lambda\right)$ and $x^{sb} = x^{*}\left(\delta, \widehat{\widehat{w}}\left(\delta, \lambda\right)\right)$ for $\delta \geq \widehat{\delta}\left(\lambda\right)$, where $x^{sb}_{\delta} = x^{*}_{\delta}\left(\delta, \widehat{\widehat{w}}\left(\delta, \lambda\right)\right) + x^{*}_{w}\left(\delta, \widehat{\widehat{w}}\left(\delta, \lambda\right)\right)\widehat{\widehat{w}}_{\delta}\left(\delta, \lambda\right) < 0$ can be verified.

non-utilitarian health authority sets a positive public reimbursement in order to constrain the supply of private practice and, thereby, the external costs imposed on public patients. Nevertheless, the interest in extracting physician surplus leads the health-authority to restrain public sector payments and, thereby, trade-off the reduction of externalities against the extraction of rents: Albeit below the laissez-faire level, private practice continues to be over-supplied. Note that the excess supply of private practice for $\delta > \Delta - c$ diminishes with $\lambda$ and vanishes for $\lambda = 1$ : a utilitarian health authority always constrains private supply to the first-best level whenever this is appropriate even if this means the payment of high rents to the physician.

If $\lambda < \frac{11c-7\Delta}{2(5c-3\Delta)}$ as in (iv) of the Proposition, then the health-authority's interest in extracting surplus always leads them to set a zero public reimbursement $w^{sb} = 0$ and, thereby, admit the laissez-faire solution. Private practice is then under-provided if and only if $\Delta - c - \delta > 0$, i.e. if and only if the net benefit of private treatment, including its time cost is positive. Otherwise private pracice is over-provided at maximal extent.

## 5.2 Banning dual practice

We have not yet analyzed the health authority's decision on whether or not to admit dual practice in the first place. Indeed, it may be in the interest of the health authority to ban dual practice altogether. Inspection of Figure 2 reveals that a ban on private practice improves social welfare whenever time costs are very high and the health authority is not fully utilitarian. Consider a situation, where for $\delta \to 2(\Delta - c)$ it would be socially optimal to reduce private provision to (an almost) zero level. In principle, this allocation can be attained by setting a public reimbursement $\widehat{w}(\delta) \to \Delta - c > 0$, but such a policy leaves maximal rents $\widehat{w}(\delta)[1 - x^*(\delta, \widehat{w}(\delta))] \to \widehat{w}(\delta) \to \Delta - c$ to the physician. While a non-utilitarian planer will therefore admit a positive second-best level of private provision, $x^{sb} > 0$, in a move to reduce the physician's rent, this implies a welfare loss on two margins: a direct loss from rental payments and an indirect loss due to the external waiting costs associated with excessive private provision. In such a case it is always efficient to ban private practice, as the first-best level of private practice (equal to zero) can be attained while, at the same time, the public remuneration can be reduced to zero. Therefore, a ban of dual practice (weakly) dominates the regulation by way of fees at the highest level of time costs. However, this result is less clear-cut at lower levels of $\delta$, as a ban now implies that (some) social value of private treatment is foregone.

ECON WPS

To arrive at a complete solution, we note that the health authority bans dual practice whenever

$$\Omega(\delta, \lambda) :\equiv W_0 - W^{sb}(\delta, \lambda) > 0,$$

where $W_0 := \frac{\underline{\theta}}{2}$ measures social surplus for the case in which dual (or, indeed, private) practice is banned, implying $x \equiv 0$, and the public treatment is offered at $w = 0$; and where

$$W^{sb}(\delta, \lambda) = \frac{\underline{\theta}\left(1 - x^{sb}\right)^2}{2} + \left[\frac{\overline{\theta}\left(2 - x^{sb}\right)}{2} - c\right] x^{sb} - \frac{\delta x^{sb}}{2}\left(1 - x^{sb}\right) - (1 - \lambda)\,\Pi\left(x^{sb}, w^{sb}\right)$$

measures social surplus for the case in which dual practice is allowed and a second-best allocation is implemented.[23]. After substitution and rearrangement we then obtain

$$\Omega(\delta, \lambda) :\equiv (1 - \lambda)\,\Pi\left(x^{sb}, w^{sb}\right) - (\Delta - \delta)\left[x^{fb}(\delta) - \frac{x^{sb}}{2}\right] x^{sb}. \tag{15}$$

Thus, broadly speaking, a ban is warranted if the loss to a non-utilitarian health authority of leaving rents to the physician exceeds a measure of the net benefit from private practice. In the Appendix we establish the following result.

**Proposition 3** *There exists a function $\widetilde{\delta}(\lambda)$ with $\widetilde{\delta}_\lambda(\lambda) > 0$, $\widetilde{\delta}(0) = \frac{\Delta - c}{2}$ and $\widetilde{\delta}(1) = 2(\Delta - c)$ such that a health-authority characterized by $\lambda \in [0, 1]$ prefers to ban dual practice if and only if $\delta > \widetilde{\delta}(\lambda)$.*

As one would expect, non-utilitarian health authorities ($\lambda < 1$) will ban dual practice if the externalities associated with the time/waiting cost $\delta$ are sufficiently high. Furthermore, the lower the weight on physician profits, the larger the range of waiting costs at which the health authority is willing to ban private practice. Given that $\widetilde{\delta}(\lambda) \in \left[\frac{\Delta - c}{2}, 2(\Delta - c)\right]$ we see that essentially two scenarios can arise, which are depicted in Figures 3 and 4.

[Insert Figures 3 and 4 about here]

Figure 3 illustrates a situation where for a high level of $\lambda$ we have $\widetilde{\delta}(\lambda) > \widehat{\delta}(\lambda) > \Delta - c$. As the figure reveals the second-best level of private care is then characterized by three regimes: For $\delta \in [0, \Delta - c)$ there is an under-supply of private care at the laissez-faire

[23] Recall for this case that $\psi^{sb} = 1$.

ECON WPS

level $x^*(\delta, 0)$; for $\delta \in \left(\Delta - c, \widetilde{\delta}(\lambda)\right]$ there is an over-supply of private care;[24] and for $\delta \in \left(\widetilde{\delta}(\lambda), 2(\Delta - c)\right]$ there is again an under-supply of private care owing to a ban. Thus, there is under-supply of private care both for low and high levels of the externality. Note, however, that the under-supply at the low end results from the inability of the health-authority to stimulate additional private supply in the presence of market power; whereas the under-supply at the high end results from the rents that have to be left to the physician to provide an incentive for downsizing private provision exceeding the full surplus from private treatment. We conclude our discussion of this first scenario by noting that a utilitarian health authority ($\lambda = 1$) never bans dual practice and, indeed, induces the first-best supply for $\delta \in [\Delta - c, 2(\Delta - c)]$. As we have argued before, this is because a utilitarian health authority is indifferent about the allocation of surplus and therefore willing to implement the first-best allocation whenever this is feasible.

Figure 4 illustrates a situation where for a low level of $\lambda$,we have $\widehat{\delta}(\lambda) > \Delta - c > \widetilde{\delta}(\lambda) \geq \frac{\Delta - c}{2}$.[25] In this case, the second-best level of private care is characterized by under-provision throughout: for $\delta \in (0, \widetilde{\delta}(\lambda)]$, private care is under-supplied at laissez-faire level; for $\delta \in (\widetilde{\delta}(\lambda), 2(\Delta - c)]$ private care is under-supplied due to the ban. This allocation may be surprising in as far as dual practice is banned even at levels of the time cost $\delta \in (\widetilde{\delta}(\lambda), \Delta - c]$, where the second-best supply of private care lies below the first-best. Thus, by banning dual practice the health authority is worsening the under-supply of private care. The rationale for such a seemingly counter-intuitive decision can be understood as follows. Consider for the sake of illustration a health authority who at $\lambda = 0$ is only interested in patient benefit and the public purse. Consider further some $\delta \in \left[\frac{\Delta - c}{2}, \Delta - c\right]$ at which $w^{sb} = 0$ and private care is under-supplied at laissez-faire level $x^*(\delta, 0)$. To understand why it is nevertheless efficient to ban private practice, we note that in contrast to the first-best perspective, where the first unit of private provision is evaluated at $\Delta - c - \delta$, the health authority evaluates private provision from the patients' perspective and, thus, at $\Delta - p^* - \delta$. Recalling from (10) that for $w^{sb} = 0$ we have $p^* = \frac{\Delta + c}{2}$, private provision is now evaluated at $\frac{\Delta - c}{2} - \delta$. Hence, for $\delta \in \left[\frac{\Delta - c}{2}, \Delta - c\right]$ additional private treatments are valuable from a first-best perspective but not from the patients' perspective once the time

[24] For $\delta \in \left(\Delta - c, \widehat{\delta}(\lambda)\right]$ at laissez-faire level $x^*(\delta, 0)$ and for $\delta \in \left(\widehat{\delta}(\lambda), \widetilde{\delta}(\lambda)\right]$ at a level below the laissez-faire $x^*\left(\delta, w^{sb} > 0\right)$.

[25] We have omitted the case, where $\widehat{\delta}(\lambda) > \widetilde{\delta}(\lambda) > \Delta - c$. This case generates the same pattern of (under-then-over-then-under) supply of private care as the previous case, the only difference being that $w^{sb} = 0$ holds for all $\delta$.

ECON WPS

cost is internalized.

While on average patients benefit from a ban under the circumstances just described, it is instructive to consider the change in surplus for different types of patients. Public patients always benefit as they are relieved from the waiting costs $t^{pub} = \frac{\delta x^{sb}}{1-x^{sb}}$. For private patients a more diverse picture emerges:

**Proposition 4** *Banning dual practice benefits all public patients and a share of private patients with (relatively) low willingness to pay. It always harms private patients with (relatively) high willingness to pay and is, therefore, not Pareto efficient even from a patient perspective.*

To see this compare the change in surplus, $p - \Delta h$, for a private patient with propensity to benefit from treatment $h$ when dual practice is banned. While the patient saves the cost of private treatment they lose the benefit from more intensive treatment. Substituting from the indifference condition (5) we obtain

$$p - \Delta h = \left[\Delta + \frac{\delta x}{1-x}\right]\widehat{h} - \Delta h = \Delta\left(\widehat{h} - h\right) + \delta\left(1 - \widehat{h}\right),$$

where the second equality follows, when recalling $\widehat{h} = 1 - x$. Thus, we have

$$p - \Delta h \geq 0 \Leftrightarrow h \leq \widehat{h} + \frac{\delta\left(1 - \widehat{h}\right)}{\Delta}. \tag{16}$$

From this, we see immediately that type $\widehat{h}$ who is indifferent between the public and private treatment under dual practice is rendered unambiguously better off by a ban. The reason is that the avoidance of waiting costs in a regime of dual practice has increased the willingness to pay over and above the 'pure' benefit from more intensive treatment $\Delta h$. As a ban of dual practice eliminates all waiting costs, the marginal patient is, therefore, gaining surplus amounting to the willingness to pay for the avoidance of waiting. Note that this gain is the larger, the larger the size of private practice in a dual regime. By continuity a range of patients with $h > \widehat{h}$ also benefit from the ban. However, this does not include the patients with the highest willingness to pay. To see this consider the patient with the highest propensity to benefit $h = 1$. As is readily verified, the inequality in (16) does not hold since $\Delta > \delta$. Hence, while benefiting some private patients a ban is never Pareto optimal from a patient perspective.

ECON WPS

# 6 Discussion

In a later version of the paper we will discuss some model extensions. We can show, for instance, that a subsidy improves the allocation should private care be undersupplied. Nevertheless, unless the health authority is utilitarian, the first-best allocation will not be implemented as a second-best optimum. Upper bounds on public wait times also prove to be effective. Finally, we will argue that our results extend to a framework where private treatment quality is endogenous.

# 7 Conclusion

We developed a simple model of dual practice where physicians can self-refer patients from the public to their private practice. Although patients have to pay for private care out of pocket they may prefer this treatment option over free public treatment. This is likely when a patient is prone to benefit from the more intensive private treatment and when public sector waiting times are long. With regard to the latter, a network effect arises: by inducing additional patients to consume the more time intensive private treatment, public waiting times tend to reinforce themselves. If the improvement in health outcomes is sufficiently large as compared to the costs of dual practice (higher treatment costs and induced waiting times) then dual practice is socially desirable. In the special case without waiting cost dual practice always dominates a pure public system. Indeed, in such a case private practice is under-supplied by a monopolistic physician. With increasing waiting costs the under-supply of private practice overturns into an excessive supply. This is because the physician - as, indeed, patients going private - do not internalize the waiting costs imposed on patients remaining with the public sector. Interestingly, this problem could be resolved, in principle, as long as the health-authority is willing to pay a public reimbursement at a level that is sufficient to induce the physician to treat patients publicly rather than privately. In reality, however, the scarcity of public funds and/or the preference of patients' over physicians' interests will typically induce the health authority to allow a certain amount of excess supply (and excessive waiting) in a move to contain both private and public payments to the physician. A strong(er) interest in patient welfare will then lead the health authority to ban dual practice at high levels of the waiting cost. With a strong focus on patient benefit this will even be the case if there is an under-supply of private practice relative to the first-best. Even if public sector payments are

ECON WPS

at their lowest (in our model: zero) level, a patient-oriented health authority takes into consideration that patient surplus always falls below total surplus by an amount equal to the mark-up on private treatments. Thus, they are willing to shut-down dual practice even at comparatively low levels of the external cost. Interestingly, a ban of dual practice will benefit even some private patients, namely those who have purchased private care for the predominant reason of avoiding public sector waiting costs.

When interpreting the social welfare weight of profit as determined by the political influence of physicians, our results also allow a political economy interpretation: While dual practice is always more likely to be allowed by health-authorities subject to greater influence of physicians, such health authorities are also better able to contain waiting costs by restricting the extent of private practice. This is because physician influence guarantees levels of public remuneration that render private practice a relatively less attractive option.

A number of limitations and potential extensions are worth mentioning. First, in the absence of waiting costs in the public system ($\delta = 0$) dual practice is always socially desirable, because there is no externality of an individual's decision to purchase private care. This result is based, however, on our assumption that patients are perfectly informed about the benefits of treatment. If patients are imperfectly informed about the benefits from private treatment, the physician may induce private demand even if this does not yield a sufficient benefit. In such a situation we would expect over-treatment and over-provision of private care. Second, when physicians are not pure profit maximizers but also derive utility from the patients' well-being, then private care is likely to be priced at lower levels. Whether or not this exacerbates a problem of over-provision depends on the extent to which the physician takes due account of the waiting costs. Overall, however, dual practice becomes more desirable and easier to regulate to the extent that physician surplus is lower. Third, we did not consider additional aspects of practice under the physician's control, such as diagnosing, patient selection and quality incentives. Selection may matter, in particular, if patients differ not only in their propensity to benefit from treatment but also in the costs of treatment. Quality incentives tend to induce additional bias towards private practice. Suppose for instance, that the physician can increase the intensity and, thus, the patient's benefit from private treatment. If the time cost of private treatment also increases in intensity, this typically implies that the physician invests excessively in the intensity of private treatment: Not only will a profit-maximizing physician disregard the increase in waiting costs imposed on public patients, but she will have an active interest in raising waiting costs as this increases demand. Thus, there is a clear incentive to over-

ECON WPS

invest in the intensity of private treatment. Fourth, we have not considered the effect of competition for patients as it may arise, in particular, within the private segment. To the extent that competition erodes the profit from private treatment this mitigates the problem of under-supply in the presence of low waiting costs, but at the same time exacerbates the problem of over-supply when waiting costs are high. In particular, this holds true if low or zero profit margins within the private segment constrain the health-authority from raising reimbursement within the public segment even if this was sought for. Finally, the health authority may, in some cases, have additional instruments at its disposition. For instance, it may be able to impose ceilings on public sector waiting times and/or it may be able to provide demand-side incentives by subsidizing or taxing the consumption of private care. We leave all of these extension for future analysis.

## References

[1] **Barros PP, Olivella P**: "Waiting Lists and Patient Selection," *Journal of Economics and Management Strategy*, 2005, 14(3), 623-646.

[2] **Bir A, Eggleston K**: "Physician Dual Practice: Access Enhancement or Demand Inducement?" Tufts University, Working Paper, 2003.

[3] **Brekke KR, Sørgard L**: "Public Versus Private Health Care in a National Health Service," *Health Economics*, 2007, 16, 579-601.

[4] **Biglaiser G, Ma A**: "Moonlighting: Public Service and Private Practice," *RAND Journal of Economics*, 2007, 38(4), 1113-1133.

[5] **Cuff K, Hurley J, Mestelman S, Muller RA, Nuscheler R**: "Public and Private Health Care Financing under Alternate Public Rationing Rules," *Health Economics*, 2012, 21, 83-100.

[6] **Eggleston K, Bir A**: "Physician Dual Practice," *Health Policy*, 2006, 78, 157-166.

[7] **Garcia-Prado A, Gonzalez P**: "Policy and Regulatory Responses to Dual Practice in the Health Sector," *Health Policy*, 2007, 84, 142-152.

[8] **Garcia-Prado A, Gonzalez P**: "Whom Do Physicians Work For? An Analysis of Dual Practice in the Health Sector," *Journal of Health Politics, Policy and Law*, 2011, 36, 265-294.

ECON WPS

[9] **Gonzalez P**: "Should Physicians' Dual Practice be Limited? An Incentive Approach," *Health Economics*, 2004, 13(6), 505-524.

[10] **Gonzalez P**: "On a Policy of Transferring Public Patients to Private Practive," *Health Economics*, 2005, 14, 513-527.

[11] **Gonzalez P, Macho-Stadler I**: "A Theoretical Approach to Dual Practice Regulations in the Health Sector," *Journal of Health Economics*, 2013, 32, 66-87.

[12] **Hoel M, Sæther EM**: "Public Health Care with Waiting Time: The Role of Supplementary Private Health Care," *Journal of Health Economics*, 2003, 22(4), 599-616.

[13] **Hotelling H**: "Stability in Competition," *Economic Journal*, 1929, 39, 41-57.

[14] **Iversen T**: "The Effect of a Private Sector on the Waiting Time in a National Health Service," *Journal of Health Economics*, 1997, 16, 381-396.

[15] **Socha KZ, Bech M**: "Physician Dual Practice: A Review of the Literature," *Health Policy*, 2011, 102, 1-7.

# Appendix

## 8 Proof of Proposition 3

To establish the existence of a boundary value $\widetilde{\delta}\left(\lambda\right)$, with $\widetilde{\delta}_{\lambda}\left(\lambda\right) > 0$, $\widetilde{\delta}\left(0\right) = \frac{\Delta - c}{2}$ and $\widetilde{\delta}\left(1\right) = 2\left(\Delta - c\right)$, such that a ban is preferred for $\delta > \widetilde{\delta}\left(\lambda\right)$, we need to show (i) that $\Omega_{\delta}(\delta, \lambda) > 0$ and $\Omega_{\lambda}(\delta, \lambda) < 0$ for all $\delta \in [0, 2\left(\Delta - c\right)]$ and $\lambda \in [0, 1]$; (ii) that $\Omega(\frac{\Delta - c}{2}, \lambda) \leq 0 \leq \Omega(2\left(\Delta - c\right), \lambda)$ for all $\lambda \in [0, 1]$, and (iii) that $\Omega(\frac{\Delta - c}{2}, \lambda) = 0$ and $\Omega(2\left(\Delta - c\right), \lambda) = 0$ hold for $\lambda = 0$ and $\lambda = 1$, respectively. Note that together with $\Omega_{\delta}(\delta, \lambda) > 0$ (ii) then implies the existence of a unique boundary value $\widetilde{\delta}\left(\lambda\right)$ such that $\Omega(\delta, \lambda) \Leftrightarrow \delta > \widetilde{\delta}\left(\lambda\right)$, while (i) implies that $\widetilde{\delta}_{\lambda}\left(\lambda\right) = -\frac{\Omega_{\lambda}(\widetilde{\delta}(\lambda), \lambda)}{\Omega_{\delta}(\widetilde{\delta}(\lambda), \lambda)} > 0$. Finally, (iii) implies $\widetilde{\delta}\left(0\right) = \frac{\Delta - c}{2}$ and $\widetilde{\delta}\left(1\right) = 2\left(\Delta - c\right)$. We now prove (i)-(iii) in turn.

ECON WPS

**Part (i):** From (15) we obtain (after rearranging) the derivatives[26]

$$\Omega_{\delta}(\delta,\lambda) = (1-\lambda)\left(1-x^{sb}\right)w_{\delta}^{sb} - (\Delta-\delta)\left[x^{fb}(\delta) - x^{sb}\right]x_{\delta}^{sb} + \frac{\left(1-x^{sb}\right)x^{sb}}{2}$$

$$\Omega_{\lambda}(\delta,\lambda) = -\Pi\left(x^{sb},w^{sb}\right) + (1-\lambda)\left(1-x^{sb}\right)w_{\lambda}^{sb} - (\Delta-\delta)\left[x^{fb}(\delta) - x^{sb}\right]x_{\lambda}^{sb},$$

which a priori cannot be signed without ambiguity. In the following, we need to distinguish two cases:

- $\delta \in \left[0,\widehat{\delta}(\lambda)\right)$ : Here, we have $w^{sb} = 0$ and $x^{sb} = x^{*}(\delta,0)$, with $x^{*}(\cdot)$ as defined in (9), implying $w_{\delta}^{sb} = w_{\lambda}^{sb} = 0$, as well as $x_{\delta}^{sb} = x_{\delta}^{*}(\delta,0) = \frac{x^{*}(\delta,0)}{\Delta-\delta} > 0$ and $x_{\lambda}^{sb} = 0$. For this case, we then obtain

$$\begin{aligned}
\Omega_{\delta}(\delta,\lambda) &= -(\Delta-\delta)\left[x^{fb}(\delta) - x^{*}(\delta,0)\right]x_{\delta}^{*}(\delta,0) + \frac{\left[1-x^{*}(\delta,0)\right]x^{*}(\delta,0)}{2} \\
&= \left[-x^{fb}(\delta) + x^{*}(\delta,0) + \frac{1-x^{*}(\delta,0)}{2}\right]x^{*}(\delta,0) \\
&= \left[-2x^{fb}(\delta) + x^{*}(\delta,0) + 1\right]\frac{x^{*}(\delta,0)}{2} \\
&= (3c-\Delta)\,\frac{x^{*}(\delta,0)}{4(\Delta-\delta)} > 0,
\end{aligned}$$

where the last equality follows when inserting from (9) and (12) and summarizing terms appropriately, and where the inequality follows from (A1). Furthermore,

$$\Omega_{\lambda}(\delta,\lambda) = -\Pi\left(x^{sb},w^{sb}\right) < 0.$$

- $\delta \in \left[\widehat{\delta}(\lambda), 2(\Delta-c)\right]$ : [27] Here, we have $w^{s}b = \widehat{\widehat{w}}(\delta,\lambda)$ as defined in (14) and $x^{s}b = x^{*}(\delta,\widehat{\widehat{w}}(\delta,\lambda))$, implying $w_{\delta}^{s}b = \frac{5-4\lambda}{3-2\lambda} > 0$ and $w_{\lambda}^{s}b = \frac{2(2c-\delta)}{(3-2\lambda)^{2}} > 0$, as well as[28]

$$\begin{aligned}
x_{\delta}^{sb} &= x_{\delta}^{*}(\delta,\widehat{\widehat{w}}(\delta,\lambda)) + x_{w}^{*}(\delta,\widehat{\widehat{w}}(\delta,\lambda))w_{\delta}^{sb} \\
&= \frac{\Delta-c-\widehat{\widehat{w}}(\delta,\lambda)}{2(\Delta-\delta)^{2}} - \frac{1}{2(\Delta-\delta)}\,\frac{5-4\lambda}{3-2\lambda} < 0
\end{aligned}$$

[26] Observe that $\Pi_{x}\left(x^{sb},w^{sb}\right) = 0$ from the envelope theorem and $\Pi_{w}\left(x^{sb},w^{sb}\right) = 1-x^{sb}$.

[27] We assume here that $\lambda \in \left(\frac{11c-7\Delta}{2(5c-3\Delta)},1\right]$, implying that $\widehat{\delta}(\lambda) \leq 2(\Delta-c)$. For $\lambda \leq \frac{11c-7\Delta}{2(5c-3\Delta)}$ we are always with the previous case.

[28] The inequality in the second line is verified when inserting from (14), collecting and canceling terms and observing (A1).

ECON WPS

and $x_\lambda^{sb} = x_w^*(\delta, \widehat{\widehat{w}}(\delta,\lambda))w_\lambda^{sb} = -\frac{1}{2(\Delta-\delta)}\frac{2(2c-\delta)}{(3-2\lambda)^2} < 0$. We then obtain

$$\begin{aligned}
\Omega_\delta(\delta,\lambda) &= (1-\lambda)\left(1-x^{sb}\right)w_\delta^{sb} - (\Delta-\delta)\left[x^{fb}(\delta)-x^{sb}\right] \\
&\quad\times\left[x_\delta^*(\delta,\widehat{\widehat{w}}(\delta,\lambda)) + x_w^*(\delta,\widehat{\widehat{w}}(\delta,\lambda))w_\delta^{sb}\right] + \frac{\left(1-x^{sb}\right)x^{sb}}{2} \\
&= \left\{(1-\lambda)\left(1-x^{sb}\right) - (\Delta-\delta)\left[x^{fb}(\delta)-x^{sb}\right]x_w^*(\delta,\widehat{\widehat{w}}(\delta,\lambda))\right\}w_\delta^{sb} \\
&\quad-(\Delta-\delta)\left[x^{fb}(\delta)-x^{sb}\right]x_\delta^*(\delta,\widehat{\widehat{w}}(\delta,\lambda)) + \frac{\left(1-x^{sb}\right)x^{sb}}{2}.
\end{aligned}$$

Consider the first term (in bracelets). Noting that $x_w^*(\delta,\widehat{\widehat{w}}(\delta,\lambda)) = \frac{-1}{2(\Delta-\delta)}$ we have

$$\begin{aligned}
&(1-\lambda)\left(1-x^{sb}\right) - (\Delta-\delta)\left[x^{fb}(\delta)-x^{sb}\right]x_w^*(\delta,\widehat{\widehat{w}}(\delta,\lambda)) \\
&= (1-\lambda)\left(1-x^{sb}\right) + \frac{x^{fb}(\delta)-x^{sb}}{2} \\
&= \frac{2(1-\lambda)+x^{fb}(\delta)-(3-2\lambda)x^{sb}}{2} \\
&= \frac{4(1-\lambda)(\Delta-\delta)+2(\Delta-c-\delta/2)-(3-2\lambda)\left(\Delta-c-w^{sb}\right)}{4(\Delta-\delta)} = 0,
\end{aligned}$$

where the last equality can be verified when inserting $w^{sb} = \widehat{\widehat{w}}(\delta,\lambda)$ from (14) and canceling terms appropriately. But then,

$$\Omega_\delta(\delta,\lambda) = -(\Delta-\delta)\left[x^{fb}(\delta)-x^{sb}\right]x_\delta^*(\delta,\widehat{\widehat{w}}(\delta,\lambda)) + \frac{\left(1-x^{sb}\right)x^{sb}}{2} \geq 0$$

since $x^{fb}(\delta) \leq x^{sb}$ for $\delta \in \left[\widehat{\delta}(\lambda), 2(\Delta-c)\right]$ and $x_\delta^*(\delta,\widehat{\widehat{w}}(\delta,\lambda)) > 0$. Finally, consider

$$\begin{aligned}
\Omega_\lambda(\delta,\lambda) &= -\Pi\left(x^{sb},w^{sb}\right) + (1-\lambda)\left(1-x^{sb}\right)w_\lambda^{sb} - (\Delta-\delta)\left[x^{fb}(\delta)-x^{sb}\right]x_\lambda^{sb} \\
&= -\Pi\left(x^{sb},w^{sb}\right) + \left\{(1-\lambda)\left(1-x^{sb}\right) - (\Delta-\delta)\left[x^{fb}(\delta)-x^{sb}\right]x_w^*(\delta,\widehat{\widehat{w}}(\delta,\lambda))\right\}w_\lambda^{sb} \\
&= -\Pi\left(x^{sb},w^{sb}\right) < 0,
\end{aligned}$$

where the third equality follows as we have just shown that the term in bracelets is zero. Consequently, $\Omega_\delta(\delta,\lambda) \geq 0$ and $\Omega_\lambda(\delta,\lambda) < 0$ for all $\delta \in [0, 2(\Delta-c)]$ and $\lambda \in [0,1]$, which completes the proof of part (i).

ECON WPS

**Parts (ii) and (iii):** Observing that $\frac{\Delta-c}{2} < \Delta - c \leq \widehat{\delta}\,(\lambda)$, we obtain from (15)

$$\begin{aligned}
\Omega(\frac{\Delta-c}{2},\lambda) &\equiv (1-\lambda)\,\Pi\left(x^*(\frac{\Delta-c}{2},0),0\right) - \frac{\Delta+c}{2}\left(x^{fb}\left(\frac{\Delta-c}{2}\right) - \frac{x^*(\frac{\Delta-c}{2},0)}{2}\right)x^*(\frac{\Delta-c}{2},0) \\
&= \left\{(1-\lambda)\,(p^*-c) - \frac{\Delta+c}{2}\left(x^{fb}\left(\frac{\Delta-c}{2}\right) - \frac{x^*(\frac{\Delta-c}{2},0)}{2}\right)\right\}x^*(\frac{\Delta-c}{2},0) \\
&= \left\{(1-\lambda)\,\frac{\Delta-c}{2} - \frac{\Delta+c}{2}\left(\frac{3\,(\Delta-c)}{2\,(\Delta+c)} - \frac{\Delta-c}{2\,(\Delta+c)}\right)\right\}x^*(\frac{\Delta-c}{2},0) \\
&= \left\{(1-\lambda)\,\frac{\Delta-c}{2} - \frac{\Delta-c}{2}\right\}x^*(\frac{\Delta-c}{2},0) \\
&= -\lambda\frac{\Delta-c}{2}x^*(\frac{\Delta-c}{2},0) \leq 0
\end{aligned}$$

with a strict equality if and only if $\lambda = 0$. Similarly, it is readily checked that

$$\Omega(2\,(\Delta-c)\,,\lambda) \equiv (1-\lambda)\,\Pi\left(x^{sb},w^{sb}\right) + \frac{2c-\Delta}{2}\left(x^{sb}\right)^2 \geq 0,$$

with a strict equality for $\lambda = 1$ since $x^{sb} = x^*(2\,(\Delta-c)\,,\widehat{\widehat{w}}\,(2\,(\Delta-c)\,,1) = x^*(2\,(\Delta-c)\,,\widehat{w}\,(2\,(\Delta-c)) = x^{fb}\,(2\,(\Delta-c)) = 0$. This completes the proof of parts (ii) and (iii).

ECON WPS

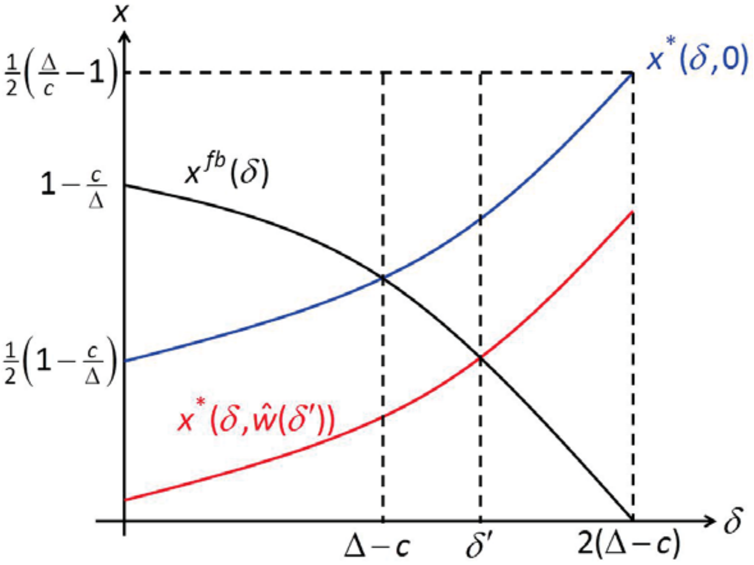

Figure 1: Laissez-faire versus first-best allocation.

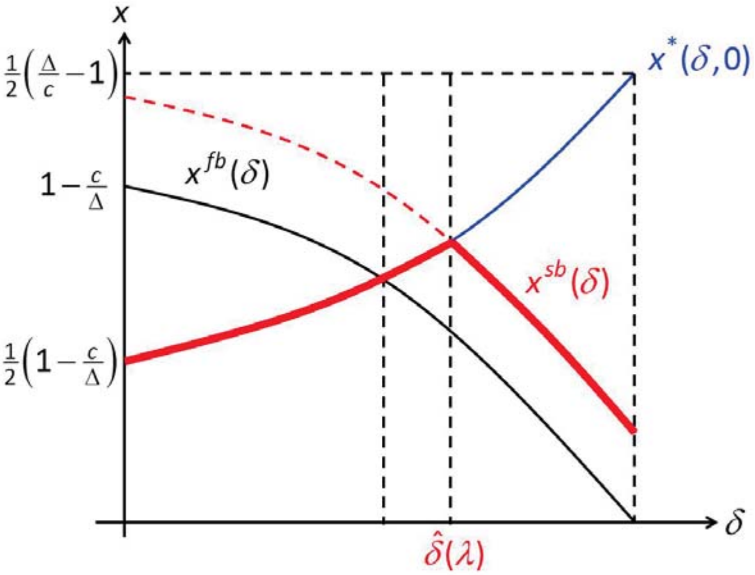

Figure 2: Second-best allocation.

ECON WPS

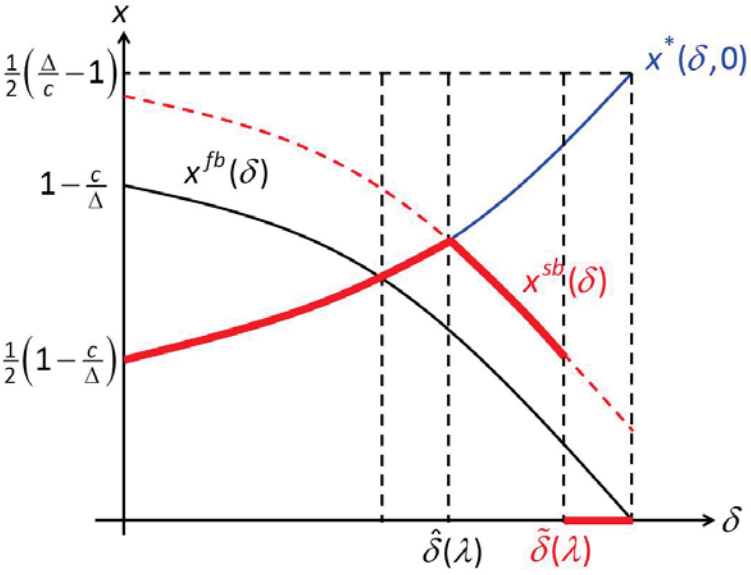

Figure 3: Ban of dual practice (high $\lambda$).

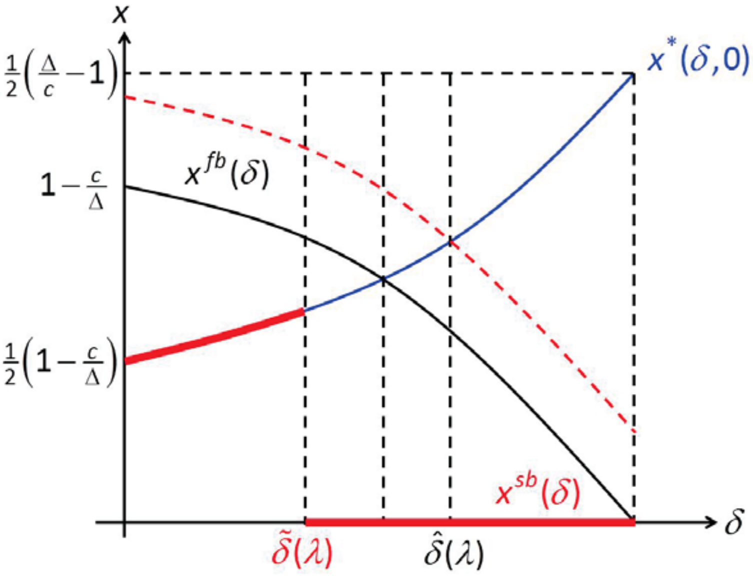

Figure 4: Ban of dual practice (low $\lambda$).

ECON WPS

ECON WPS

Working Papers published in the Series - State September 2013

## Published Working Papers

WP 02/2013: Saving the public from the private? Incentives and outcomes in dual practice

WP 01/2013: The Age-Productivity Pattern: Do Location and Sector Affiliation Matter?

WP 05/2012: The Public Reallocation of Resources across Age: A Comparison of Austria and Sweden

WP 04/2012: Quantifying the role of alternative pension reforms on the Austrian economy

WP 03/2012: Growth and welfare effects of health care in knowledge based economies

WP 02/2012: Public education and economic prosperity: semi-endogenous growth revisited

WP 01/2012: Optimal choice of health and retirement in a life-cycle model

WP 04/2011: R&D-based Growth in the Post-modern Era

WP 03/2011: Ageing, productivity and wages in Austria

WP 02/2011: Ageing, Productivity and Wages in Austria: evidence from a matched employer-employee data set at the sector level

WP 01/2011: A Matched Employer-Employee Panel Data Set for Austria: 2002 - 2005

Please cite working papers from the ECON WPS like this example:
Freund, I., B. Mahlberg and A. Prskawetz. 2011. "A Matched Employer-Employee Panel Data Set for Austria: 2002-2005." *ECON WPS 01/2011*. Institute of Mathematical Methods in Economics, Vienna University of Technology.

ECON WPS

Vienna University of Technology Working Papers
in Economic Theory and Policy

ISSN 2219-8849 (online)
http://www.econ.tuwien.ac.at/wps/

September 2013

**The Series "Vienna University of Technology Working Papers in Economic Theory and Policy" is published by the**

Research Group Economics
Institute of Mathematical Methods in Economics
Vienna University of Technology

**Contact**

Research Group Economics
Institute of Mathematical Methods in Economics
Vienna University of Technology

Argentinierstraße 8/4/105-3