THE UNIVERSITY *of York*

CHE
Centre For Health Economics

Price Adjustment in the Hospital Sector

CHE Research Paper 41

PRICE ADJUSTMENT IN THE HOSPITAL SECTOR

Marisa Miraldo[1]
Luigi Siciliani[1,2,3]
Andrew Street[1]

[1]Centre for Health Economics, University of York, UK
[2]Department of Economics and Related Studies, University of York, UK
[3]CEPR, London

August 2008

Background

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

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

Acknowledgements

The project was funded by the Department of Health in England as part of a programme of policy research at the Centre for Health Economics, University of York. The views expressed are those of the authors and may not reflect those of the funder.

Disclaimer

Papers published in the CHE Research Paper (RP) series are intended as a contribution to current research. Work and ideas reported in RPs may not always represent the final position and as such may sometimes need to be treated as work in progress. The material and views expressed in RPs are solely those of the authors and should not be interpreted as representing the collective views of CHE research staff or their research funders.

Further copies

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

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

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

Abstract

Prospective payment systems are currently used in many OECD countries, where hospitals are paid a fixed price for each patient treated. We develop a theoretical model to analyse the properties of the optimal fixed prices to be paid to hospitals when no lump-sum transfers are allowed and when the price can differ across providers to reflect observable exogenous differences in costs (for example land, building and staff costs). We find that: a) when the marginal benefit from treatment is decreasing and the cost function is the (commonly used) power function, the optimal price adjustment for hospitals with higher costs is positive but partial; if the marginal benefit from treatment is constant, then the price is identical across providers; b) if the cost function is exponential, then the price adjustment is positive even when the marginal benefit from treatment is constant; c) the optimal price is lower when lump-sum transfers are not allowed, compared to when they are allowed; d) higher inequality aversion of the purchaser is associated with an increase in the price for the high-cost providers and a reduction in the price of the low-cost providers.

Keywords: Price adjustment, Hospitals, DRGs.
JEL classification: I11, I18.

1 Introduction

Over the last three decades most OECD countries have made a transition from cost-based reimbursement of hospitals, via negotiated budgets, to prospective payment systems under which hospitals are paid on the basis of the volume and type of patients treated (Mossialos, 2002).

Under cost reimbursement hospital payments are paid retrospectively on the basis of the cost incurred by each individual patient. This reimbursement system operated in the United States during the 1960s and 1970s. This fuelled escalation in health care costs as hospitals engaged in a “medical arms race”, spending ever more on technologies and facilities to attract patients. Under cost reimbursement, providers were able to claim the costs back from health insurance companies and Medicare and Medicaid, the public insurance programmes for the elderly and the poor.

Prospective payment systems aim to overcome the deficiencies of cost-based reimbursement. Incentives for cost control and efficient behaviour are introduced by relating payment directly to activity and by ensuring that hospitals cannot influence the price they face. There are two key features of prospective payment systems. First, activity is described using some form of diagnosis related groups (DRGs)[1] rather than for each individual patient (under cost-based reimbursement). Second, the price per DRG is fixed in advance and independent of the costs incurred by the hospital.

Several theoretical studies have analysed the design of optimal payment systems to induce providers to behave optimally, i.e. to provide the optimal level of quality, cost containment effort and the optimal number of treatments (see for example Chalkley and Malcomson, 1998a and 1998b; Ma, 1994; Rickman and McGuire, 1999; Ellis and McGuire, 1986; Ellis, 1998). A common assumption in this literature is that the payer pays a price that is not provider specific, combined with a fixed lump-sum transfer in order to ensure providers' participation in the market.

However, in practice, lump-sum transfers are infrequent.[2] Instead, in countries like

[1] DRGs are standard groupings of clinical treatments, which use similar levels of healthcare resources.

[2] Moreover, if implemented, lump-sum transfers (or fixed-budget component) are positive, while the

England (under Payment by Results) and the US (within Medicare), each hospital receives a price for every patient treated. Moreover, prices are adjusted to incorporate providers' exogenous differences in costs. These adjustments are justified on the basis that they compensate for costs that hospitals incur because of the environment in which they are located or the constraints they have on the organisational structure. Such costs are considered out of the hospital's control.

These constraints bind more tightly in socialised health systems, where hospitals have highly restricted choice about where they are located, or the population they serve, and – at least in the short term – have limited discretion about their size and the mix of specialties they have. These constraints may impact on the cost of service provision, irrespective of how "efficient" the hospital is. For example, in England under Payment by Results (PbR), hospitals are paid an HRG price (the English version of DRG prices) based on national average costs adjusted by a provider specific index, the Market Factor Forces (MFF). The MFF adjusts the national price for local unavoidable differences in factor prices for staff, land and building costs. The staff index is built using data on private sector wages and is calculated to account for wage variation and indirect costs of employing staff. The buildings index is based on a rolling average of tender prices for all public and private contracts. The land index is calculated for each hospital in the National Health Service (NHS) and Primary Care Trust (PCTs), using data from the Valuation Office on the NHS estate in 2004. These sub-indices are then combined into a single overall index, known as the MFF index, which is built by multiplying each providers normalized sub-indices by the national proportionate usage of these inputs (Department of Health 2002a).

Also in the US, where providers are less restricted in their choices, the DRG payment system allows for adjustments to the average cost based on providers' characteristics, for example to adjust for wage variation, cost variations between urban or rural areas and teaching status (Shwartz, Merrill and Blake 1984).

To the best of our knowledge only the analysis by Mougeot and Naegelen (2005)

theory predicts negative lump-sum transfers whenever the marginal cost is increasing (as it is commonly assumed in the literature).

considers the properties of optimal hospital regulation when no lump-sum transfers are allowed. More precisely, the authors study whether a fixed-price regime complemented with global expenditure caps can induce first-best quality and cost-containment effort levels when lump-sum transfers are not allowed. While considering that no lump-sum transfers are allowed, our analysis departs from theirs by focusing on the properties of fixed prices when they can be adjusted to reflect exogenous cost differences between providers.

If lump-sum transfers are allowed, the optimal incentive scheme is straightforward. The purchaser can obtain allocative efficiency by setting the price equal to the marginal benefit of provision, so that in equilibrium activity is chosen such that the marginal benefit is equal to the marginal cost. The purchaser does not have to worry about leaving a rent to the provider as, through the use of lump-sum transfers, the purchaser can extract rents and leave providers with zero profits. If providers differ in costs and the marginal benefit is decreasing, then the price for the high-cost provider is higher than for the low-cost provider: since the optimal activity from the purchaser's perspective is such that marginal benefit is equal to the marginal cost, the optimal activity for the high-cost provider is lower than for the low-cost provider, while the marginal benefit (evaluated at the optimal activity) and therefore the optimal price is higher.

If lump-sum transfers are not allowed, the purchaser has only one instrument (prices) to obtain two goals, i.e. allocative efficiency and rent extraction. Since providing a higher price to the high-cost provider increases its rent, it is not straightforward anymore that providing a higher price to the high-cost provider is still the optimal solution. The main result of this study is that in an imperfect setting where purchasers cannot use lump-sum transfers, under reasonable assumptions, it is still the case that the optimal price for the high-cost provider is higher than the optimal price for the low-cost provider.

More precisely, we find that: a) when the marginal benefit from treatment is decreasing and the cost function is the (commonly used) power function, the optimal price adjustment for hospitals with higher costs is positive but it is partial, i.e. the price adjustment is smaller than the additional marginal cost; if the marginal benefit from treatment is constant, then the price is identical across providers; b) if the cost function is exponential, then the price adjustment is positive even when the marginal benefit from treatment is constant; c) we further show that the optimal price is lower when lump-sum transfers are not allowed, compared to when they are allowed; moreover, the price adjustment when lump-sum transfers are allowed is also higher than the price adjustment when lump-sum transfers are not allowed if the cost function is not too convex; otherwise the comparison is indeterminate; d) finally we show that higher inequality aversion of the purchaser is associated with an increase in the price for the high-cost providers and a reduction in the price of the low-cost providers.

The paper is organized as follows. Section 2 introduces the main assumptions of the model and derives the optimal pricing policy. Section 3 compares the optimal price adjustment when lump-sum transfers are allowed and when they are not. Section 4 extends the analysis by assuming that the purchaser is averse to inequality. Section 5 extends the basic model with quality. Section 6 presents the concluding remarks.

2 The Model

2.1 Provider

Define q as the number of patients treated by each provider. The provider receives from the purchaser a price p for each patient treated. We interpret p as the reimbursement per 1.0 DRG equal, for example, to roughly US$ 4000 under the Medicare Programme during the mid-1990s.

We assume that providers differ in costs. The cost function of provider θ is $C(\theta, q)$. We assume $C_q > 0$ and $C_{qq} > 0$: cost is increasing in quantity at an increasing rate.[3] We also assume that $C_\theta > 0$ and $C_{q\theta} > 0$: hospitals with higher θ have higher cost and higher *marginal* cost of treatment. We assume that θ is observable to the purchaser.

[3] As the activity volume approaches the provider's capacity it is plausible to assume that the marginal cost increases due to congestion costs arising on limited capacity. For example, treating patients in congested conditions is more demanding and stressful as more effort by the doctor is required to treat an extra patient.

In England θ can be interpreted as unavoidable cost differences between providers, for example accounting for the fact that land, building and staff costs are considerably higher in some areas rather than others (for example London area versus the Midlands area).

Hospitals are taken to be profit-maximisers (or surplus maximisers). The utility of the provider U is given by $U(\theta, q) = pq - C(\theta, q)$. This assumption may seem unrealistic for hospitals operating in a publicly-funded health care system, since public hospitals have constraints on the distribution of profits. However, they may add to their reserves the financial surplus obtained. Alternatively, managers may spend the surplus to pursue other objectives such as increasing physician staff, expanding the range of services, or even increasing managerial perks (see Dranove and White, 1994; De Fraja, 2000; Chalkley and Malcomson, 1998a and 1998b; Rickman and McGuire, 1999). For example Foundation Trusts in England, despite being public providers, are the residual claimants of their surpluses and, therefore, are considered to be profit maximizers (Department of Health, 2002b). Each provider θ maximises its utility U by optimally choosing quantity q such that the following First Order Condition (FOC) is satisfied:

$$p = C_q(\theta, q) \tag{1}$$

That is, the optimal quantity chosen is such that the price equals the marginal cost of treatment. The Second Order Condition (SOC) is $\partial^2 U(\theta, q)/\partial q^2 = -C_{qq} < 0$, which is always satisfied. Totally differentiating (1) we obtain:

$$q_p := \frac{\partial q}{\partial p} = \frac{1}{C_{qq}} > 0,\ q_\theta := \frac{\partial q}{\partial \theta} = -\frac{C_{q\theta}}{C_{qq}} < 0 \tag{2}$$

A higher price increases quantity, and a higher marginal cost reduces quantity. We also obtain:

$$q_{p\theta} \ : \ = \frac{\partial^2 q}{\partial \theta \partial p} = \frac{\partial^2 q}{\partial p \partial \theta} = \left[\frac{C_{q\theta}C_{qqq}}{C_{qq}^2} - \frac{C_{qq\theta}}{C_{qq}}\right] q_p = \frac{C_{q\theta}C_{qqq}}{C_{qq}^3} - \frac{C_{qq\theta}}{C_{qq}^2} \tag{3}$$

$$q_{pp} \ : \ = \frac{\partial^2 q}{\partial p^2} = -\frac{C_{qqq}}{C_{qq}^2} q_p = -\frac{C_{qqq}}{C_{qq}^3} \tag{4}$$

which will be useful below. If $q_{p\theta} < 0$ then providers with higher costs respond less to an increase in price than providers with lower costs. q_{pp} indicates whether the responsiveness of quantity with respect to price increases or decreases for higher levels of price. In general, the sign of these expressions depends on the specific functional form of total cost. Suppose that the cost function is the power function $C(\theta, q) = \theta q^{\gamma}/\gamma$ with $\gamma > 1$ (note that for $\gamma = 2$ we obtain the commonly used quadratic cost function $\theta q^2/2$), then $q_{p\theta} = -\frac{1}{\theta^2(\gamma-1)^2 q^{\gamma-2}} < 0$ and providers with higher costs respond less to a marginal increase in price, and $q_{pp} = -\frac{\gamma-2}{\theta^2(\gamma-1)^2 q^{2\gamma-3}}$, which is negative for $\gamma > 2$: the responsiveness of activity to price decreases with price.

2.2 Purchaser

Define W^s as the utility of the purchaser of health services.[4] The purchaser buys medical care from the provider at a price p. We assume that the purchaser's utility is a weighted sum of patients' utility (or consumers' surplus) and providers' utility net of the transfer to the provider (weighted by the opportunity cost of public funds). More precisely, define $B(q)$ as patients' benefit, with $B_q > 0$, $B_{qq} \leq 0$: benefit is increasing in quantity at a (weakly) decreasing rate; $\delta \in [0, 1]$ as the weight attached to the provider; and $\lambda > 0$ as the shadow cost of public funds, i.e. for each \$1 levied to subsidize health care expenditure, distortionary taxation generates \$$(1+\lambda)$ disutility for the taxpayers. The purchaser's utility is then $W^s = B(q) + \delta U(\theta, q) - (1+\lambda) pq$, which after substitution of U gives

$$W^s = B(q) - (1 + \lambda - \delta) pq - \delta C(\theta, q). \tag{5}$$

This specification clusters three polar cases of welfare functions. 1) The first accrues to the scenario where public funds are not costly ($\lambda = 0$) and the purchaser attaches zero weight to the provider's utility ($\delta = 0$). In this case the welfare function coincides with net consumer surplus: $W = B(q) - pq$. 2) For a purchaser attaching equal weight to consumer and provider surplus ($\delta = 1$), in the absence of distortionary effects from raising public

[4] Superscript s will be used throughout the paper to denote the scenario with no lump-sum transfers.

funds ($\lambda = 0$), we obtain a utilitarian welfare function: $W = B(q) - C(\theta, q)$. 3) Finally, if raising public funds is costly ($\lambda > 0$) and $\delta = 1$, then $W = B(q) - C(\theta, q) - \lambda pq$.

For each provider θ the regulator sets the price $p(\theta)$. The optimal price p^s to be paid to provider θ is then characterized by the following FOC ($\partial W/\partial p(p^s) = 0$):

$$B_q q_p = [(1-\delta) + \lambda](q + q_p p) + \delta C_q q_p \tag{6}$$

Dividing by q_p and rearranging, we obtain:

$$B_q = [(1-\delta) + \lambda]\, p \left(1 + \frac{1}{\varepsilon_p^q}\right) + \delta C_q \tag{7}$$

where $\varepsilon_p^q = q_p(p/q) > 0$ is the elasticity of quantity with respect to price. At the optimum, the marginal benefit from treatment for the patients is equal to the marginal cost. The marginal cost includes two components: i) the cost associated with the purchaser payment to the hospital, weighted by the shadow cost of public funds; ii) the treatment costs multiplied by the weight attached to the provider's utility. Notice how a higher provider's elasticity of quantity to price implies a lower marginal cost. For the special cases defined above, we obtain that if the purchaser disregards provider's utility and health expenditure is funded out of non-distortionary taxation ($\lambda = 0, \delta = 0$) the optimal price is such that the marginal benefit of treatment equals the marginal cost of the transfer from the purchaser to the provider, $B_q = p\left(1 + \frac{1}{\varepsilon_p^q}\right)$. With an utilitarian purchaser ($\delta = 1$) results are twofold: if raising public funds is not costly ($\lambda = 0$) the optimal price is set such that the marginal benefit is equal to the marginal cost of treatment, $B_q = C_q$. Otherwise, if public funds are costly the optimal price is such that the marginal benefit from treatment equals the marginal cost of the transfer from the purchaser to the provider plus the marginal cost from treatment, i.e. $B_q = \lambda p\left(1 + \frac{1}{\varepsilon_p^q}\right) + C_q$. Note that the optimal price is higher under the former, as, for the same benefits, social costs are lower.

Rearranging the FOC of the purchaser's problem, and setting $C_q = p$, we obtain

$$p^s = \frac{B_q}{1+\lambda} - \frac{(1+\lambda-\delta)}{1+\lambda}\frac{q}{q_p} \tag{8}$$

which establishes that the optimal price is set below the marginal benefit. The first term is the marginal benefit from treatment discounted for the opportunity cost of public funds. Since leaving a rent to the provider is costly, the second term is negative and implies a lower price. Note that a higher responsiveness of activity (q_p) to price implies a higher optimal price.

The Second Order Condition is $\partial^2 W/\partial p^2 < 0$ and is always satisfied (see Appendix A). By the implicit function theorem we obtain $dp^s/d\theta = -\frac{\partial^2 W}{\partial p \partial \theta}/\frac{\partial^2 W}{\partial p^2}$. Therefore the sign of $dp^s/d\theta$ depends on the sign of $\frac{\partial^2 W}{\partial p \partial \theta}$. Totally differentiating with respect to θ we obtain:

$$W_{p\theta} := \frac{\partial^2 W}{\partial p \partial \theta} = B_{qq} q_p q_\theta - (1+\lambda-\delta)\, q_\theta + \frac{(1+\lambda-\delta)\, q}{q_p} q_{p\theta} \tag{9}$$

There are three main terms. The first term is positive. Since high-cost providers provide lower activity, the marginal benefit from an increase in activity is higher and therefore the price should be higher (*treatment effect)*. The second term is also positive. Again, since high-cost providers provide lower activity, the rent for high-cost providers, and the associated cost, is lower. Therefore the price should be higher (*rent effect*). The third term is negative whenever high-cost providers respond less to an increase in price, i.e. when $q_{p\theta} < 0$, which is the case for many well-behaved cost functions. Since an increase in price is less effective in boosting activity for high-cost providers, then the optimal price should be lower (*responsiveness effect*). Therefore, the price for high-cost providers is higher than for low-cost providers only if the first two effects dominate the third.

By substitution, we obtain (see Appendix B):

$$\frac{dp^s}{d\theta} = -\frac{\partial^2 W}{\partial p \partial \theta} / \frac{\partial^2 W}{\partial p^2} = \frac{-B_{qq} C_{q\theta} + (1+\lambda-\delta)\left\{\left(C_{q\theta} - qC_{qq\theta}\right) C_{qq} + qC_{q\theta} C_{qqq}\right\}}{-B_{qq} + (1+\lambda-\delta)\, qC_{qqq} + (2-\delta+2\lambda)\, C_{qq}} \tag{10}$$

The denominator of $dp^s/d\theta$ is positive when the SOC of the purchaser's problem is satisfied (i.e. $W_{pp} < 0$). The sign of $dp^s/d\theta$ is determined by the sign of the numerator which is in general indeterminate and will differ according to the functional form of the cost function. Proposition 1 provides a condition that guarantees that the price is higher for providers with higher costs.

Proposition 1 *Suppose that: (a) costs behave according to the power function: $C(\theta, q) = \theta q^\gamma/\gamma$, with $\gamma > 1$; and (b) the marginal benefit from treatment is decreasing ($B_{qq} < 0$); then the optimal price is such that:*

$$0 < \frac{dp^s}{d\theta} = -\frac{\partial^2 W}{\partial p \partial \theta} / \frac{\partial^2 W}{\partial p^2} = \frac{-B_{qq} C_{q\theta}}{-B_{qq} + \theta\left(\gamma - 1\right) q^{\gamma-2}\left(\gamma\left(1-\delta\right) + \delta + \lambda\gamma\right)} < C_{q\theta}. \tag{11}$$

Hospitals with higher costs receive a higher price. Moreover, the price increase (price adjustment) for providers with higher costs is smaller than the additional marginal cost.

Proof. >From $C(\theta, q) = \theta q^\gamma/\gamma$ with $\gamma > 1$, we obtain $C_q = \theta q^{\gamma-1} > 0$, $C_{qq} = \theta\left(\gamma-1\right) q^{\gamma-2} > 0$, $C_{q\theta} = q^{\gamma-1} > 0$, $C_{qqq} = \theta\left(\gamma-1\right)\left(\gamma-2\right) q^{\gamma-3}$, $C_{qq\theta} = \left(\gamma-1\right) q^{\gamma-2}$. By substitution we obtain $C_{q\theta} - qC_{qq\theta} + q\frac{C_{q\theta}C_{qqq}}{C_{qq}} = q^{\gamma-1} - q\left(\gamma-1\right) q^{\gamma-2} + q\frac{q^{\gamma-1}\theta(\gamma-1)(\gamma-2)q^{\gamma-3}}{\theta(\gamma-1)q^{\gamma-2}} = 0$. Substituting in (10) for $C_{qq} = \theta\left(\gamma-1\right) q^{\gamma-2}$, $C_{qqq} = \theta\left(\gamma-1\right)\left(\gamma-2\right) q^{\gamma-3}$ we obtain $\left[\left(1-\delta\right) + \lambda\right] qC_{qqq} + \left(2 - \delta + 2\lambda\right) C_{qq} = \theta\left(\gamma-1\right) q^{\gamma-2}\left[\gamma\left(1-\delta\right) + \delta + \lambda\gamma\right] > 0$. ∎

Therefore, for a broad class of cost functions with decreasing returns to scale, like the power function, we have that the providers with higher costs θ are paid higher prices when the marginal benefit from treatment is decreasing ($B_{qq} < 0$). Intuitively, if the marginal benefit is decreasing with quantity, then the regulator is willing to pay a higher price to the high-cost provider in order to induce the treatment of patients that, compared to patients receiving care in the low-cost provider, benefit more from treatment, i.e. the *treatment effect* is stronger for high-cost providers. Notice that for this specification the expression in the curly brackets of Eq.(10) is zero and the *rent effect* is, perhaps surprisingly, exactly offset by the *responsiveness effect*.

As a corollary, note that if the marginal benefit from treatment is instead constant ($B_{qq} = 0$), it is optimal for the purchaser to set the same price for each provider regardless of their costs, so that $dp^s/d\theta = 0$ when the cost function is the power function. Indeed, a lower activity by a high-cost provider no longer has an impact on the marginal benefit from treatment and therefore there is no incentive to pay providers differently. This might be the case for interventions, such as diagnostic procedures, where the benefit of intervention is not known in advance.

A related issue is whether the price adjustment designed by the purchaser should be proportional to the increase in the marginal cost faced by high-cost providers. Proposition 1 establishes that such adjustment is smaller than the marginal cost. Even though, intuitively one would expect that the price adjustment paid to high-cost providers should cover the cost difference between providers, given that the marginal benefit is decreasing, the price adjustment is only partial. As a limit case, note that if $\gamma \to 1$ (constant marginal cost), then the price adjustment is instead proportional to the increase in the marginal cost faced by high-cost providers, as $lim_{\gamma\to 1} dp^s/d\theta = C_{q\theta}$.

Proposition 2 shows the price adjustment for a different cost function.

Proposition 2 *Suppose that the cost function of the provider is exponential:* $C(\theta, q) = \theta e^{\gamma q}$*, then the optimal price is such that:*

$$\frac{dp^s}{d\theta} = -\frac{\partial^2 W}{\partial p \partial \theta} / \frac{\partial^2 W}{\partial p^2} = \frac{-B_{qq}C_{q\theta} + (1+\lambda-\delta)\,\theta\gamma^3 e^{2\gamma q}}{-B_{qq} + (1+\lambda-\delta)\,qC_{qqq} + (2-\delta+2\lambda)\,C_{qq}} > 0 \qquad (12)$$

Proof. Suppose that $C(\theta, q) = \theta e^{\gamma q}$ with $\gamma > 0$, then, $C_q = \theta\gamma e^{\gamma q}$, $C_{qq} = \theta\gamma^2 e^{\gamma q}$, $C_{q\theta} = \gamma e^{\gamma q} > 0$, $C_{qqq} = \theta\gamma^3 e^{\gamma q} > 0$, $C_{qq\theta} = \gamma^2 e^{\gamma q}$. By substitution we obtain that $\left(C_{q\theta} - qC_{qq\theta} + q\frac{C_{q\theta}C_{qqq}}{C_{qq}}\right) C_{qq} = \theta\gamma^3 e^{2\gamma q} > 0$. Since $B_{qq} < 0$ and $C_{q\theta} > 0$ the numerator is positive. The denominator is also positive implying $dp/d\theta > 0$. ■

When the cost function is exponential, it is still the case that hospitals with a higher cost receive a higher price. Notice that this result holds even if the marginal benefit from treatment is constant ($B_{qq} = 0$). This result is in contrast with the previous example, when the cost follows the power function, in which case we have shown that $dp^s/d\theta = 0$ for $B_{qq} = 0$.

Finally we investigate how activity varies across providers in equilibrium. Since $q^s =$

$q(\theta, p^s(\theta))$, then hospitals with higher costs provide less activity if: $\frac{dq^s}{d\theta} = \frac{\partial q}{\partial \theta} + \frac{\partial q}{\partial p^s}\frac{\partial p^s}{\partial \theta} < 0$, where $\partial q/\partial \theta < 0$ while $\partial q/\partial p^s > 0$ and $\partial p^s/\partial \theta > 0$. More extensively, after substitution and simplifying, we obtain

$$\frac{dq^s}{d\theta} = \frac{-(1+\lambda) C_{q\theta} - (1+\lambda-\delta) q C_{qq\theta}}{-B_{qq} + (1+\lambda-\delta) q C_{qqq} + (2-\delta+2\lambda) C_{qq}} \tag{13}$$

The denominator is positive when the SOC of the purchaser's problem is satisfied. The first term in the numerator is negative. The second term is also negative whenever $C_{qq\theta} > 0$, i.e. whenever the cost function is more concave for providers with higher costs. Suppose that the cost function is multiplicative separable $C(\theta, q) = \theta \widetilde{C}(q)$, with $\widetilde{C}_q > 0$, $\widetilde{C}_{qq} > 0$. Then $C_{qq\theta} = \widetilde{C}_{qq} > 0$ and hospitals with higher costs provide less activity in equilibrium. This condition is clearly satisfied for the two cost functions of propositions 1 and 2, when the cost function is respectively the power function or the exponential function (as they are both multiplicatively separable in θ). Providers with higher costs provide a lower quantity.

Finally, high-cost providers may have a higher or lower profit in equilibrium, since the sign of $\partial U/\partial \theta = q\partial p/\partial \theta - C_\theta$ is ambiguous.[5] Indeed, $\partial U/\partial \theta$ depends on the sign and magnitude of $\partial p/\partial \theta$, and of the cost and benefit functions. For the special case when the price does not vary across providers, providers with higher costs have lower profit. If instead $\partial p/\partial \theta > 0$, and for the case of the power cost function, high-cost providers have higher profits only when the benefit function is sufficiently concave (proof omitted).

In summary, when lump-sum transfers cannot be used as regulatory instruments, providers are allowed economic rents. Since these rents are costly then the regulator faces a trade-off between efficiency and rent extraction and therefore a fixed price policy will result in a second-best allocation. The next section compares the solution with a scenario where lump-sum transfers are allowed.

[5] Note that

$$\frac{dU}{d\theta} = \frac{\partial p}{\partial \theta} q + \left(p - \frac{\partial C}{\partial q}\right)\left(\frac{\partial q}{\partial \theta} + \frac{\partial q}{\partial p}\frac{\partial p}{\partial \theta}\right) - \frac{\partial C}{\partial \theta}$$

By the envelope theorem $p - \frac{\partial C}{\partial q} = 0$ so that $\frac{dU}{d\theta} = \frac{\partial p}{\partial \theta} q - \frac{\partial C}{\partial \theta}$.

3 Comparison with First Best

We define as First Best a scenario where the purchaser can use prices as well as lump-sum transfers to remunerate the provider. We also assume that, like prices, lump-sum transfers can differ across providers. Define $T(\theta)$ as the lump-sum transfer received by provider θ in conjunction with the price $p(\theta)$ for each patient treated. The utility of the provider is now given by $U^f(\theta, q) = pq + T - C(\theta, q)$. The optimal quantity chosen by the provider is the one that satisfies: $p = C_q(\theta, q)$, which for a given price coincides with the solution found in section 2.1.

As in the previous section the purchaser problem is to maximise $W^f = B(q) + \delta U^f(\theta, q) - (1+\lambda)(pq + T)$ subject to $U^f = pq + T - C(\theta, q) \geq 0$.[6] From the definition of U^f we can write $pq = U^f + C(\theta, q)$, which substituted into W^f, gives $W^f = B(q) - (1+\lambda)C(\theta, q) - (1+\lambda-\delta)U^f$. Since public funds are costly ($\lambda > 0$), and the weight attached to the utility of the provider is weakly less than one, $\delta \leq 1$, leaving a rent to the provider is costly from the purchaser's perspective (as $\partial W^f/\partial U^f < 0$). The optimal transfer will be then set at the minimum needed to ensure that the participation constraint of the provider is satisfied, so that $U^f = 0$ (and $T = C(\theta, q) - pq$).

The purchaser objective function is given by:

$$\max_p W^f = B(q(p)) - (1+\lambda)C(\theta, q(p)) \tag{14}$$

Maximizing W^f with respect to p the optimal price, denoted as p^f, is then characterized by the following FOC:

$$B_q q_p = (1+\lambda)C_q q_p \tag{15}$$

The price is set such that the marginal benefit from treatment is equal to its marginal cost ($B_q = (1+\lambda)C_q$).[7] Using the FOC for the optimal quantity of the provider, $p = C_q$, we also establish from the FOC of the purchaser that: $p^f = B_q/(1+\lambda)$, i.e. the price is

[6] Superscript f will be used to denote the first-best scenario.

[7] The SOC is given by: $W^f_{pp} = B_{qq}q_p^2 - (1+\lambda)C_{qq}q_p^2 + q_{pp}(B_q - (1+\lambda)C_q) < 0$.

equal to the marginal benefit discounted for the opportunity cost of public funds.

Proposition 3 compares the optimal price when lump-sum transfers are allowed and when they are not.

Proposition 3 *When lump-sum transfers are allowed the optimal price paid to a provider of type θ is larger than the optimal price when no lump-sum transfers are allowed, i.e.* $p^f > p^s$.

Proof. When no lump-sum transfers are allowed the optimal price is characterized by the first order condition: $B_q q_p = [(1-\delta)+\lambda](q+q_p p)+\delta C_q q_p$. Given that $p = C_q$, by substitution and rearranging, we obtain $B_q q_p = (1+\lambda) C_q q_p + (1+\lambda-\delta) q$. For $(1+\lambda-\delta) > 0$ the price solving this condition is necessarily lower than the one solving $B_q q_p = (1+\lambda) C_q q_p$. ■

Intuitively, when lump-sum transfers are not allowed, the provider obtains a positive rent. In contrast when they are allowed the rent is zero. To reduce such rents it is optimal for the purchaser to set a lower price. Since quantity is monotonically increasing in price, we can also establish that the quantity of care provided will be lower when lump-sum transfers are not allowed, i.e. $q^f \equiv q(p^f) > q(p^s) \equiv q^s$.

Moreover, we obtain:[8]

$$\frac{dp^f}{d\theta} = -\frac{\partial^2 W}{\partial p \partial \theta} / \frac{\partial^2 W}{\partial p^2} = C_{q\theta} \frac{-B_{qq}}{-B_{qq}+(1+\lambda) C_{qq}} > 0 \tag{16}$$

If the marginal benefit is decreasing, providers with higher θ are paid higher tariffs. However, the price increase (or tariff adjustment) for providers with higher costs is smaller than the additional marginal cost, i.e. $dp^f/d\theta < C_{q\theta}$. Again, if the marginal benefit is

[8] Totally differentiating with respect to θ, and recalling $B_q - (1+\lambda) C_q = 0$, we obtain:

$$\begin{aligned} W^f_{p\theta} \quad &: \quad = \frac{\partial^2 W}{\partial p \partial \theta} = [B_{qq} - (1+\lambda) C_{qq}] q_p q_\theta - (1+\lambda) C_{q\theta} q_p \\ &= \quad -[B_{qq} - (1+\lambda) C_{qq}] \frac{C_{q\theta}}{C^2_{qq}} - (1+\lambda) \frac{C_{q\theta}}{C_{qq}} = -B_{qq} \frac{C_{q\theta}}{C^2_{qq}} < 0 \\ W^f_{pp} \quad &: \quad = \frac{\partial^2 W^f}{\partial p^2} = B_{qq} q_p^2 - (1+\lambda) C_{qq} q_p^2 < 0. \end{aligned}$$

constant then the optimal price will be equal across all providers, i.e. $dp^f/d\theta = 0$. This result is in contrast with proposition 2, where we showed that providers that face a higher cost receive a higher price even when the marginal benefit is constant.

We have already established that the optimal price chosen by the purchaser is lower when lump-sum transfers are not allowed. We want now to establish whether the adjustment is also smaller, ie whether $dp^f/d\theta > dp^s/d\theta$.

Suppose the benefit and cost function are quadratic: $C(\theta, q) = \theta q^2/2$; $B(q) = aq - (b/2)q^2$. Then,

$$\frac{dp^f}{d\theta} = \frac{bq^f}{b + (1+\lambda)\,\theta} > 0, \; \frac{dp^s}{d\theta} = \frac{bq^s}{b + (2 - \delta + 2\lambda)\,\theta} > 0 \tag{17}$$

Note that the denominator of $dp^s/d\theta$ is larger while the numerator is smaller compared to $dp^f/d\theta$ and therefore we conclude that the price adjustment is also smaller when lump-sum transfers are not allowed, i.e $dp^f/d\theta > dp^s/d\theta$.

Suppose now that the cost function is the more general $C(\theta, q) = \theta q^\gamma/\gamma$ with $\gamma > 1$ (we still assume a quadratic benefit function). Then,

$$\frac{dp^f}{d\theta} = \frac{b\left(q^f\right)^{\gamma-1}}{b + \theta\,(\gamma-1)\,(1+\lambda)\,(q^f)^{\gamma-2}} > 0, \tag{18}$$

$$\frac{dp^s}{d\theta} = \frac{b\,(q^s)^{\gamma-1}}{b + \theta\,(\gamma-1)\,[\gamma\,(1+\lambda) + \delta\,(1-\gamma)]\,(q^s)^{\gamma-2}} > 0 \tag{19}$$

Given that $q^f > q^s$ the numerator of $dp^f/d\theta$ is larger than the numerator of $dp^s/d\theta$. Given that $\gamma\,(1+\lambda) + \delta\,(1-\gamma) > (1+\lambda)$, for $\gamma > 1$, then also the denominator of $dp^f/d\theta$ is smaller when $1 < \gamma \leq 2$, and $dp^f/d\theta > dp^s/d\theta$. Otherwise the comparison is ambiguous when $\gamma > 2$. In summary, the price adjustment is smaller when no lump-sum transfers are allowed when the cost function is not too convex, i.e $1 < \gamma \leq 2$. The comparison is ambiguous when $\gamma > 2$.

Similarly, suppose that the cost function is exponential: $C(\theta, q) = \theta e^{\gamma q}$ with $\gamma > 0$.

Then,

$$\frac{dp^f}{d\theta} = \frac{b\gamma e^{\gamma q^f}}{b+(1+\lambda)\theta\gamma^2 e^{\gamma q^f}} > 0, \tag{20}$$

$$\frac{dp^s}{d\theta} = \frac{b\gamma e^{\gamma q^s} + [(1-\delta)+\lambda]\,\theta\gamma^3 e^{2\gamma q^s}}{b+(1+\lambda)\,\theta\gamma^2 e^{\gamma q^s} + \theta e^{\gamma q^s}\gamma^2\,(1-\delta+\lambda)\,(1+\gamma q^s)} > 0. \tag{21}$$

Comparing the numerators of both expressions given that $q^f > q^s$, the comparison is indeterminate. The comparison of the denominators is also indeterminate. Therefore the relation between $dp^f/d\theta$ and $dp^s/d\theta$ is ambiguous. Note that if the marginal benefit is constant (i.e. $B_{qq} = 0$) then $dp^s/d\theta > dp^f/d\theta = 0$: the price adjustment is smaller when lump-sum transfers are not allowed.

The analysis above establishes that when lump-sum transfers are allowed, the price is higher for hospitals with higher costs. How does the optimal lump-sum transfer differ across the providers? Recall that $T(\theta) = C\left(\theta, q(\theta, p^f(\theta))\right) - p^f(\theta)\, q(\theta, p^f(\theta)) < 0$. Differentiating T with respect to the cost parameter θ, using $p = C_q$, we find:

$$\frac{dT}{d\theta} = C_\theta - q^f \frac{dp^f}{d\theta} = C_\theta + \frac{q^f C_{q\theta} B_{qq}}{-B_{qq} + (1+\lambda)C_{qq}} \tag{22}$$

Given that $dp^f/d\theta > 0$ the sign of $dT/d\theta$ is ambiguous and will depend on the cost and benefit functions. If the marginal benefit is constant then $dT/d\theta > 0$. If the marginal benefit is decreasing, the price adjustment increases with the degree concavity of the benefit function. Since higher concavity implies that a marginal increase in activity from a high-cost provider leads to a large increase in benefit, it is more worthwhile to increase prices. However, a higher price implies more rent which will be extracted through a higher lump-sum transfer (in absolute values). Note that these results mirror the findings of the previous discussion on $\partial U/\partial\theta$.

Suppose that the cost function takes the form of the power function: $C(\theta, q) = \theta q^\gamma/\gamma$. Then,

$$\frac{dT}{d\theta} = q^\gamma\,(\gamma-1)\,\frac{B_{qq} + (1+\lambda)\theta q^{\gamma-2}}{-B_{qq}\gamma + (1+\lambda)\theta\gamma\,(\gamma-1)\,q^{\gamma-2}} \gtrless 0 \tag{23}$$

The denominator is positive by the SOC of the purchaser problem. Therefore, providers with higher costs have a lower lump-sum transfer only when the benefit function is sufficiently concave ($dT/d\theta < 0$), otherwise the effect is positive ($dT/d\theta > 0$). Note that since the lump-sum transfer is negative ($T(\theta) < 0$), if $dT/d\theta < 0$ then providers with higher costs have to pay to the purchaser a larger transfer than the providers with lower costs.

Finally, we investigate how activity varies across providers in equilibrium. Since $q^f = q(\theta, p^f(\theta))$, then hospitals with lower costs provide more activity if: $\frac{dq^f}{d\theta} = \frac{\partial q}{\partial \theta} + \frac{\partial q}{\partial p^f}\frac{\partial p^f}{\partial \theta} < 0$. More extensively, after substitution and simplifying, we obtain

$$\frac{dq^f}{d\theta} = -C_{q\theta}\frac{1+\lambda}{-B_{qq} + (1+\lambda) C_{qq}} < 0 \tag{24}$$

and hospitals with higher costs provide a lower quantity.

4 Extension with inequality aversion

Consider now that the purchaser maximizes a welfare function which is parameterized by a degree of social aversion to inequality. This could arise because of a desire to equalise geographical access to hospital services.[9] The benefit of the purchaser when patients receive care from provider θ is given by:

$$W(q) = \begin{cases} \frac{B(q)^{1-\rho}}{1-\rho} & \rho \neq 1 \\ \ln(B(q)) & \rho = 1 \end{cases} \tag{25}$$

where ρ is an index of social aversion to inequality in the provision of quantity. Higher ρ implies more aversion to inequality. In the limit, an infinite degree of aversion to inequality corresponds to the Rawlsian egalitarian social preferences (i.e. the maximization of the least-favoured). For $\rho = 0$ we obtain the utilitarian welfare function.

To keep the analysis simple, we assume that there are two types of providers such that

[9] Note that, implicitly, we are assuming that there is no mobility of patients across providers and that, being local monopolists, health care is always commissioned to all providers irrespectively of their efficiency level.

$\theta_i \in \{\theta_1, \theta_2\}$. The proportion of type 1 is ω and the proportion of type 2 is $(1-\omega)$, with $\omega \in (0,1)$. Without loss of generality we assume that $\theta_2 > \theta_1$: provider 2 faces higher exogenous costs. The marginal welfare gain from an increase in the benefit for patients in hospital i is given by $W_B = B^{-\rho}$ and $W_{B\rho} = -\ln(B)/B^{\rho}$.

We assume that the purchaser has a global budget K, which can be used to finance the two types of provider. As in section 2, we assume that lump-sum transfers are not allowed. The purchaser's maximisation problem is:

$$\max_{\{p_1,p_2\}} \omega \frac{B(q_1(\theta_1,p_1))^{1-\rho}}{1-\rho} + (1-\omega)\frac{B(q_2(\theta_2,p_2))^{1-\rho}}{1-\rho} \tag{26}$$

$$s.t. \;\; \omega p_1 q_1(\theta_1,p_1) + (1-\omega) p_2 q_2(\theta_2,p_2) \leq K \tag{27}$$

The Lagrangian of the problem is:

$$\mathcal{L} = \omega \frac{B(q_1)^{1-\rho}}{1-\rho} + (1-\omega)\frac{B(q_2)^{1-\rho}}{1-\rho} - \mu\left[\omega p_1 q_1 + (1-\omega) p_2 q_2 - K\right] \tag{28}$$

where $\mu \geq 0$ is the Lagrange multiplier. The optimal prices paid to the providers satisfy the following FOCs:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial p_1} &= \omega B(q_1)^{-\rho} B_{q_1} q_{p_1} - \omega\mu(q_1 + p_1 q_{p_1}) = 0 \\
\frac{\partial \mathcal{L}}{\partial p_2} &= (1-\omega) B(q_2)^{-\rho} B_{q_2} q_{p_2} - (1-\omega)\mu(q_2 + p_2 q_{p_2}) = 0 \\
\frac{\partial \mathcal{L}}{\partial \mu} &= -\omega p_1 q_1 - (1-\omega) p_2 q_2 + K = 0
\end{aligned} \tag{29}$$

In analogy to the results obtained in section 2, the price for the high-cost provider is higher than the price for the low-cost provider, i.e. $p_2 > p_1$.[10] Rearranging $\partial\mathcal{L}/\partial p_i = 0$, and dividing by q_{p_i}, we obtain:

$$B(q_i)^{-\rho} B_{q_i} = \mu\left(p_i + \frac{q_i}{q_{p_i}}\right) \tag{30}$$

[10] The proof for this result is omitted for brevity but available from the authors.

The marginal benefit from treatment is equal to the marginal cost in terms of the opportunity cost associated with the budget constraint. The second term takes into account the effect of price on the rents.

The optimal price is such that:

$$p_i = \frac{B\,(q_i)^{-\rho}\,B_{q_i}}{\mu} - \frac{q_i}{q_{p_i}} \tag{31}$$

The first term is the marginal benefit from treatment discounted for the opportunity cost of the budget constraint. Comparative statics with respect to the degree of inequality aversion suggests that:

$$\frac{dp_1}{d\rho} = \frac{-\,(1-\omega)\,(q_1+p_1 q_{p_1})\,(q_2+p_2 q_{p_2})^2\,\mu\,(\ln(B_1)-\ln(B_2))}{J} < 0 \tag{32}$$

$$\frac{dp_2}{d\rho} = \frac{-\omega\,(q_2+p_2 q_{p_2})\,(p_1 q_{p_1}+q_1)^2\,\mu\,(\ln(B_2)-\ln(B_1))}{J} > 0 \tag{33}$$

where J is a positive expression (see Appendix C). Given that the logarithm function is monotonic and increasing in its argument it follows that $\ln(B_1) > \ln(B_2)$ and therefore, given that $\omega \in (0,1)$ and $\theta_2 > \theta_1$, then $dp_1/d\rho < 0$ while $dp_2/d\rho > 0$. When inequality aversion is higher, the price for the high-cost provider is higher, while the price of the low-cost provider is lower. Since $p_2 > p_1$, this also implies that $d(p_2 - p_1)/d\rho > 0$: the price differential between the high-cost provider and the low-cost provider increases with the degree of inequality aversion.

The difference between the benefits of the patients treated within the two types of hospital become more relevant for the purchaser when the inequality aversion increases. Since provider 2 has a lower health benefit, the price paid to provider 2 is increased and the one paid to provider 1 is decreased so that differences in health outcomes between the two providers are reduced.

5 Extension with quality

In this section we extend the model by introducing quality, and we show that the results using this more general specification are similar to the ones obtained above. We follow the approach suggested by Ma (1994) and Chalkley and Malcomson (1998b). Define m as the quality generated by the provider. The cost function of the provider is $C(\theta, q, m)+\varphi(q, m)$. C includes the monetary cost, which increases with quality and activity: $C(\theta, q, m)$, with $C_q > 0$ and $C_m > 0$. φ is the non-monetary cost, or disutility, which increases with activity and quality: $\varphi(q, m)$, with $\varphi_q > 0$ and $\varphi_m > 0$.

We also assume that the demand for treatment depends positively on quality so that $q = q(m)$ with $q_m > 0$ and $q_m \leq 0$. This assumption implies $q = q(m) \Leftrightarrow m = m(q)$, $m_q > 0$. Therefore by contracting activity the purchaser can implicitly contract the level of quality. The benefit function of the patients is $B = B(q, m)$ with $B_q > 0$ and $B_m > 0$. Since quality is a positive function of activity, we can also write $B = B(q, m(q))$ with $\frac{\partial B}{\partial q} = \frac{\partial B}{\partial q} + \frac{\partial B}{\partial m}\frac{\partial m}{\partial q} > 0$. The provider's utility is given by the surplus: $U = py - C(q, m(q)) - \varphi(q, m(q))$. The purchaser's utility is $B\left(q, m(q)\right) - (1 + \lambda - \delta)\, pq - \delta \left[C\left(\theta, q, m(q)\right) + \varphi(q, m(q))\right]$.

Since the benefit function is increasing and concave in quantity while the cost function (the sum of the monetary and non-monetary cost) is increasing and convex in quantity, the same qualitative results of sections 2-4 are obtained.

6 Conclusions

We have investigated the optimal pricing system when hospitals differ in costs, and such differences are observable to the purchaser of health services. Costs might vary because of unavoidable differences in factor prices faced by hospitals in different locations. These differences are taken into account by the purchasers (regulators or governments) in the design of the optimal price. For example the Department of Health in England adjusts the price to reflect, at least to some extent, differences in costs. We have derived the optimal properties of such adjustments, when purchasers can use only prices to reimburse hospitals for the activity performed, i.e. lump-sum transfers are not allowed. If price is the only instrument to pay for healthcare services, then providers might hold some rents. Since rents are costly from the purchaser (and society's) perspective, the design of the optimal price needs to take into account also the potential effect of variations in prices on such rents.

We have shown that in a constrained (and realistic) institutional setting, where price is the only instrument of the purchaser (which we term the second-best scenario), providers with higher costs will be remunerated with higher tariffs if the cost function is the power function or the exponential function. This result is qualitatively similar to what we might obtain in a first-best setting where lump-sum transfers are allowed and providers never hold a rent. However, we have shown that the price in the second best is typically lower than in the first best: since a higher price implies a higher rent, the purchaser optimally sets a lower price.

We have also shown that the positive price adjustment for hospitals with higher costs is typically smaller than the additional marginal cost, whenever the marginal benefit from treatment is decreasing. While in the first best the presence of constant marginal benefit implies a constant price across providers, in the second best the same result holds if the cost function is the power function. If the cost function is exponential, the price adjustment might be positive in the second best even when the marginal benefit is constant.

Finally, we have shown that higher inequality aversion will imply an increase in the price for the high-cost providers and a reduction in the price of the low-cost providers. It also implies that the difference in the price of the high-cost provider and the price of the low-cost provider increases with inequality aversion. In other words, when inequality aversion matters, purchasers of health services are more willing to pay a higher price for the high-cost provider at the cost of reducing the price for the low-cost one.

Our conclusions remain qualitatively unchanged when the purchaser is concerned not only about quantity but also about quality.

References

[1] Chalkley M., Malcomson J.M. (1998a), "Contracting for Health Services with unmonitored Quality", The Economic Journal, 108, 1093-1110.

[2] Chalkley M., Malcomson J.M. (1998b), "Contracting for Health Services when patient demand does not reflect quality", Journal of Health Economics, 17, 1-19.

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

[4] Department of Health. (2002a), "Reforming NHS Financial flow: introducing Payment by Result."

[5] Department of Health. (2002b), "A guide to NHS Foundation Trusts", Department of Health: London. Also available at http://www.dh.gov.uk/assetRoot/04/06/04/80/04060480.pdf

[6] Department of Health. (2004), "Implementing the new system of financial flows - Payment by results: Technical Guidance 2003/04".

[7] Dranove D. (1987), "Rate-setting by DRG and hospital specialization", Rand Journal of Economics, 18(3), 417-27.

[8] Dranove D., White W.D., (1994), "Recent Theory and Evidence on Competition in Hospital Markets", Journal of Economics and Management Strategy, 3, 169-209.

[9] Ellis R. P., McGuire T.G. (1986), "Provider Behaviour under prospective reimbursement: cost sharing and supply", Journal of Health Economics, 5, 129-51.

[10] Ellis R. P. (1998), "Creaming, skimping and dumping: provider competition on intensive and extensive margins", Journal of Health Economics, 17(5), 537-55.

[11] Ma A.C. (1994), "Health care payment systems: cost and quality incentives", Journal of Economics and Management Strategy, 3(1), 93-112.

[12] Mossialos E., (2002), "Funding health care: options for Europe", Buckingham: Open University Press, European Observatory on Health Care Systems series.

[13] Mougeot M., Naegelen F., (2005), "Hospital Price regulation and expenditure cap policy", Journal of Health Economics, 24, 55-72.

[14] Rickman N., McGuire A., (1999), "Providers' reimbursement in the health care market", Scottish Journal of Political Economy, 46(1), 53-71.

[15] Shleifer A. (1985), "A Theory of Yardstick Competition", RAND Journal of Economics, The RAND Corporation, 16(3), 319-327.

[16] Shwartz M., Merrill J., Blake L. (1984), "DRG-Based Case Mix and Public Hospitals", Medical Care, 22(4), 283-299.

A Second Order Condition

The SOC is given by:

$$W_{pp} := \frac{\partial^2 W}{\partial p^2} = (B_{qq} - \delta C_{qq})\, q_p^2 + (B_q - \delta C_q)\, q_{pp} - [(1-\delta)+\lambda]\,(2q_p + p q_{pp}) < 0. \quad (34)$$

Substituting for q_p, q_{pp} we obtain:

$$W_{pp} = \frac{B_{qq} - \delta C_{qq}}{(C_{qq})^2} - (B_q - \delta C_q)\,\frac{C_{qqq}}{(C_{qq})^3} - [(1-\delta)+\lambda]\left(\frac{2}{C_{qq}} - p\frac{C_{qqq}}{(C_{qq})^3}\right) < 0. \quad (35)$$

B Effect of θ on price: $dp^s/d\theta$

By the implicit function theorem, $\frac{\partial^2 W}{\partial p \partial \theta} d\theta + \frac{\partial^2 W}{\partial p^2} dp = 0$ so that $\frac{dp^s}{d\theta} = -\frac{\partial^2 W}{\partial p \partial \theta} / \frac{\partial^2 W}{\partial p^2}$. Totally differentiating with respect to θ we obtain:

$$\begin{aligned} W_{p\theta} \quad : \quad &= \frac{\partial^2 W}{\partial p \partial \theta} & (36) \\ &= (B_{qq} - \delta C_{qq})\, q_p q_\theta + \{B_q - [(1-\delta)+\lambda]\, p - \delta C_q\}\, q_{p\theta} - \delta C_{q\theta} q_p - [(1-\delta)+\lambda]\, q_\theta \end{aligned}$$

Notice that: i) $-\delta C_{qq} q_p q_\theta - \delta C_{q\theta} q_p = 0$; ii) from the FOC of the purchaser (6), we have that $B_q - (1+\lambda-\delta)\, C_q - \delta C_q = \frac{(1+\lambda-\delta)q}{q_p}$; and Eq.(37) is obtained. Moreover, substituting q_p, q_θ, and $q_{p\theta}$ into $W_{p\theta}$ (see Equations 2, 3), generates

$$W_{p\theta} = -B_{qq}\frac{C_{q\theta}}{(C_{qq})^2} + \frac{(1+\lambda-\delta)}{C_{qq}}\left[C_{q\theta} - qC_{qq\theta} + q\frac{C_{q\theta}C_{qqq}}{C_{qq}}\right] \quad (37)$$

Substituting in $\frac{dp^s}{d\theta} = -\frac{\partial^2 W}{\partial p \partial \theta} / \frac{\partial^2 W}{\partial p^2}$, the result is obtained.

C Comparative statics with respect to inequality aversion

Recall from (29) that the optimal prices must satisfy the following first order conditions:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial p_1} &= \omega B\left(q_1\right)^{-\rho} B_{q_1} q_{p_1} - \omega\mu\left(q_1 + p_1 q_{p_1}\right) = 0 \\
\frac{\partial \mathcal{L}}{\partial p_2} &= \left(1-\omega\right) B\left(q_2\right)^{-\rho} B_{q_2} q_{p_2} - \left(1-\omega\right)\mu\left(q_2 + p_2 q_{p_2}\right) = 0 \\
\frac{\partial \mathcal{L}}{\partial \mu} &= -\omega p_1 q_1 - \left(1-\omega\right) p_2 q_2 + K = 0
\end{aligned}$$

Let $F_1 \equiv \frac{\partial \mathcal{L}}{\partial p_1}$, $F_2 \equiv \frac{\partial \mathcal{L}}{\partial p_2}$, and $F_3 \equiv \frac{\partial \mathcal{L}}{\partial \mu}$. By the implicit function theorem we have that:

$$\frac{\partial F_i}{\partial \rho} + \frac{\partial F_i}{\partial p_1}\frac{\partial p_1}{\partial \rho} + \frac{\partial F_i}{\partial p_2}\frac{\partial p_2}{\partial \rho} + \frac{\partial F_i}{\partial \mu}\frac{\partial \mu}{\partial \rho} = 0 \quad i = 1,2,3$$

In matrix format:

$$\begin{bmatrix} \frac{\partial F_1}{\partial p_1} & 0 & \frac{\partial F_1}{\partial \mu} \\ 0 & \frac{\partial F_2}{\partial p_2} & \frac{\partial F_2}{\partial \mu} \\ \frac{\partial F_3}{\partial p_1} & \frac{\partial F_3}{\partial p_2} & 0 \end{bmatrix} \begin{bmatrix} \partial p_1 \\ \partial p_2 \\ \partial \mu \end{bmatrix} = - \begin{bmatrix} \frac{\partial F_1}{\partial \rho} \\ \frac{\partial F_2}{\partial \rho}\ \partial \rho \\ 0 \end{bmatrix}$$

Using the Cramer's rule we obtain:

$$\frac{\partial p_1}{\partial \rho} = \frac{A_1}{J}, \quad \frac{\partial p_2}{\partial \rho} = \frac{A_2}{J}$$

where:

$$J = \begin{vmatrix} \frac{\partial F_1}{\partial p_1} & 0 & \frac{\partial F_1}{\partial \mu} \\ 0 & \frac{\partial F_2}{\partial p_2} & \frac{\partial F_2}{\partial \mu} \\ \frac{\partial F_3}{\partial p_1} & \frac{\partial F_3}{\partial p_2} & 0 \end{vmatrix}, \ A_1 = \begin{vmatrix} -\frac{\partial F_1}{\partial \rho} & 0 & \frac{\partial F_1}{\partial \mu} \\ -\frac{\partial F_2}{\partial \rho} & \frac{\partial F_2}{\partial p_2} & \frac{\partial F_2}{\partial \mu} \\ 0 & \frac{\partial F_3}{\partial p_2} & 0 \end{vmatrix}, \ A_2 = \begin{vmatrix} \frac{\partial F_1}{\partial p_1} & -\frac{\partial F_1}{\partial \rho} & \frac{\partial F_1}{\partial \mu} \\ 0 & -\frac{\partial F_2}{\partial \rho} & \frac{\partial F_2}{\partial \mu} \\ \frac{\partial F_3}{\partial p_1} & 0 & 0 \end{vmatrix}$$

Therefore, computing the partial derivatives in A_1, A_2 and J we obtain:

$$\begin{aligned}
\frac{\partial F_1}{\partial p_1} &= \omega q_{p_1}^2 \frac{\left(B_{q_1 q_1} - \rho B_{q_1}/B_1\right)}{B_1^{\rho}} - \omega\mu\left(2q_{p_1} + p_1 q_{p_1 p_1}\right) < 0 \\
\frac{\partial F_2}{\partial p_2} &= (1-\omega)\, q_{p_2}^2 \frac{\left(B_{q_2 q_2} - \rho B_{q_2}/B_2\right)}{B_2^{\rho}} - \mu\,(1-\omega)\left(2q_{p_2} + p_2 q_{p_2 p_2}\right) < 0 \\
\frac{\partial F_1}{\partial \mu} &= -\omega\left(q_1 + p_1 q_{p_1}\right) < 0;\ \frac{\partial F_2}{\partial \mu} = -\,(1-\omega)\left(q_2 + p_2 q_{p_2}\right) < 0 \\
\frac{\partial F_3}{\partial p_1} &= -\omega\left(p_1 q_{p_1} + q_1\right) < 0;\ \frac{\partial F_3}{\partial p_2} = -\,(1-\omega)\left(p_2 q_{p_2} + q_2\right) < 0 \\
\frac{\partial F_1}{\partial \rho} &= -\frac{B_{q_1} q_{p_1} \ln(B_1)}{B_1^{\rho}};\ \frac{\partial F_2}{\partial \rho} = -\frac{B_{q_2} q_{p_2} \ln(B_2)}{B_2^{\rho}}
\end{aligned}$$

Plugging then into A_1, A_2, A_3 and J we have:

$$\begin{aligned}
A_1 &= \frac{\partial F_3}{\partial p_2}\left(\frac{\partial F_1}{\partial \rho}\frac{\partial F_2}{\partial \mu} - \frac{\partial F_1}{\partial \mu}\frac{\partial F_2}{\partial \rho}\right) \\
&= -\,(1-\omega)\left(q_2 + p_2 q_{p_2}\right)\left(\ln(B_1)\frac{B_{q_1} q_{p_1}}{B_1^{\rho}}\left(q_2 + p_2 q_{p_2}\right) - \frac{B_{q_2} q_{p_2} \ln(B_2)}{B_2^{\rho}}\left(q_1 + p_1 q_{p_1}\right)\right) \\
&= -\,(1-\omega)\left(q_1 + p_1 q_{p_1}\right)\left(q_2 + p_2 q_{p_2}\right)^2\left(\ln(B_1)\frac{B_1^{-\rho} B_{q_1} q_{p_1}}{q_1 + p_1 q_{p_1}} - \ln(B_2)\frac{B_2^{-\rho} B_{q_2} q_{p_2}}{q_2 + p_2 q_{p_2}}\right)
\end{aligned}$$

>From the FOCs we have: $\frac{B(q_1)^{-\rho} B_{q_1} q_{p_1}}{q_1 + p_1 q_{p_1}} = \frac{B(q_1)^{-\rho} B_{q_1} q_{p_1}}{q_1 + p_1 q_{p_1}} = \mu$ and

$$A_1 = -\,(1-\omega)\left(q_1 + p_1 q_{p_1}\right)\left(q_2 + p_2 q_{p_2}\right)^2 \mu\left(\ln(B_1) - \ln(B_2)\right)$$

Similarly,

$$\begin{aligned}
A_2 &= \frac{\partial F_3}{\partial p_1}\left(\frac{\partial F_1}{\partial \mu}\frac{\partial F_2}{\partial \rho} - \frac{\partial F_1}{\partial \rho}\frac{\partial F_2}{\partial \mu}\right) \\
&= -\omega\left(p_1 q_{p_1} + q_1\right)\left(\left(q_1 + p_1 q_{p_1}\right)\frac{B_{q_2} q_{p_2} \ln(B_2)}{B_2^{\rho}} - \frac{B_{q_1} q_{p_1} \ln(B_1)}{B_1^{\rho}}\left(q_2 + p_2 q_{p_2}\right)\right) \\
&= -\omega\left(q_2 + p_2 q_{p_2}\right)\left(p_1 q_{p_1} + q_1\right)^2\left(\ln(B_2)\frac{B_2^{-\rho} B_{q_2} q_{p_2}}{q_2 + p_2 q_{p_2}} - \ln(B_1)\frac{B_1^{-\rho} B_{q_1} q_{p_1}}{q_1 + p_1 q_{p_1}}\right) \\
&= -\omega\left(q_2 + p_2 q_{p_2}\right)\left(p_1 q_{p_1} + q_1\right)^2 \mu\left(\ln(B_2) - \ln(B_1)\right)
\end{aligned}$$

$$J = -\frac{\partial F_1}{\partial \mu}\frac{\partial F_2}{\partial p_2}\frac{\partial F_3}{\partial p_1} - \frac{\partial F_1}{\partial p_1}\frac{\partial F_2}{\partial \mu}\frac{\partial F_3}{\partial p_2} > 0$$

Therefore, $dp_1/d\rho$ and $dp_2/d\rho$ simplify to:

$$\begin{aligned}
\frac{dp_1}{d\rho} &= \frac{-\left(1-\omega\right)\left(q_1+p_1q_{p_1}\right)\left(q_2+p_2q_{p_2}\right)^2\mu\left(\ln(B_1)-\ln(B_2)\right)}{J} < 0 \\
\frac{dp_2}{d\rho} &= \frac{-\omega\left(q_2+p_2q_{p_2}\right)\left(p_1q_{p_1}+q_1\right)^2\mu\left(\ln(B_2)-\ln(B_1)\right)}{J} > 0.
\end{aligned}$$