

Tangerås, Thomas P.

Working Paper

Regulation of Cost and Quality under Yardstick Competition

IUI Working Paper, No. 573

Provided in Cooperation with:

Research Institute of Industrial Economics (IFN), Stockholm

Suggested Citation: Tangerås, Thomas P. (2002) : Regulation of Cost and Quality under Yardstick Competition, IUI Working Paper, No. 573, The Research Institute of Industrial Economics (IUI), Stockholm

This Version is available at:

<https://hdl.handle.net/10419/94924>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

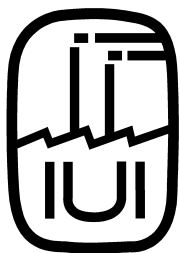
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



THE RESEARCH INSTITUTE OF INDUSTRIAL ECONOMICS

Working Paper No. 573, 2002

Regulation of Cost and Quality under Yardstick Competition

by Thomas P. Tangerås

IUI, The Research Institute of Industrial Economics
P.O. Box 5501
SE-114 85 Stockholm
Sweden

Regulation of cost and quality under yardstick competition*

Thomas P. Tangerås[†]

January 15, 2002

Abstract

This paper analyses simultaneous regulation of cost and quality when firms have private, correlated information about productivity and the regulator receives a signal about quality. It is shown that managerial effort and expenditures on quality are positively correlated in the optimal contract. The higher is firm productivity the more should the firm spend on quality improvement and the more efficiently should it produce. Optimal yardstick competition reduces distortion of both effort and quality. Under product market competition expenditures on quality should be increasing in the firm's own productivity and decreasing in the competitor's productivity.

JEL Classification: D82, I11, I18, L51.

Keywords: Yardstick competition, quality, regulation.

*The major part of this research was undertaken while I was visiting University des Sciences Sociales in Toulouse. I am very grateful to colleagues and staff for their hospitality during my stay there and especially to Emmanuelle Auriol and Jean Tirole for their helpful comments on an earlier version of this paper. I also thank seminar participants at IUI for their comments and David Sundén for computer programming. The financial support of the Jan Wallander and Tom Hedelius Research Foundation is gratefully acknowledged.

[†]The Research Institute of Industrial Economics (IUI), Box 5501, S-114 85 Stockholm, Sweden. Tel: +46 (0)8 665 45 26, Fax: +46 (0)8 665 45 99. *Mailto:* thomas.tangeras@iui.se.

1. Introduction

In 1983 the US government in an attempt to curb escalating hospital costs,¹ replaced their retrospective reasonable-charge reimbursement system by the more high-powered Prospective Payment System (PPS). Under the PPS hospitals receive a flat rate for each admission, the amount depending on the diagnostic group to which the patient is assigned. The rate is based on a nationwide average of the cost incurred for that specific diagnostic group. As opposed to the case within the old cost-reimbursement system, hospitals are now residual claimants to their own cost-savings, which is supposed to provide strong incentives for cost-reduction. The average nationwide costs of treatment should thus be expected to have decreased following the introduction of the PPS - and the reimbursement levels with them - since the rates are calculated on the basis of the cost average. In this way, cost savings are de facto passed onto the tax payers. This is known as yardstick competition or relative performance evaluation.² Encouraged by its apparent success in reducing Medicare expenses on inpatients,³ federal government in the year 2000 went on to implementing a PPS even for outpatients.⁴

¹In the period 1960-84 the growth in overall Medicare outlays outperformed the growth in the consumer price index by a factor of 6 (Feinglass and Holloway, 1991).

²See Shleifer (1985) for a formal analysis of the PPS system. The seminal papers on yardstick competition were Baiman and Demski (1982) and Holmström (1982). They were later followed by Nalebuff and Stiglitz (1983), Demski and Sappington (1984), Mookherjee (1984) and Cremer and McLean (1988). Recent contributions include Auriol and Laffont (1992), Auriol (1993), Dalen (1998), Sobel (1999) and Laffont and Martimort (2000).

³For example, the average length of stay for Medicare patients dropped by a fourth from 1980 to 1985 (Feinglass and Holloway, 1991) without any sign of discharged patients being sicker than before (Carroll and Erwin, 1987).

⁴The use of yardstick competition in the US health sector has been viewed with interest even in Europe. Currently, Belgium and France are considering introducing yardstick competition in the health sector. On a broader level, the EU commission has initiated a program to investigate the use of benchmarking, i.e. best-practice rules, for public agencies. Benchmarking is less formal than yardstick competition, but both build on the idea that performance is correlated across firms or public agencies and that all information should be used in order to improve efficiency and reduce cost.

Providing incentives for cost-reduction only, concern has been raised as to the effect of the PPS on hospitals' incentives for providing quality (e.g. Johnson, 1984; Broyles and Rosko, 1985). The fear is that managers will devote most of their attention to reducing cost and that doctors will cut corners in medical treatment, for example by discharging patients prematurely, so as to achieve ambitious cost-targets. The idea that economic agents assigned to perform multiple tasks (e.g. provide high quality health care at low cost) tend to put too much effort into tasks for which they receive the highest compensation, was formally developed and analyzed by Holmström and Milgrom (1991). One of their findings is that the compensation scheme should be low-powered if the agent is assigned tasks that are very difficult to measure (quality provision in health care constituting an archetypal task). Since the PPS rewards cost-reduction more heavily than does a retrospective cost-plus system, one would expect quality of care to have declined following the introduction of the PPS. The picture is a bit more complicated, however. The PPS pays a fixed rate for every patient admitted. This means that a *sales incentive* arises if a hospital can influence demand for its services by investing in quality improvement. In general this may lead to under- or over-provision of quality, depending on the elasticity of demand with respect to quality (Tirole, 1988).

In view of the arguments above, one would expect the net effect of the PPS on quality to have been ambiguous. This is in line with the existing empirical evidence. In a survey of the literature, Feinglass and Holloway (1991, p.107) conclude: "To date, there is little direct, generalizable evidence that PPS has reduced the quality of care for Medicare patients." What should be obvious, is that it would be entirely coincidental if the provided level of quality turned out

to be the socially optimal one. This calls for joint regulation of cost and quality.

This paper is the first, to my knowledge, to study the effect of yardstick competition on optimal regulation of quality. Since the health sector is a major candidate for implementation of yardstick competition, the following two key assumptions are made. First, there are no prices. Demand for a firm's product is determined entirely by the quality of its product as perceived by potential customers. This reflects the fact that in many countries, a large part of the expenditures on health care is covered by the government budget or by insurance companies.⁵ Second, quantity is unregulated. This is the case with the PPS.⁶ A few additional assumptions are made. Information about quality comes to customers in the shape of a signal. We can think of this as a quality indicator compiled by the government. Firm productivity is private information, but correlated across firms, which allows the regulator to implement yardstick competition to extract rent. Management exerts unobservable effort to contain cost.

I study first the case with independent regulation of regional monopolies, a case which brings out some general results. Optimal regulation of cost and quality are found to be explicably linked. Raising expenditures on quality improvement boosts expected demand for the firm's product, which, in turn, increases the scope for cost savings. Thus, *the higher is expected quality of its output, the more efficient the firm is required to be*. Furthermore, *at the social optimum, expenditures on quality are increasing in productivity*. The reason is two-fold. First, marginal production cost is decreasing in productivity, making the supply of quality cheaper. Second, the more productive is a firm, the less compensation it requires

⁵For example, the data for the US for 1993 show that a mere 2.8% of total expenditures on hospital care were out-of-pocket payments (Table 17.2, Folland, Goodman and Stano, 1997).

⁶Moreover, it is probably politically impossible to implement a regulatory policy that regulates, say, the number of emergency operations to undertake during a year.

in order to be willing to supply quality. This implies that the informational rent necessary to preserve truth-telling incentives under quality provision is decreasing in productivity, hence more quality is supplied.

Effort and quality are both distorted downward so as to reduce informational rent to firms. As is well known, yardstick competition reduces informational rent by filtering out private information. This calls for less distortion of both quality and effort. Thus, *expected quality is higher and managerial effort is distorted less under yardstick competition than under independent regulation of firms*. The savings in rent achieved by yardstick competition are spent both on increasing efficiency in production and increasing quality.

When firms are competing for customers, the regulatory authority faces the problem of how to distribute expenditures on quality improvement across firms. Should they all spend equally much, and if not, who should spend more? The answer to this question depends on firms' relative productivity. As in the case with regional monopoly, expenditures on quality improvement are increasing in the firm's own productivity. The new result is that *a firm should spend more on quality improvement, the less productive is the other firm*. From society's point of view, the most productive firm should serve the majority of the customers. This is achieved by increasing expenditures on quality improvement by the most productive firm and thereby its expected market share.

The fundamental lesson to be drawn from the results above is that *firms should be treated unequally depending on their relative efficiency*. The provision of high quality goods should be left to the high-productivity firms that are also to be given strong incentives to contain cost. At the other end of the scale, low-productivity firms produce goods of low expected quality and are allowed more slack in pro-

duction. The intuition is straightforward. Production of quality is cheaper the more productive is a firm, and the cost of supplying quality should be mitigated by providing management with stronger incentives for producing efficiently.⁷

The remainder of this paper is organized as follows: section 2 formulates the model. Section 3 characterizes and compares optimal regulatory contracts in an increasingly complex environment. The first part considers independent regulation of regional monopolies, the second analyses yardstick competition among regional monopolies, the final part introduces competition among firms into the yardstick competition scheme. Section 4 discusses practical implementation. Section 5 takes a primary look at data from US acute hospitals to study the effects of yardstick competition. Finally, section 6 concludes. Tedious proofs and derivations are collected in the appendix.

2. The model

2.1. Producers

There are two firms, indexed by $i \in \{1, 2\}$, each of whom produces a good in amount q_i . The goods may differ in terms of perceived quality (more on that

⁷So far little work has been done on regulation of quality. An early contribution in the field was Lewis and Sappington (1988) who examined the effect of verifiability versus non-verifiability (i.e. whether quality is contractible) on optimal regulation. Laffont and Tirole (1993) consider two models, one with ex ante and one with ex post observable, but unverifiable quality. They find necessary and sufficient conditions for incomplete information to lead to downward distortion of quality and study how the power of the incentive scheme changes as a function of demand and supply parameters. The models above consider the monopoly case, hence cannot capture the effects of yardstick and product market competition. Further, the relationship between quality and cost efficiency has remained unexamined until now and so has the effect of productivity changes on optimal regulation. More recently Auriol (1998) has analysed the effect of competition on quality provision. She derives the optimal market structure when there is a concern for quality and quality displays public good-like features. In the present setting, the market structure is exogenously given and quality is a private good.

below), but are otherwise homogeneous. The firms considered could either be regional monopolies or competitors. Examples of the first type include elementary schools or regional hospitals, and the second type include hospitals in a metropolitan area or universities competing for students on a national level. Firm i 's total production cost is

$$C_i = (\beta_i + s_i - e_i)q_i. \quad (2.1)$$

It operates at constant marginal cost $c(\beta_i, s_i, e_i)$ that depends positively on an exogenous productivity parameter β_i , positively on expenditures s_i on service or quality improvement and negatively on effort e_i exerted by management to keep costs down.⁸ The regulator observes each producer's aggregate cost C_i and demand q_i only. He cannot disaggregate β_i , s_i and e_i without an appropriate incentive contract. Thus, quality is unverifiable. Management's disutility of effort is

$$\psi(e_i), \text{ with } \psi \geq 0, \psi' > 0, \psi'' > 0, \psi''' \geq 0.$$

Disutility is always non-negative, it is increasing in effort at an increasing rate. Non-negativity of the third derivative is sufficient for concavity of the social planner's problem.⁹

2.1.1. Productivity

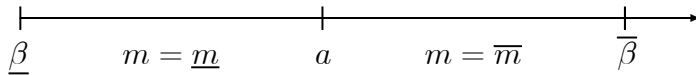
In the education sector as well as the health sector production cost is to some extent determined by factors common to all firms in the industry. For example,

⁸All results derived in this paper carry over to the case for which the cost of quality improvement is fixed, i.e. $C_i = (\beta_i - e_i)q_i + s_i$.

⁹The model presented here, with the cost of services s_i being monetary, is formally equivalent to one in which quality improvement is a result of managerial effort and embedded in the disutility function as $\psi(e_i + s_i)$, see chapter 4 of Laffont and Tirole (1993).

the variation across hospitals in the treatment of well-known diseases is probably low and operation techniques and procedures there for all to use. At the same time firms are unlikely to be identical since production cost is affected by local conditions also. Under the US Medicare system, for example, hospital reimbursements are adjusted to account for idiosyncrasies in location (urban versus rural) and differences in local wages. In the model these effects are captured by assuming that the cost or productivity parameter β_i consists of a common part m and an idiosyncratic part ε_i . Total productivity is given by the weighted average $\beta_i = \alpha m + (1 - \alpha)\varepsilon_i$. *Industry-specific (IS) productivity* m is high ($m = \underline{m}$) with probability v and low ($m = \overline{m} > \underline{m}$) with probability $1 - v$. *Firm-specific (FS) productivity* ε_i is continuously distributed on the interval $[\underline{\varepsilon}, \overline{\varepsilon}]$ with strictly positive density $g(\cdot)$ and cumulative distribution $G(\cdot)$. The hazard rate G/g is assumed to be increasing, and it is further assumed that m , ε_1 and ε_2 are stochastically independent. Finally, α is common knowledge and equal to $\alpha = (\overline{\varepsilon} - \underline{\varepsilon})/(\overline{m} - \underline{m} + \overline{\varepsilon} - \underline{\varepsilon})$. Thus, β_i is distributed over two connected intervals as illustrated in the figure below:

Figure 1



When IS productivity is high, both firms have total productivity somewhere in the interval $[\underline{\beta}, a]$, with $\underline{\beta} = \alpha \underline{m} + (1 - \alpha)\underline{\varepsilon}$ and $a = \alpha \underline{m} + (1 - \alpha)\overline{\varepsilon}$. Conversely, low IS productivity implies both firms having total productivity in $[a, \overline{\beta}]$, with

$a = \alpha \overline{m} + (1 - \alpha) \underline{\varepsilon}$ and $\overline{\beta} = \alpha \overline{m} + (1 - \alpha) \overline{\varepsilon}$. The marginal distribution of β_i on $[\underline{\beta}, \overline{\beta}]$ is a convolution of the distributions of m and ε_i . Denote by $f(\cdot)$ and $F(\cdot)$ the density function and cumulative distribution of β_i , respectively. The hazard rate F/f is assumed to be increasing. Finally, write $\widehat{F}(\cdot)$ the joint cumulative distribution of $\beta = (\beta_1, \beta_2)$.¹⁰

The regulator can take advantage of the fact that productivity is linked across firms to extract information about productivity. Consider the following yardstick competition scheme: a firm that outperforms the rest of the industry (appears to be of type $\beta_i \in [\underline{\beta}, a]$ whenever the other firm appears to be of type $\beta_{-i} \in [a, \overline{\beta}]$) is rewarded a bonus. Under this scheme nobody wants to be the under-performer if the bonus is sufficiently high, hence both firms perform well if their type is $\beta_i \in [a, \overline{\beta}]$. This is achieved at no cost to society since the firms compete away the bonus. At the same time the bonus can be set sufficiently low so as to prevent low-productivity types from "over-working" (appear to be of type $\beta_i \in [\underline{\beta}, a]$ when their type is in fact $\beta_i \in [a, \overline{\beta}]$).¹¹ If the regulator considers each firm independently, he has no information about their type. In particular, a high productivity type $\beta_i \in [\underline{\beta}, a]$ can relax and credibly pretend to be a low productivity type $\beta_i \in [a, \overline{\beta}]$, which is impossible under yardstick competition. In this way yardstick competition reduces firms' informational advantage.

¹⁰The stochastic structure described above was first utilized by Auriol and Laffont (1992). They show that monotonicity of G/g and the condition $vg(\overline{\varepsilon}) \geq (1 - v)g(\underline{\varepsilon})$ are sufficient to guarantee monotonicity of F/f under the described set-up. The structure has later been used to study aspects of yardstick competition such as investment incentives (Dalen, 1998) and collusion (Tangerås, 2001). See also Auriol (1993 and 2000).

¹¹This, owing to the fact that it is more costly, in terms of disutility of effort, for a low-productivity firm to perform well than for a high-productivity one.

2.1.2. Quality

Denote by σ_i the quality of firm i 's product as perceived by customers. σ_i may be real quality or an informative signal thereof. We can think of σ_i as a quality indicator constructed by the authorities on the basis of information about the firm. For example, a quality indicator for hospitals could be based on patients' assessment of their treatment,¹² on accounts of injuries caused by medical management and on external reviews of doctors' competence. Within the education sector, teaching could be evaluated on the basis of student achievement and progression, curriculum design and organization, to name a few possibilities.¹³

Firm i has a dichotomous choice between spending a lot ($s_i = \bar{s}$) on improving quality or nothing at all ($s_i = 0$). I assume that perceived quality is high (the signal is favourable: $\sigma_i = \sigma^h$) with probability $\theta(s_i)$ and low (the signal is unfavourable: $\sigma_i = \sigma^l$) with probability $1 - \theta(s_i)$. Expected quality is increasing in s_i , i.e. $\theta(\bar{s}) > \theta(0)$. Signals are assumed to be independent across firms.

Why would hospitals and universities invest in quality? Apart from the fact that they may take pride in providing quality, profits may depend directly on their assessed quality. First, high-quality producers could receive more transfers than low-quality ones.¹⁴ Second, quality may affect demand. Since patients presumably prefer the best hospitals and prospective students the best universities, one would expect a high quality rating to bolster demand.

¹²Under the US Medicare system, Peer Review Organizations and the hospitals themselves conduct utilization reviews to detect whether the admission and length-of-stay are appropriate and whether patients receive proper care.

¹³In England and Northern Ireland, the Quality Assessment Agency reviews universities and colleges based on these and other aspects.

¹⁴This is the case with higher education in the UK. The higher is an institution's score in the teaching and research assessment, the more funding it subsequently receives.

2.2. Customers

Increased expenditures on quality have two effects on customers. Existing customers are better off because they are offered a product of higher expected quality, e.g. patients receive better care and students benefit from improved teaching and tutoring. This effect alone may be sufficient for society to prefer high quality hospitals, schools and universities. Second, demand may increase. If the prospects of getting a good job increases following an upgrade of the university, for example, more students are likely to enrol. Moreover, firms may be in competition with one another for patients or students.

In many European countries education and hospital treatment (acute treatment, in particular) are free in the sense that these services are paid over the tax bill. In the US many kinds of treatments are paid over the health insurance. Under such a system there are no prices to govern demand; it is determined entirely by perceived quality. Let q^{hl} be demand for a good that is perceived to be of high quality given that the competitor's good is perceived to be of low quality, and define q^{lh} , q^{hh} and q^{ll} in a similar fashion according to perceived quality. $q^{hh} \geq q^{lh}$ and $q^{hl} \geq q^{ll}$ owing either to a competition effect and/or a general positive elasticity of demand with respect to quality. Expected demand facing firm i is

$$q_i(\mathbf{s}) = \theta(s_i)[\theta(s_{-i})q^{hh} + (1 - \theta(s_{-i}))q^{hl}] \\ + (1 - \theta(s_i))[\theta(s_{-i})q^{lh} + (1 - \theta(s_{-i}))q^{ll}]$$

and a function of the two firms' expenditures $\mathbf{s} = (s_i, s_{-i})$ on quality improvement. It follows from $\theta(\bar{s}) > \theta(0)$ and the assumptions on realized demand that expected demand is non-decreasing in expenditures on quality, i.e. $q_i(\bar{s}, s_{-i}) \geq q_i(0, s_{-i})$ for all $s_{-i} \in \{0, \bar{s}\}$. When the firms in question are regional monopolies, there is

no competition effect, hence $q^{hh} = q^{hl} = q^h$ and $q^{lh} = q^{ll} = q^l$. Consequently, expected demand $q_i(\mathbf{s})$ simplifies to $q(s_i) = \theta(s_i)q^h + (1 - \theta(s_i))q^l$ in the no-competition case.

Expected consumer surplus is given by

$$V(\mathbf{s}) = \theta(s_i)[\theta(s_{-i})V^{hh} + (1 - \theta(s_{-i}))V^{hl}] \\ + (1 - \theta(s_i))[\theta(s_{-i})V^{lh} + (1 - \theta(s_{-i}))V^{ll}]$$

with $V^{hl} = V^{lh}$ being consumer surplus when the two firms offer different qualities, and V^{hh} (V^{ll}) consumer surplus when both firms offer a high (low) quality product. Consumer surplus is increasing in quality ($V^{hh} > V^{hl}$ and $V^{lh} > V^{ll}$) since existing customers are offered a better product and potentially more customers are attracted into the market.¹⁵ It is easy to check that expected consumer surplus is increasing in expenditures on quality s_i , i.e. $V(\bar{s}, s_{-i}) > V(0, s_{-i})$ for all $s_{-i} \in \{0, \bar{s}\}$. In the no-competition case, consumer surplus in each market equals $V^{hh} = V^{hl} = V^h$ or $V^{lh} = V^{ll} = V^l < V^h$, and expected consumer surplus in market i is $V(s_i) = \theta(s_i)V^h + (1 - \theta(s_i))V^l$.

2.3. The regulatory setting

The scope of regulation is to induce firms to serve their customers in a cost-efficient way and provide a certain level of service or quality at minimum cost to society. It is assumed that demand is unregulated. The assumption of unregulated demand seems particularly appropriate for the health sector. A regulatory policy specifying in advance the number of heart transplants to be performed or child deliveries

¹⁵Consider this a reduced form of a Hotelling location model with two firms located at a certain geographical distance from each other and buyers distributed in between. Customers select producers based on quality considerations and their relative distance from the producer. The higher is the perceived quality of a firm's product the more customers find the trip to that firm worthwhile. Transportation cost is embedded in consumer surplus V .

to be assisted by a hospital during a period sounds politically infeasible.¹⁶ To avoid unnecessary complications, it is further assumed that each firm has sufficient capacity to meet its demand.

By the Revelation Principle, the regulatory authority can restrict itself to offering a direct revelation mechanism $D = \{\mathbf{t}, \mathbf{c}\}$. D is a menu of contracts that for each possible productivity report $\mathbf{b} = (b_1, b_2)$ specifies a vector of average cost targets $\mathbf{c} = (c_1(\mathbf{b}), c_2(\mathbf{b}))$ and a vector $\mathbf{t} = (t_1^l(\mathbf{b}), t_1^h(\mathbf{b}), t_2^l(\mathbf{b}), t_2^h(\mathbf{b}))$ of unit transfers. Transfers depend not only on \mathbf{b} , but also on the quality signal: firm i receives unit transfer $t_i^l(\mathbf{b})$ whenever $\sigma_i = \sigma^l$, and $t_i^h(\mathbf{b})$ otherwise, given that the two firms have already reported productivity \mathbf{b} .

I study a static game in which there is production only once. First, productivity is revealed and is the private information of each firm. Second, the regulator commits to the regulatory contract D . Third, productivity is reported and firms receive their specified contract if they agree to produce. The regulatory contract chosen by each firm is common knowledge. A firm may choose to shut down and receive reservation utility 0. To simplify matters, it is assumed that production by both firms is always profitable in equilibrium.¹⁷ Fourth, firms select the amount of cost-reducing effort to undertake and decide on the level of expenditures on quality. Finally, the signal of quality is revealed, demand is realized, and transfers are paid out as specified by the regulatory contract.

Firms are risk-neutral. Firm rent U_i is thus given by operating profit minus

¹⁶Under the US Medicare system hospitals are paid a fixed amount for each procedure undertaken. Quantity is indirectly regulated through Certificate-of-Need laws under which hospitals are required to obtain official approval before undertaking investments in capacity. The purpose is to prevent investments in over-capacity.

¹⁷Actually, one of the purposes of the PPS was to get rid of unprofitable hospitals. The regulatory authority should weight the cost of an inefficient hospital up against the benefit of reduced transportation costs for patients and the future ability to make performance comparisons. The latter affects the incentive for cost reduction and quality provision.

disutility of effort:

$$U_i = (t_i - c(\beta_i, s_i, e_i))q_i - \psi(e_i). \quad (2.2)$$

Welfare equals:

$$W = V + \sum_i U_i - (1 + \lambda) \sum_i t_i q_i, \quad (2.3)$$

consumer surplus V plus the sum of firm utility U_i minus the social cost of transfers $t_i q_i$. λ is the positive shadow price on public funds. As is now standard (2.1), (2.2) and (2.3) are manipulated to yield:

$$W = V - (1 + \lambda) \sum_i ((\beta_i + s_i - e_i)q_i + \psi(e_i)) - \lambda \sum_i U_i \quad (2.4)$$

3. Regulation

The regulator influences by the choice of transfers and cost targets the amounts of cost-reducing effort $e_i(\cdot)$ and expenditures on quality $s_i(\cdot)$ undertaken by the two firms. Let $q_i^w(\mathbf{b}) = \theta(s_{-i}(\mathbf{b}))q^{wh} + (1 - \theta(s_{-i}(\mathbf{b})))q^{wl}$ be firm i 's expected demand contingent on $\sigma_i = \sigma^w$ ($w \in \{l, h\}$), on the productivity reports \mathbf{b} and on $-i$'s subsequent choice $s_{-i}(\mathbf{b})$ of expenditures in quality improvement. Firm i 's expected profit is $\pi_i^w(\mathbf{b}) = (t_i^w(\mathbf{b}) - c_i(\mathbf{b}))q_i^w(\mathbf{b})$.

Suppose the vector of true types is $\boldsymbol{\beta}$, that firm $-i$ has truthfully reported its type, that i has reported b_i and holds the belief that the competitor will spend $s_{-i}(b_i, \beta_{-i})$ on quality improvement. Firm i subsequently chooses expenditures $s_i \in \{0, \bar{s}\}$ on quality improvement so as to maximize rent

$$U_i(s_i, b_i, \boldsymbol{\beta}) = \theta(s_i)\pi_i^h(b_i, \beta_{-i}) + (1 - \theta(s_i))\pi_i^l(b_i, \beta_{-i}) - \psi(\beta_i - c_i(b_i, \beta_{-i}) + s_i). \quad (3.1)$$

Given that the firm spends s_i on quality, management must exert effort $\beta_i - c_i(b_i, \beta_{-i}) + s_i$ to reach the cost target $c_i(b_i, \beta_{-i})$. Hence, the disutility of effort

in eq. (3.1). Let $s_i(b_i, \beta)$ be the choice of service that maximizes (3.1) and write $U_i(b_i, \beta) = U_i(s_i(b_i, \beta), b_i, \beta)$. By a natural extension of this notation, $U_i(\beta) = U_i(\beta_i, \beta)$, $s_i(\beta) = s_i(\beta_i, \beta)$, and $e_i(\beta)$ is the effort it takes to reach cost-target $c_i(\beta)$ given $s_i(\beta)$.

The regulator chooses the mechanism D so as to maximize expected welfare

$$W = \int_{\underline{\beta}}^{\bar{\beta}} \int_{\underline{\beta}}^{\bar{\beta}} [V(\mathbf{s}(\beta)) - (1 + \lambda) \sum_i ((\beta_i + s_i(\beta) - e_i(\beta)) q_i(\mathbf{s}(\beta)) + \psi(e_i(\beta))) - \lambda \sum_i U_i(\beta)] d\hat{F}(\beta) \quad (3.2)$$

subject to three constraints.¹⁸ First, the regulator must provide firms with the correct incentive for quality improvement:

$$U_i(\beta) = \max\{U_i(\bar{s}, \beta), U_i(0, \beta)\} \quad \forall \beta \in [\underline{\beta}, a]^2 \cup [a, \bar{\beta}]^2, i \in \{1, 2\}. \quad (\text{QIC})$$

The incentive compatibility constraint QIC on service provision states that it must be optimal for both firms and all types to provide the socially preferred level of services, conditional on the expectation that the other firm does so and on truthful productivity reports. Second, firms must prefer truth-telling to lying:

$$E[U_i(\beta)|\beta_i] \geq E[U_i(b_i, \beta)|\beta_i] \quad \forall (b_i, \beta_i) \in [\underline{\beta}, \bar{\beta}]^2, i \in \{1, 2\}. \quad (\text{TIC})$$

$E[\cdot|\beta_i]$ is the expectations operator over β_{-i} , conditional on the type β_i . The incentive compatibility constraint TIC on truth-telling states that the firm cannot strictly benefit from misrepresenting its type, conditional on its information about the other type, on the expectation that the competitor truthfully reports its type and on QIC. Third, firms must prefer operating to shutting down:

$$E[U_i(\beta)|\beta_i] \geq 0 \quad \forall \beta_i \in [\underline{\beta}, \bar{\beta}], i \in \{1, 2\}. \quad (\text{IR})$$

¹⁸Note that the regulator implements a pure strategy mechanism. Firms do not randomize between \bar{s} and 0. As we shall see below, the optimal regulatory mechanism does indeed prescribe pure strategies for almost all types.

The assumption that firms choose expenditures on quality improvement after having observed the regulatory contracts, creates an interesting signalling problem when firms are in direct competition. By lying about productivity a firm might commit to spending a lot on quality improvement, owing to the transfer structure induced by the selected regulatory contract. This might scare the manager of the competing firm into reducing expenditures on quality, i.e. giving up market shares. Thus, lying about productivity may gain a firm the competitive edge in a market. Suppose firms have reported \mathbf{b} and that the regulator prefers the two firms to spend $\{s_1(\mathbf{b}), s_2(\mathbf{b})\}$ on quality improvement if the vector of true types is indeed \mathbf{b} . Achieving this may be potentially difficult owing to the beliefs firms may have about the other's subsequent actions. As shown in appendix A.1, signalling poses no real problem in the analysis, for the following reason. For any cost-target $c_i(\mathbf{b})$, the regulator achieves the preferred division between cost-reduction $e_i(\mathbf{b})$ and quality improvement $s_i(\mathbf{b})$ by manipulating the variance in profits (i.e. changing $t_i^h(\mathbf{b})$ and $t_i^l(\mathbf{b})$). If the variance in profits is zero, the firm invests nothing in quality improvement since it is fully insured against shocks to demand. Conversely, the firm invests a lot ($s_i = \bar{s}$) if the variance is large. Importantly, the division can be achieved irrespectively of the firm's true type and its subjective beliefs about the competitor's actions. Since firms are risk-neutral, it is unnecessary to compensate them for changes in risk. The implication of this is that QIC is non-binding. Given that TIC and IR are met, the regulator can implement the preferred level of services $\{s_1(\mathbf{b}), s_2(\mathbf{b})\}$ by the appropriate choice of transfers.¹⁹

¹⁹This also means that the regulator cannot benefit from withholding information about one firm from the other. There is no loss in welfare in publishing the regulatory contracts.

3.1. Independent regulation of regional monopolies

This section considers cost and quality regulation when firms are regulated separately, and there is no competition among them for customers. Apart from being relevant on its own, the analysis brings out some general features about cost/quality regulation that prove useful in understanding the effect of yardstick competition.

The regulator offers a menu of contracts $D = \{t_h(b), t_l(b), c(b)\}$ that determines unit transfers as a function of reported productivity and perceived quality, and a marginal cost target as a function of reported productivity.²⁰ From the previous discussion, we know that QIC is non-binding. Further, Laffont and Tirole (1993) show that for the monopoly case, TIC and IR reduce to the following three necessary and sufficient conditions: (i) firm rent be given by

$$U(\beta) = \int_{\beta}^{\bar{\beta}} \psi'(e(x))dx + U(\bar{\beta}); \quad (3.3)$$

(ii) $U(\bar{\beta}) \geq 0$ and (iii) $\beta - e(\beta)$ be non-decreasing in β .

Due to risk-neutrality, rent is independent of the level of quality supplied by the firm. This does not, however, imply that quality is free. For a given cost-target $c(\beta)$, an increase in $s(\beta)$ implies an increase in $e(\beta)$, which is costly for the manager. Consequently, the regulator must allow the firm a higher operating profit in order to induce it to supply quality. This change does not affect the rent of type β , but it increases the rent of all types that are more efficient than β , owing to the extra profit firms can make from mimicking the type β firm. Thus, the choice of quality affects informational rent to more efficient firms.

²⁰Subscript i used to identify firms is dropped here since both firms are identical ex ante, hence receive identical regulatory contracts.

Rent is costly, hence the regulator optimally sets $U(\bar{\beta}) = 0$. Use this and (3.3) to obtain an expression for expected rent

$$\int_{\underline{\beta}}^{\bar{\beta}} U(\beta) dF(\beta) = \int_{\underline{\beta}}^{\bar{\beta}} \int_{\underline{\beta}}^{\bar{\beta}} \psi'(e(x)) dx dF(\beta) = \int_{\underline{\beta}}^{\bar{\beta}} \psi'(e(\beta)) \frac{F(\beta)}{f(\beta)} dF(\beta),$$

where the second equality follows from an integration by parts. Substitute this for expected rent in W to get expected welfare under independent regulation:

$$W^I = \int_{\underline{\beta}}^{\bar{\beta}} [V(s(\beta)) - (1 + \lambda)((\beta + s(\beta) - e(\beta))q(s(\beta)) + \psi(e(\beta))) - \lambda\psi'(e(\beta))\frac{F(\beta)}{f(\beta)}] dF(\beta).$$

Differentiate W^I with respect to $e(\beta)$, ignoring for the moment constraint (iii), and rewrite to obtain optimal effort level of firm i for given expenditures on quality improvement $s \in \{0, \bar{s}\}$:

$$\psi'(e^I(\beta, s)) = q(s) - \frac{\lambda}{1 + \lambda} \frac{F(\beta)}{f(\beta)} \psi''(e^I(\beta, s)) \quad \forall \beta \in [\underline{\beta}, \bar{\beta}]. \quad (3.4)$$

Equation (3.4) is the familiar one²¹ and shows that effort is distorted below the first best level [$\psi'(e(\beta, s)) = q(s)$] so as to reduce informational rent to more efficient types. More interesting for our purpose is the implication that optimal managerial effort is increasing in expenditures on quality improvement, everything else held equal.²² The reason is that an increase in quality leads to a boost in expected demand from $q(0)$ to $q(\bar{s})$ and thus to an increased scope for cost savings. There are returns to scale in effort. The more resources management devotes to increasing efficiency (the organization's fixed cost being $\psi(e)$) the lower is marginal cost $\beta + s - e$. As the firm grows bigger in terms of units sold, the higher are the returns to scale.

²¹See, for example, chapter 2 of Laffont and Tirole (1993).

²²The right-hand side of (3.4) non-increasing (by $\psi''' \geq 0$) and the left-hand side strictly increasing (by $\psi'' > 0$) in e imply effort strictly [weakly] increasing in s by $q(\bar{s}) > q(0)$ [$q(\bar{s}) \geq q(0)$].

Consider next optimal regulation of quality. The net benefit $\Delta W^I(\beta)$ in expected welfare of increasing quality of a firm of type β is given by

$$\begin{aligned} \Delta W^I(\beta) = & \{V(\bar{s}) - V(0)\} - (1 + \lambda)\{(\beta + \bar{s} - e^I(\beta, \bar{s}))q(\bar{s}) - (\beta - e^I(\beta, 0))q(0) \\ & + \psi(e^I(\beta, \bar{s}) - \psi(e^I(\beta, 0)))\} - \lambda\{\psi'(e^I(\beta, \bar{s})) - \psi'(e^I(\beta, 0))\}\frac{F(\beta)}{f(\beta)}. \end{aligned} \quad (3.5)$$

Increased spending on quality improvement benefits firms and consumers alike. Expected consumer surplus increases (the first term in curly brackets) owing to a boost in expected demand and to an improvement in expected quality. Firms get more informational rents owing to more efficient production. The social cost of increased informational rent is given by the last term. The effect on social production cost (the middle term) is ambiguous since marginal cost may increase or decrease following increased expenditures on quality improvement. The net welfare effect is thus ambiguous. To gain further insight, consider the effect on $\Delta W^I(\beta)$ of a marginal increase in productivity:

$$-\frac{d\Delta W^I(\beta)}{d\beta} = (1 + \lambda)\{q(\bar{s}) - q(0)\} + \lambda\{\psi'(e^I(\beta, \bar{s})) - \psi'(e^I(\beta, 0))\}\frac{d}{d\beta}\frac{F(\beta)}{f(\beta)}. \quad (3.6)$$

The expression is positive whenever elasticity of demand with respect to quality is positive [$q(\bar{s}) > q(0)$]. Thus, the social value of quality improvement is increasing in firm productivity. The reason is two-fold. First, marginal production cost is lower the higher is firm productivity. As productivity increases it becomes cheaper to supply the additional expected demand following increased spending on quality improvement. The effect on production cost is captured by the first term in the expression above. Second, the more productive is a firm the less compensation it requires to be willing to supply quality. This implies that the informational rent necessary to preserve truth-telling incentives under quality provision is decreasing in productivity. That effect is captured by the second term. Thus, optimal

regulation requires that stronger incentives for quality improvement be given to high productivity firms than to low productivity ones. Note also that the second term vanishes under perfect information. The net benefit of quality improvement is lower under asymmetric information than under perfect information. Thus, effort and quality are both distorted downward under asymmetric information. We collect the above results in a proposition:

Proposition 3.1. *Asymmetric information leads to downward distortions of effort and quality. The optimal contract under independent regulation has the following characteristics: (i) the more productive is a firm the more it spends on quality improvement ($\exists \beta^I \in [\underline{\beta}, \bar{\beta}] : s^I(\beta) = \bar{s} \forall \beta < \beta^I$ and $s^I(\beta) = 0 \forall \beta > \beta^I$); (ii) the more firms spend on quality improvement, the more effort is devoted to reducing cost ($e^I(\beta, \bar{s}) \geq e^I(\beta, 0)$, with effort implicitly given by eq. (3.4)).²³*

The analysis above suggests dividing firms into categories of "good" and "bad". Good firms are the high productivity ones. They produce goods of high average quality, and management works hard to keep costs down. Conversely, bad firms have low productivity, produce goods of low quality have a lot of "slack" in production.

Remark 1. *There is a positive correlation between measured quality and cost-efficiency. This gives us a statistical test to detect sub-optimal regulatory policies: if the correlation of quality and cost is negative and significant after controlling*

²³But what about constraint (iii) that $\beta - e^I(\beta)$ be non-decreasing? $e^I(\beta) = e^I(\beta, \bar{s})$ for all $\beta \in [\underline{\beta}, \beta^I]$ and $e^I(\beta) = e^I(\beta, 0)$ otherwise. $e^I(\beta, \bar{s})$ and $e^I(\beta, 0)$ are both weakly decreasing in β by the assumption of an increasing hazard rate and the properties of $\psi(\cdot)$, hence an additional condition necessary and sufficient to obtain global monotonicity is $e^I(\beta^I, \bar{s}) \geq e^I(\beta^I, 0)$. This follows directly from (ii) of the proposition.

for expenditures on quality improvement, the regulatory policy cannot be optimal because correlation should display the opposite sign.

3.2. Yardstick competition

Productivity is correlated across firms: they both have productivity in $[\underline{\beta}, a]$ if the common part is favourable ($m = \underline{m}$) or in $[a, \bar{\beta}]$ if the common part is unfavourable ($m = \bar{m}$). The two firms know this and so does the regulator. Yardstick competition allows the regulator to reduce informational rent: a transfer scheme that punishes incompatible productivity reports (e.g. $b_1 \in [\underline{\beta}, a]$ and $b_2 \in [a, \bar{\beta}]$) forces firms to deliver reports in identical intervals. The regulator ensures productivity reports in the true interval by rewarding (punishing) firms that perform relatively well (poor). Under such a contract high productivity firms produce efficiently (report $b_i \in [\underline{\beta}, a]$) so as to win the reward, whereas low-productivity firms find the bonus insufficient to cover the cost of having to work hard (hence report $b_i \in [a, \bar{\beta}]$). In particular, firms subjected to this type of yardstick competition cannot credibly understate productivity to $[a, \bar{\beta}]$ if their true type is in fact in $[\underline{\beta}, a]$. Since incompatible reports are not observed in equilibrium, the reduction in informational asymmetry is attained at no cost.

Expected rent, contingent on truth-telling, is $E[U_i(\boldsymbol{\beta})|\beta_i]$. Utilizing the envelope theorem and the fact that the distribution of β_{-i} is locally independent of β_i ,²⁴ yields $dE[U_i(\boldsymbol{\beta})|\beta_i] = -E[\psi'(e_i(\boldsymbol{\beta}))|\beta_i]d\beta_i$. Integrating over β_i gives an expression for expected rent under TIC

$$E[U_i(\boldsymbol{\beta})|\beta_i] = \begin{cases} E[\int_{\underline{\beta}_i}^a \psi'(e_i(x, \beta_{-i}))dx|\beta_i] + E[U_i(a, \beta_{-i})|a] & \forall \beta_i \in [\underline{\beta}, a] \\ E[\int_{\beta_i}^{\bar{\beta}} \psi'(e_i(x, \beta_{-i}))dx|\beta_i] + E[U_i(\bar{\beta}, \beta_{-i})|\bar{\beta}] & \forall \beta_i \in (a, \bar{\beta}] \end{cases} \quad (3.7)$$

²⁴The density function of β_{-i} conditional on $\beta_i = \alpha m + (1-\alpha)\varepsilon_i$ is given by $g((\beta_{-i} - \alpha m)/(1-\alpha)) / (1-\alpha)$ for all $\beta_{-i} \in [\alpha m + (1-\alpha)\underline{\varepsilon}, \alpha m + (1-\alpha)\bar{\varepsilon}]$ and zero otherwise.

Further, $E[U_i(a, \beta_{-i})|a] \geq 0$ and $E[U_i(\bar{\beta}, \beta_{-i})|\bar{\beta}] \geq 0$ follow directly from IR. Assuming these necessary conditions to be sufficient (which they are shown in appendix A.2 to be), one obtains expected rent

$$E[U_i(\boldsymbol{\beta})|\beta_i] = \int_{\underline{\beta}}^{\bar{\beta}} \int_{\underline{\beta}}^{\bar{\beta}} \psi'(e_i(\boldsymbol{\beta})) \frac{F(\beta_i) - I(\beta_i)F(a)}{f(\beta_i)} d\hat{F}(\boldsymbol{\beta})$$

after setting $E[U_i(a, \beta_{-i})|a] = 0$ and $E[U_i(\bar{\beta}, \beta_{-i})|\bar{\beta}] = 0$ and performing an integration by parts. $I(\beta)$ is an indicator function, unity for all $\beta \in (a, \bar{\beta}]$ and zero otherwise. Plug the expression for expected firm rent into the welfare function (3.2) to obtain

$$W^{yc} = \int_{\underline{\beta}}^{\bar{\beta}} \int_{\underline{\beta}}^{\bar{\beta}} [V(\mathbf{s}(\boldsymbol{\beta})) - (1 + \lambda) \sum_i ((\beta_i + s_i(\boldsymbol{\beta}) - e_i(\boldsymbol{\beta})) q_i(\mathbf{s}(\boldsymbol{\beta})) + \psi(e_i(\boldsymbol{\beta}))) - \lambda \sum_i \psi'(e_i(\boldsymbol{\beta})) \frac{F(\beta_i) - I(\beta_i)F(a)}{f(\beta_i)}] d\hat{F}(\boldsymbol{\beta}) \quad (3.8)$$

Differentiate W^{yc} to get optimal effort under yardstick competition for given expenditures \mathbf{s} on quality improvement :

$$\psi'(e_i^{yc}(\beta_i, \mathbf{s})) = q_i(\mathbf{s}) - \frac{\lambda}{1 + \lambda} \frac{F(\beta_i) - I(\beta_i)F(a)}{f(\beta_i)} \psi''(e_i^{yc}(\beta_i, \mathbf{s})) \quad \forall \beta_i \in [\underline{\beta}, \bar{\beta}]. \quad (3.9)$$

3.2.1. Regional monopolies

When firms operate in independent markets, expected demand $q_i(\mathbf{s})$ for a firm's goods depends entirely on its own service s_i and is independent of the expenditures s_{-i} on quality improvement undertaken by the other firm, i.e. $q_i(\mathbf{s}) = q(s_i)$. The effect of yardstick competition on managerial incentives can in this case be observed by comparing cost-reducing effort (3.9) under yardstick competition with cost-reducing effort (3.4) under independent regulation. Nothing changes for the high-productivity types $\beta \in [\underline{\beta}, a]$ whereas low-productivity types $\beta \in (a, \bar{\beta}]$ produce more efficiently (effort is less distorted) under yardstick competition than

under independent regulation.²⁵ Yardstick competition thus leads to more high-powered incentives, everything else held equal. How can this be explained? Effort is distorted below first-best in order to extract rents from high productivity firms. Under independent regulation rent must be given up to firms with high industry-specific productivity ($\beta \in [\underline{\beta}, a]$) so as to prevent them from understating productivity to $[a, \overline{\beta}]$, which is unnecessary under yardstick competition. Total and marginal rent extraction is larger for low-productivity types $\beta \in [a, \overline{\beta}]$ under independent regulation than under yardstick competition. This calls for larger distortions of effort in the former than the latter case. Yardstick competition leaves optimal regulation of high-productivity types unaffected since more efficient types to the same extent as before have to be compensated for not understating productivity.

In the previous section, the optimal provision of quality was found to be increasing in productivity, one of the reasons being a decreasing amount of informational rent necessary to induce supply of quality. As we have just seen, yardstick competition reduces informational rent. Combining these two results, one would expect yardstick competition to lead to a higher optimal supply of quality. Presumably, the rent-reduction achieved by yardstick competition would be optimally spent on quality improvement as well as production efficiency. The analysis below establishes that this intuition is correct: the net benefit of expenditures on quality is (at least weakly) higher under yardstick competition than under independent regulation, leading to more provision of quality under the former than the latter regulatory regime.

²⁵ Auriol and Laffont (1992) were the first to observe that yardstick competition does not affect incentives in high-productivity firms and leads to less distortion in low-productivity firms under the imposed stochastic structure. See also Auriol (1993) and Dalen (1998).

The socially optimal level of quality is based on the same trade-off as in the previous section; increased consumer surplus is balanced against increased informational rent, with the net effect on production cost being either positive or negative:

$$\begin{aligned} \Delta W^{yc}(\beta) = & \{V(\bar{s}) - V(0)\} - (1 + \lambda)\{(\beta + \bar{s} - e^{yc}(\beta, \bar{s}))q(\bar{s}) - (\beta - e^{yc}(\beta, 0))q(0) \\ & + \psi(e^{yc}(\beta, \bar{s}) - \psi(e^{yc}(\beta, 0)))\} - \lambda\{\psi'(e^{yc}(\beta, \bar{s})) - \psi'(e^{yc}(\beta, 0))\} \frac{F(\beta) - I(\beta)F(a)}{f(\beta)}. \end{aligned} \quad (3.10)$$

The fundamental difference between this expression and eq. (3.5) defining the optimal supply of quality under independent regulation, is the rent-extraction effect of yardstick competition. Increased rent-extraction reduces the social cost of supplying quality (the proof is in appendix A.3):

Proposition 3.2. *In the regional monopoly case, the social value of quality improvement is higher under yardstick competition than under independent regulation ($\Delta W^{yc}(\beta) \geq \Delta W^I(\beta) \forall \beta \in [\underline{\beta}, \bar{\beta}]$).*

The implication is that more should be spent on quality improving measures under yardstick competition than under independent regulation. To see this, consider the effect on types that display low IS-productivity, but high FS-productivity, i.e. β close to, but above a . Under independent regulation the marginal production cost combined with the informational rent necessary to induce truth-telling may render quality improvement for these types too costly. Under yardstick competition, on the other hand, informational rent is relatively low since the common part is filtered out. The effect may be sufficiently strong to induce expenditures on quality despite the high marginal production cost. To summarize the results:

Proposition 3.3. *At the social optimum with regional monopoly, firms produce more efficiently and spend more on quality improvement under yardstick competition than under independent regulation.*

3.2.2. Competition

The analysis so far has concerned itself with regional monopolies. This is an appropriate description of the relationship between elementary schools responsible for education in geographically separate areas or between hospitals serving patients in segmented markets. In other instances it is more appropriate to consider competition. Universities compete on a national level for students. Hospitals in urban areas compete for patients. This section considers regulation when firms are in direct competition with one another. Under this market structure, optimal spending on cost-reducing effort is still given by (3.9). The new thing to consider is quality. To simplify the analysis, attention is restricted to the case with inelastic total demand, i.e. $q_1(\mathbf{s}) + q_2(\mathbf{s}) = Q$. Should two competing hospitals spend equally much on quality improvement, and if not, who should spend more/less? Intuitively, the answer depends on their relative productivity. Quality provision is cheaper the more productive is a hospital. Hence, the more productive firm in the region should be given stronger incentives for quality improvement than the less productive one. This means that competing firms that are sufficiently different in terms of productivity should produce goods of different quality (the proof is in appendix A.4):

Proposition 3.4. *Expenditures on quality improvement under yardstick competition and product-market competition are increasing in the firm's own productivity and decreasing in the competitor's productivity.*

Proposition 3.4 along with eq. (3.9) emphasise the necessity of sometimes treating firms unequally even if they are operating in the same market. When productivity differs significantly across firms, some producers (the high performers) should be given stronger incentives for cost-reduction and quality improvement than others (the low performers). Such a regulatory policy is probably non-controversial when it comes to cost-efficiency. The implication for quality provision is less appealing. For the health sector, for example, the optimal policy encourages variation in the quality of care across hospitals as a function of their relative productivity. As a norm, a patient admissiomed to a high productivity hospital should receive better medical treatment than a patient hospitalized in a less productive one. Two people suffering an accident in the same city might expect to receive unequal treatment depending on the hospital to which they are admitted. To be sure, this is the reality in many metropolitan areas. Some hospitals are considered better than others. The new result here is that this situation may in fact be desirable from a welfare point of view.

Since inequality of hospitals may be in conflict with a principle of equality of treatment and thus politically controversial, it is well worth recapitulating why differentiating hospitals with respect to quality may be optimal. As a general rule, patients should be treated at the cheapest, i.e. most productive, hospital, all other things held equal (see, e.g. Auriol and Laffont, 1992). Hence, there is a motive for patient-shifting from inefficient to efficient hospitals within a metropolitan area. Since patients are free to choose hospital (at least in this model they are) patient-shifting is achieved by encouraging investment in quality improvement in high-productivity hospitals so as to attract patients to them. The subsequent increase in production cost is partially off-set by stronger demands on management

to keep costs down. If in addition there are returns to scale in investments on quality, e.g. expenditures on quality improvement are fixed, an additional effect plays in. Holding fixed the amount of patients admitted to each hospital at the new levels, the value of investing in quality is higher in the larger hospital since there are more patients treated there and the impact on consumer surplus higher.

4. Implementation

This paper has shown that quality and cost concerns are inherently intertwined. Under optimal regulation firms are polarized according to their relative productivity. High-productivity firms should produce high-quality goods and management work hard to keep costs down. At the other end of the scale, low-productivity firms should produce lesser-quality goods and be allowed more slack in production so as to increase rent-extraction. The majority of the market would thus be served by the most efficient (in terms of productivity) firms.

Optimal regulation puts strong requirements on data availability, such as detailed productivity reports and reliable quality indicators - data to which the regulatory body seldom has access. Under the US Prospective Payment System (PPS), for example, the authorities solely collect and use cost data from hospitals. Thus the optimal scheme is difficult to implement in practice. Practical implementation requires simpler mechanisms. In light of the insights provided above, this section addresses two questions. First, how do we expect a yardstick competition scheme like the PPS to perform in terms of providing incentives for quality improvement? Second, if incentives are believed to be insufficient, how can the simple mechanism be augmented so as to take quality considerations explicitly

into account?

Under the PPS and for each diagnostic group, hospital i is paid a transfer $t\bar{c}_{-i} > 0$ per admission which is a fraction of the average cost \bar{c}_{-i} across the other hospitals of performing operations within that specific diagnostic group.²⁶ Write $E[\bar{c}_{-i}(\boldsymbol{\beta}_{-i})|\beta_i]$ i 's expectation over \bar{c}_{-i} , contingent on productivity β_i . Suppose now that the transfer structure is augmented by the following bonus structure. In addition to the fixed payment, hospital i receives a bonus $b^h(\bar{c}_{-i} - c_i)$ if it produced services of high quality that period ($\sigma_i = \sigma^h$) and $b^l(\bar{c}_{-i} - c_i)$ otherwise ($\sigma_i = \sigma^l$), with $b^h \geq b^l$. Let $E[q_i^w(\boldsymbol{\beta}_{-i})|\beta_i]$ be expected demand facing firm i given $\sigma_i = \sigma^w$ ($w \in \{l, h\}$), and assume $E[q_i^h(\boldsymbol{\beta}_{-i})|\beta_i] \geq E[q_i^l(\boldsymbol{\beta}_{-i})|\beta_i]$. The inequality is strict if demand is elastic with respect to quality. The assumption of constant marginal cost given by $c_i = \beta_i + s_i - e_i$ is maintained throughout. Under this regulatory scheme expected firm rent is

$$E[U_i(s_i, e_i, \boldsymbol{\beta})|\beta_i] = (1 - \theta(s_i))\{(t + b^l)E[\bar{c}_{-i}(\boldsymbol{\beta}_{-i})|\beta_i] - (1 + b^l)c_i\}E[q_i^l(\boldsymbol{\beta}_{-i})|\beta_i] \\ + \theta(s_i)\{(t + b^h)E[\bar{c}_{-i}(\boldsymbol{\beta}_{-i})|\beta_i] - (1 + b^h)c_i\}E[q_i^h(\boldsymbol{\beta}_{-i})|\beta_i] - \psi(e_i).$$

The FOC for managerial effort is:

$$\theta(s_i)(1 + b^h)E[q_i^h(\boldsymbol{\beta}_{-i})|\beta_i] + (1 - \theta(s_i))(1 + b^l)E[q_i^l(\boldsymbol{\beta}_{-i})|\beta_i] = \psi'(e_i).$$

It is straightforward to verify that managerial effort is increasing in expenditures in quality improvement - a necessary condition for optimality of the regulatory contract. The PPS provides, at least in theory, good incentives for cost containment. What about quality considerations? By utilising the envelope theorem and the assumption that the distribution of $\boldsymbol{\beta}_{-i}$ is locally independent of β_i , one finds:

$$-\frac{\delta}{\delta\beta_i}[E[U_i(\bar{s}, e_i(\bar{s}), \boldsymbol{\beta})|\beta_i] - E[U_i(0, e_i(0), \boldsymbol{\beta})|\beta_i]] = \\ \frac{[\theta(\bar{s}) - \theta(0)][(1 + b^h)E[q_i^h(\boldsymbol{\beta}_{-i})|\beta_i] - (1 + b^l)E[q_i^l(\boldsymbol{\beta}_{-i})|\beta_i]]}{\theta(\bar{s}) - \theta(0)} \geq 0.$$

²⁶In reality the average is over all hospitals, including hospital i . To simplify the analysis, I ignore this marginal effect in the subsequent analysis.

Spending on quality improvement is increasing in productivity, exactly as required of an optimal mechanism. This leads, in turn, to stronger focus on cost savings.

The proposed mechanism displays the qualitative feature required of an optimal scheme, namely stronger focus on cost control and quality provision the more efficient is the hospital. Observe that this is prevalent even under the current PPS system ($b^h = b^l = 0$) provided demand is elastic with respect to quality. Thus, reinforcing the sales incentive by increasing demand elasticity probably is a useful instrument in mitigating quality problems. One obvious thing to do would be to publish systematic quality comparisons between hospitals so as to increase transparency and patient awareness. Further, the authorities also could subsidize ambulance services so as to reduce patient transportation cost between hospitals and thereby increase competition for patients in geographically disperse areas.

Sometimes the regulatory authorities cannot do much in terms of affecting elasticity of demand. Demand for some services offered by regional hospitals, such as mending broken legs and delivering babies, is inelastic. Under the current PPS system hospitals most likely under-invest in the quality of these services since income is independent of quality. However, this could be corrected by utilising the proposed bonus system and setting $b^h > b^l$. It is easy to verify that the larger is the difference between b^h and b^l , the more will hospitals spend on quality improvement.²⁷ How could this be achieved in practice? Within the UK education sector, for example, transfers to universities depend on the measured quality of their research and education programs relative to that of other universities during the previous period. This could be applied to the health sector too, given that reliable quality data could be collected.

²⁷As is readily apparent, $b^h - b^l$ can be used to boost overall expenditures on quality improvement even when demand is elastic so as to correct for underinvestment.

Under the proposed yardstick competition scheme, transfers are explicitly connected to relative cost as the quality of output. To my knowledge, these have never been combined in practice, but there seems to be no conceptual difficulty involved in doing so. The greatest obstacle to implementation appears to be finding good objective quality measures and cost data.

5. A first look at the data

US hospitals have been regulated by means of yardstick competition, the PPS, for quite some time now. The analysis above predicts the PPS to have had an impact on quality and cost containment across hospitals even in its current form. Specifically, it predicts the separation of high-productivity and low-productivity hospitals into high-quality, efficient suppliers versus lesser quality, less efficient suppliers. Can we find any trace of separation in the data? Unfortunately, no systematic quality comparisons across hospitals and over time have been made. One has to look for indirect evidence.

For most treatments there are probably minimum requirements in terms of investments needed for hospitals to perform treatments of acceptable quality. Under the old cost-reimbursement system, it was relatively inexpensive even for inefficient hospitals to supply costly treatments since the public paid the bill anyhow. Under the PPS however, pressure is put on hospitals to reduce costs. Only a few, efficient hospitals now have the size and competence to offer adequate services at competitive costs for some expensive treatments. An expected outcome of a yardstick competition scheme such as the PPS would thus be an increased degree of specialization. In metropolitan areas one might expect to find a few large hos-

pitals (in terms of patient numbers) supplying health services for which expensive, specialized equipment is required in order to guarantee quality of treatment.

Each year the authorities collect data on the types of cases treated by the individual hospitals, categorized by the *Diagnostic Related Groups* (DRGs) to which the cases belong. A hospital that has had a high fraction of its patients in costly DRGs receive a high score (so-called *Case-Mix*) and conversely for a hospital that has provided inexpensive treatments to its patients. A national Case-Mix average is computed and each hospital is ranked relative to that average. This is the *Case-Mix Index* (CMI), which is a measure of the costliness of cases treated by a hospital relative to the cost of the national average of all Medicare hospital cases. The CMI forms the basis on which hospitals are reimbursed, and is the data I study the effect of the PPS on the degree of specialization among hospitals.

First, I studied the distribution of the CMI across acute hospitals for each state for the two years 1988 and 2000. For each state and year the average CMI across hospitals was calculated and the fraction of hospitals that had a CMI above the state average computed and used as a measure of the degree of specialization. A reduction in this number from 1988 to 2000 reveals a decrease in the relative amount of hospitals offering the costliest treatments within that state and is interpreted as an increase in specialization.²⁸ An increase indicates reduced specialization in expensive treatments.²⁹ Table 5.1 below shows the average fraction

²⁸In California, for example, 37% of the hospitals offered more expensive treatments than the state average in 1988. This number had reduced to 8% by 2000. Specialization in expensive treatments thus increased in California from 1988 to 2000 by my definition.

²⁹In Colorado, 15% of the hospitals offered more expensive treatments than the state average in 1988. This number had increased to 28% by 2000. Specialization in expensive treatments thus decreased in Colorado from 1988 to 2000.

	1988	2000
Average fraction	0,32	0,19
Sample variance	0,037	0,017

Table 5.1: Specialization across states

	1988	2000
Average Fraction	0,31	0,34
Sample variance	0,024	0,035

Table 5.2: Specialization across cities

and sample variation across states.³⁰

In 1988, an average of 32% of the US acute hospitals offered treatments that were more costly than the state average. This figure had reduced to 19% by 2000. The decrease is significant on the 1 % level.³¹ Hence, there seems to have been *increased* specialization on the state level from 1988 to 2000.

Second, I considered the distribution of the CMI across acute hospitals in the 50 largest US cities³² for the same years and calculated for each city and year the fraction of hospitals that had a CMI above the city average. Table 5.2 below shows the average and variation across cities:

In the metropolitan areas the picture is different. If anything, specialization appears to have *decreased* from 1988 to 2000, although the effect fails to prove

³⁰Data for 1988 was supplied by the Health Care Financing Administration (HCFA). Data for 2000 was downloaded from the HCFA website <http://www.hcfa.gov/stats/pufiles.htm> on 9. January 2002.

³¹The difference is 0,13 [= 0,32 − 0,19] and the standard error (assuming independence across years) 0,033 [= $\sqrt{(0,037 + 0,017)/52}$], which gives a t-statistic 4,24 [= 0,14/0,033] significant on the 1% level.

³²A city was exclusively defined by the three first digits in the ZIP code. Only those hospitals with ZIP code 100XX were considered New York hospitals. Hospitals located in the Bronx (ZIP code 104XX), for example, were excluded. On the other hand, Baltimore (ZIP code 212XX) includes hospitals located in Towson because Baltimore and Towson share the first three ZIP code digits. For a list of the 50 largest US cities, see appendix A.5.

significant.³³ Note that the degree of specialization in cities and across states was roughly the same in 1988, indicating that the largest cities and the less populated areas were roughly similar in terms of specialization at that time. This had changed by 2000, when the fraction of hospitals treating the most costly cases had become much larger in the big cities than on state level. It is tempting to conclude that metropolitan hospitals took over expensive, specialized treatment from rural hospitals during the period.

The above results should be interpreted with caution. A change in specialization from one year to another does not by itself constitute a *trend*. Ideally, one would like to consider changes over a range of time, including the period prior to the introduction of the PPS in 1983.³⁴ Further, changes in the composition of hospitals, should they appear to be a trend, cannot necessarily be attributed to the introduction of the PPS. Other variables should be considered too, such as the general economic climate and urbanization, which probably affect the demand for as well as supply of health services. The intention has been to show that *something* has appeared to have happened following the introduction of the PPS. A detailed analysis is left for future research.

6. Conclusion

This paper has studied simultaneous regulation of cost and quality when the regulator has access to yardstick competition and quality is unverifiable. The crucial assumptions were unregulated quantity and demand governed entirely by

³³The difference is 0,03 [= 0,34 - 0,31] and the standard error (assuming independence across years) 0,034 [= $\sqrt{(0,024 + 0,035)/50}$], which gives a t-statistic 0,88 [= 0,03/0,034].

³⁴This being difficult, owing to data availability. The assumption maintained here is that changes have occurred with a lag.

anticipated quality, typically the case in the health sector.

A general principle has emerged. Managerial effort and expenditures on quality should be positively correlated: the higher is firm productivity the more efficiently should the firm produce and the more should be spent on quality improvement. Yardstick competition reduces informational rent, which should be used to increase production efficiency as well as quality. The effect of adding product market competition is to connect competitors' performance tighter to one another. Expenditures on quality and managerial effort should be increasing in the firm's own productivity and decreasing in the competitor's productivity.

A few policy implications can be drawn. Yardstick competition offers a cheap possibility for efficiency as well as quality improvements, given that quality indicators and cost-reimbursement rules are carefully constructed and designed. Furthermore, the analysis delivers a statistical test to detect sub-optimal regulatory policies in the kind of industry studied (e.g. the health sector): if the correlation of quality and cost is negative and significant after controlling for expenditures on quality improvement, the regulatory policy is sub-optimal.

Implementing the socially optimal regulatory policy will most likely be politically challenging, for two reasons. First, consumers are better off and firms obtain more informational rent, the higher is the quality of the services offered. On the basis of this one would expect patient organizations to join forces with hospitals to lobby for maximum quality health care. The loser in the game would be the anonymous taxpayer financing an over-expensive, over-quality health service. Second, the analysis stresses the need for distinguishing between high-productivity and low-productivity producers under yardstick competition and product market competition. The implication for the health industry is that average quality of

treatment be allowed to vary across hospitals as a function of their relative productivity. A policy encouraging differences in treatment would most likely meet opposition and perhaps be politically infeasible. To this should be added that productivity is endogenous. Whether yardstick competition and product market competition lead to convergence or divergence in productivity and thus in performance in the long run, remains to be seen.³⁵

A. Appendix

A.1. Expenditures on quality improvement

Suppose firms have reported \mathbf{b} and that the regulator wants to implement $\{s_i(\mathbf{b})\}_{i=1,2}$ if \mathbf{b} is in fact the true type. This appendix shows that the regulator can costlessly induce any preferred $\{s_i(\mathbf{b})\}_{i=1,2}$ irrespective of beliefs and of the true type β . Let

$$\hat{q}_i^w(\mathbf{b}) = \hat{\mu}_i(\mathbf{b})(\theta(\bar{s})q^{wh} + (1 - \theta(\bar{s}))q^{wl}) + (1 - \hat{\mu}_i(\mathbf{b}))(\theta(0)q^{wh} + (1 - \theta(0))q^{wl})$$

be expected demand for firm i 's goods contingent on $\sigma_i = \sigma^w$ ($w \in \{l, h\}$), on the report \mathbf{b} and on i 's subjective belief that $s_{-i} = \bar{s}$ with probability $\hat{\mu}_i(\mathbf{b})$ and zero otherwise. Expected rent is

$$\hat{U}_i(s_i, \mathbf{b}, \beta_i) = \theta(s_i)(t_i^h(\mathbf{b}) - c_i(\mathbf{b}))\hat{q}_i^h(\mathbf{b}) + (1 - \theta(s_i))(t_i^l(\mathbf{b}) - c_i^l(\mathbf{b}))\hat{q}_i^l(\mathbf{b}) - \psi(\beta_i - c_i(\mathbf{b}) + s_i).$$

Define the vector of transfers

$$t_i^h(\mathbf{b}) = c_i(\mathbf{b}) + \frac{\theta(s_i(\mathbf{b}))R_i(b_i) + \delta_i(\mathbf{b})(1 - \theta(s_i(\mathbf{b})))}{\theta(s_i(\mathbf{b}))(\theta(s_{-i}(\mathbf{b}))q^{hh} + (1 - \theta(s_{-i}(\mathbf{b})))q^{hl})}$$

³⁵See Dalen (1998) for an analysis of the effect of yardstick competition on investment incentives in a market without quality considerations.

$$t_i^l(\mathbf{b}) = c_i(\mathbf{b}) + \frac{R_i(b_i) - \delta_i(\mathbf{b})}{\theta(s_{-i}(\mathbf{b}))q^{lh} + (1 - \theta(s_{-i}(\mathbf{b})))q^{ll}}.$$

for some $R_i(b_i)$ and with $\delta_i(\mathbf{b})$ yet to be defined. Under the proposed transfer structure, i 's incentive for supplying quality $\Delta \hat{U}_i(\mathbf{b}, \beta_i)$ is given by

$$\begin{aligned} \Delta \hat{U}_i(\mathbf{b}, \beta_i) = & (\theta(\bar{s}) - \theta(0)) \left\{ \frac{\theta(s_i(\mathbf{b}))R_i(b_i)\hat{q}_i^h(\mathbf{b}) + \delta_i(\mathbf{b})(1 - \theta(s_i(\mathbf{b})))\hat{q}_i^h(\mathbf{b})}{\theta(s_i(\mathbf{b}))(\theta(s_{-i}(\mathbf{b}))q^{hh} + (1 - \theta(s_{-i}(\mathbf{b})))q^{hl})} \right. \\ & \left. + \frac{\delta_i(\mathbf{b})\hat{q}_i^l(\mathbf{b}) - R_i(b_i)\hat{q}_i^l(\mathbf{b})}{\theta(s_{-i}(\mathbf{b}))q^{lh} + (1 - \theta(s_{-i}(\mathbf{b})))q^{ll}} \right\} - \psi(\beta_i - c_i(\mathbf{b}) + \bar{s}) + \psi(\beta_i - c_i(\mathbf{b})) \end{aligned}$$

The RHS of the equation is strictly increasing in $\delta_i(\mathbf{b})$. Thus, the regulator can implement $s_i(\mathbf{b}) = \bar{s}$ [$s_i(\mathbf{b}) = 0$] independently of the subjective beliefs $\hat{\mu}_i(\mathbf{b})$ and of the true type β_i by selecting $\delta_i(\mathbf{b})$ sufficiently high [low]. Hence, all types choose $\{s_i(\mathbf{b})\}_{i=1,2}$, given the previous report \mathbf{b} . In equilibrium, it is required that beliefs be consistent: firm i is not allowed to hold beliefs about firm $-i$'s actions that are incompatible with utility maximizing behaviour of firm $-i$. Under the appropriate choice of $\delta_1(\mathbf{b})$ and $\delta_2(\mathbf{b})$ the unique set of consistent beliefs $\{\mu_1(\mathbf{b}), \mu_2(\mathbf{b})\}$ is $\mu_i(\mathbf{b}) = 1$ for $s_{-i}(\mathbf{b}) = \bar{s}$ and $\mu_i(\mathbf{b}) = 0$ for $s_{-i}(\mathbf{b}) = 0$. It is straightforward to verify that expected rent is

$$U_i(\mathbf{b}, \beta_i) = R_i(b_i) - \psi(\beta_i - c_i(\mathbf{b}) + s_i(\mathbf{b})) \quad (\text{A.1})$$

under consistent beliefs, which is independent of $\delta_i(\mathbf{b})$. Thus, the regulator can freely choose $\delta_1(\mathbf{b})$ and $\delta_2(\mathbf{b})$ so as to obtain QIC. ■

A.2. Global truth-telling under yardstick competition

Since the regulator punishes firms for delivering incompatible productivity reports, a firm expecting the other one to truthfully reveal its type, will choose to report its type in the truthful interval. Under the proposed yardstick competition contract, expected profits for the type b_i is

$$E[\pi_i(b_i, \beta_{-i})|b_i] = E\left[\int_{b_i}^{\bar{\beta}(b_i)} \psi'(e_i(x, \beta_{-i}))dx + \psi(e_i(b_i, \beta_{-i})|b_i]\right] \quad (\text{A.2})$$

under truthful revelation of types, with $\bar{\beta}(b_i) = a \ \forall b_i \in [\underline{\beta}, a]$ and $\bar{\beta}$ otherwise.

Expected rent of reporting b_i in the truthful interval is given by

$$\begin{aligned} E[U_i(b_i, \beta_{-i})|\beta_i] &= E[\pi_i(b_i, \beta_{-i})|\beta_i] - E[\psi(\beta_i - b_i + e_i(b_i, \beta_{-i}))|\beta_i] \\ &= E[\pi_i(b_i, \beta_{-i})|b_i] - E[\psi(\beta_i - b_i + e_i(b_i, \beta_{-i}))|\beta_i] \\ &= E[\int_{b_i}^{\bar{\beta}(b_i)} \psi'(e_i(x, \beta_{-i}))dx + \psi(e_i(b_i, \beta_{-i})|b_i] - E[\psi(\beta_i - b_i + e_i(b_i, \beta_{-i}))|\beta_i] \\ &= E[\int_{b_i}^{\bar{\beta}(\beta_i)} \psi'(e_i(x, \beta_{-i}))dx + \psi(e_i(b_i, \beta_{-i})|\beta_i] - E[\psi(\beta_i - b_i + e_i(b_i, \beta_{-i}))|\beta_i] \end{aligned}$$

when the true type is in fact β_i . Equalities two and four stem from the fact that the types b_i and β_i contain identical information about the distribution of β_{-i} , and the third equality comes from substituting the RHS of (A.2) for $E[\pi_i(b_i, \beta_{-i})|b_i]$ into the expression. Collecting terms and simplifying yield the net benefit

$$E[U_i(b_i, \beta_{-i})|\beta_i] - E[U_i(\beta)|\beta_i] = E[\int_{b_i}^{\beta_i} \int_{x-e_i(x, \beta_{-i})}^{b_i-e_i(b_i, \beta_{-i})} \psi''(x-y)dydx|\beta_i]$$

of misrepresenting one's type. A sufficient (although not necessary) condition for this expression to be non-positive is that $\beta_i - e_i(\beta)$ be non-decreasing in β_i . Let $s_1^{yc}(\beta)$ and $s_2^{yc}(\beta)$ be the socially optimal levels of \mathbf{s} . By definition $e_i^{yc}(\beta) = e_i^{yc}(\beta_i, s_1^{yc}(\beta), s_2^{yc}(\beta))$. From the monotone hazard rate assumption and the properties of ψ , we know that $e_i^{yc}(\beta_i, \mathbf{s})$ implicitly defined in (3.9) is weakly decreasing in β_i , increasing in s_i and decreasing in s_{-i} . By proposition 3.4, $s_i^{yc}(\beta)$ is non-increasing and $s_{-i}^{yc}(\beta)$ non-decreasing in β_i , which completes the proof that $\beta_i - e_i^{yc}(\beta)$ is non-decreasing in β_i for the product-market competition case. In the regional monopoly case $e_i^{yc}(\beta_i, \mathbf{s})$ is independent of s_{-i} and $s_i^{yc}(\beta_i)$ is non-increasing in β_i , which can be seen from differentiating (3.10) wrt β_i . This implies $\beta_i - e_i^{yc}(\beta)$ non-decreasing in β_i also in the regional monopoly case. ■

A.3. Proof of proposition 3.2

Define a pseudo hazard rate $H(\beta, x) = (F(\beta) - x)/f(\beta)$, define implicitly $e(\beta, s, x)$ by

$$\psi'(e(\beta, s, x)) = q(s) - \frac{\lambda}{1 + \lambda} H(\beta, x) \psi''(e(\beta, s, x)) \quad (\text{A.3})$$

and $\Delta W(\beta, x)$ by

$$\Delta W(\beta, x) = \{V(\bar{s}) - V(0)\} - (1 + \lambda) \{(\beta + \bar{s} - e(\beta, \bar{s}, x))q(\bar{s}) - (\beta - e(\beta, 0, x))q(0) + \psi(e(\beta, \bar{s}, x) - \psi(e(\beta, 0, x)))\} - \lambda \{\psi'(e(\beta, \bar{s}, x)) - \psi'(e(\beta, 0, x))\} H(\beta, x).$$

By definition, $e(\beta, s, 0) = e^I(\beta, s)$, $e(\beta, s, I(\beta)F(a)) = e^{yc}(\beta, s)$, (i) $\Delta W(\beta, 0) = \Delta W^I(\beta)$ and (ii) $\Delta W(\beta, I(\beta)F(a)) = \Delta W^{yc}(\beta)$. Differentiate and use (A.3) to get:

$$\frac{\partial}{\partial x} \Delta W(\beta, x) = \lambda \frac{\psi'(e(\beta, \bar{s}, x)) - \psi'(e(\beta, 0, x))}{f(\beta)} \geq 0 \forall x \leq F(\beta)$$

The inequality follows from $e(\beta, \bar{s}, x) \geq e(\beta, 0, x) \forall x \leq F(\beta)$ and $\psi'' > 0$. Integrating over x and using (i) and (ii) yields

$$\Delta W^{yc}(\beta) - \Delta W^I(\beta) = \int_0^{I(\beta)F(a)} \lambda \frac{\psi'(e(\beta, \bar{s}, x)) - \psi'(e(\beta, 0, x))}{f(\beta)} dx.$$

$\Delta W^{yc}(\beta) = \Delta W^I(\beta) \forall \beta \in [\underline{\beta}, a]$ because $I(\beta) = 0 \forall \beta \in [\underline{\beta}, a]$ and $\Delta W^{yc}(\beta) \geq \Delta W^I(\beta) \forall \beta \in (a, \bar{\beta}]$ because $I(\beta) = 1 \forall \beta \in (a, \bar{\beta}]$ and $x \leq F(a) < F(\beta)$. ■

A.4. Proof of proposition 3.4

Let $W^{yc}(\beta, \mathbf{s})$ be welfare under truthful revelation of β , when \mathbf{s} is spent on quality improvement and when managerial effort $e_i^{yc}(\beta_i, \mathbf{s})$ is implicitly given by (3.9):

$$W^{yc}(\beta, \mathbf{s}) = V(\mathbf{s}) - (1 + \lambda) \sum_i ((\beta_i + s_i - e_i^{yc}(\beta_i, \mathbf{s}))q_i(\mathbf{s}) + \psi(e_i^{yc}(\beta_i, \mathbf{s}))) - \lambda \sum_i \psi'(e_i^{yc}(\beta_i, \mathbf{s})) \frac{F(\beta_i) - I(\beta_i)F(a)}{f(\beta_i)}.$$

Differentiate wrt β_i , using the envelope theorem

$$\frac{\partial W^{yc}(\boldsymbol{\beta}, \mathbf{s})}{\partial \beta_i} = -(1 + \lambda)q_i(\mathbf{s}) - \lambda \psi'(e_i^{yc}(\beta_i, \mathbf{s})) \frac{\partial}{\partial \beta_i} \frac{F(\beta_i) - I(\beta_i)F(a)}{f(\beta_i)}. \quad (\text{A.4})$$

By assumption $q_i(\bar{s}, 0) \geq q_i(\bar{s}, \bar{s}) = q_i(0, 0) \geq q_i(0, \bar{s})$. From (3.9), this implies $e_i^{yc}(\beta_i, \bar{s}, 0) \geq e_i^{yc}(\beta_i, \bar{s}, \bar{s}) = e_i^{yc}(\beta_i, 0, 0) \geq e_i^{yc}(\beta_i, 0, \bar{s})$. Thus, both terms on the RHS of (A.4) are non-increasing in s_i and non-decreasing in s_{-i} . Hence,

$$\frac{\partial W^{yc}(\boldsymbol{\beta}, \bar{s}, 0)}{\partial \beta_i} \leq \frac{\partial W^{yc}(\boldsymbol{\beta}, \bar{s}, \bar{s})}{\partial \beta_i} = \frac{\partial W^{yc}(\boldsymbol{\beta}, 0, 0)}{\partial \beta_i} \leq \frac{\partial W^{yc}(\boldsymbol{\beta}, 0, \bar{s})}{\partial \beta_i}. \quad (\text{A.5})$$

Moreover

$$W^{yc}(\boldsymbol{\beta}, \bar{s}, \bar{s}) - W^{yc}(\boldsymbol{\beta}, 0, 0) = V(\bar{s}, \bar{s}) - V(0, 0) - (1 + \lambda)\bar{s}Q$$

is independent of $\boldsymbol{\beta}$. Assume $W^{yc}(\boldsymbol{\beta}, 0, 0) \geq W^{yc}(\boldsymbol{\beta}, \bar{s}, \bar{s})$. The line of proof for the $W^{yc}(\boldsymbol{\beta}, 0, 0) < W^{yc}(\boldsymbol{\beta}, \bar{s}, \bar{s})$ case is similar and thus omitted. Define two new functions $L_i(\boldsymbol{\beta}) = W^{yc}(\boldsymbol{\beta}, \bar{s}, 0) - \max\{W^{yc}(\boldsymbol{\beta}, 0, 0), W^{yc}(\boldsymbol{\beta}, 0, \bar{s})\}$ and $L_{-i}(\boldsymbol{\beta}) = W^{yc}(\boldsymbol{\beta}, 0, \bar{s}) - \max\{W^{yc}(\boldsymbol{\beta}, \bar{s}, 0), W^{yc}(\boldsymbol{\beta}, 0, 0)\}$. Obviously, $s_i^{yc}(\boldsymbol{\beta}) = \bar{s}$ iff $L_i(\boldsymbol{\beta}) \geq 0$ and $s_{-i}^{yc}(\boldsymbol{\beta}) = \bar{s}$ iff $L_{-i}(\boldsymbol{\beta}) \geq 0$. It follows immediately from (A.5) that $\partial L_i(\boldsymbol{\beta})/\partial \beta_i \leq 0$ and that $\partial L_{-i}(\boldsymbol{\beta})/\partial \beta_i \geq 0$. Consequently, $s_i^{yc}(\boldsymbol{\beta})$ is non-increasing and $s_{-i}^{yc}(\boldsymbol{\beta})$ non-decreasing in β_i . ■

A.5. List of the 50 largest US cities 2000³⁶

Ranked by order of descending population size: New York NY, Los Angeles CA, Chicago IL, Houston TX, Philadelphia PA, Phoenix AZ, San Diego CA, Dallas TX, San Antonio TX, Detroit MI, San Jose CA, Indianapolis IN, San Francisco CA, Jacksonville FL, Columbus OH, Austin TX, Baltimore MD, Memphis TN,

³⁶Downloaded from <http://www.infoplease.com/ipa/A0763098.html> 14 January 2002.

Milwaukee WI, Boston MA, Washington DC, Nashville TN, El Paso TX, Seattle WA, Denver CO, Charlotte NC, Fort Worth TX, Portland OR, Oklahoma City OK, Tucson AZ, New Orleans LA, Las Vegas NV, Cleveland OH, Long Beach CA, Albuquerque NM, Kansas City MO, Fresno CA, Virginia Beach VI, Atlanta GA, Sacramento CA, Oakland CA, Mesa AZ, Tulsa OK, Omaha NE, Minneapolis MN, Honolulu HI, Miami FL, Colorado Springs CO, St. Louis MO, Wichita KS.

References

- [1] Auriol, Emmanuelle (1993): "Monopole ou duopole: L'effet de comparaison," *Annales d'Economie et de Statistique* 31: 1–31.
- [2] Auriol, Emmanuelle (1998): "Deregulation and quality," *International Journal of Industrial Organization* 16: 169–194.
- [3] Auriol, Emmanuelle (2000): "Concurrence par comparaison: Un point de vue normatif," *Revue Economique* 51: 621–634.
- [4] Auriol, Emmanuelle and Jean-Jacques Laffont (1992): "Regulation by duopoly," *Journal of Economics and Management Strategy* 1: 507–533.
- [5] Baiman, Stanley and Joel S. Demski (1980): "Economically optimal performance evaluation and control systems," *Journal of Accounting Research* 18 (supp): 184–220.
- [6] Broyles, Robert W. and Michael D. Rosko (1985): "A qualitative assessment of the Medicare Prospective Payment System," *Social Science and Medicine* 20: 1185–1190.
- [7] Carroll, Norman V. and W. Gary Erwin (1987): "Patient shifting as a response to Medicare prospective payment," *Medical Care* 25: 1161–1167.
- [8] Cremer, Jacques and Richard P. McLean (1988): "Full extraction of the surplus in Bayesian and dominant strategy auctions," *Econometrica* 56: 1247–1257.
- [9] Dalen, Dag M. (1998): "Yardstick competition and investment incentives," *Journal of Economics and Management Strategy* 7: 105–126.
- [10] Demski, Joel S. and David E.M. Sappington (1984): "Optimal incentive contracts with multiple agents," *Journal of Economic Theory* 53: 152–171.

- [11] Feinglass, Joe and James J. Holloway (1991): "The initial impact of the Medicare Prospective Payment System on U.S. health care: a review of the literature," *Medical Care Review* 48: 91-115.
- [12] Folland, Sherman, Allen C. Goodman and Miron Stano (1997): *The Economics of Health and Health Care*, 2nd. ed., Upper Saddle River, NJ: Prentice-Hall.
- [13] Holmström, Bengt (1982): "Moral hazard in teams," *Bell Journal of Economics* 13: 324-340.
- [14] Holmström, Bengt and Paul Milgrom (1991): "Multitask Principal-Agent analyses: incentive contracts, asset ownership and job design," *Journal of Law, Economics and Organization* 7: 24-52.
- [15] Johnson, Dana E. (1984): "Life, death and the dollar sign: medical ethics and cost containment," *Journal of the American Medical Association* 252: 223-224.
- [16] Laffont, Jean-Jacques and David Martimort (2000): "Mechanism design with collusion and correlation," *Econometrica* 68: 309-342.
- [17] Laffont, Jean-Jacques and Jean Tirole (1993): *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press.
- [18] Lewis, Tracy R. and David E.M. Sappington (1991): "Incentives for monitoring quality," *Rand Journal of Economics* 22: 370-84.
- [19] Mookherjee, Dilip (1984): "Optimal incentive schemes with many agents," *Review of Economic Studies* 51: 433-446.
- [20] Nalebuff, Barry J. and Joseph E. Stiglitz (1983): "Prizes and incentives: towards a general theory of compensation and competition," *Bell Journal of Economics* 14: 21-43.
- [21] Shleifer, Andrei (1985): "A theory of yardstick competition," *RAND Journal of Economics* 16: 319-327.
- [22] Sobel, Joel (1999): "A reexamination of yardstick competition," *Journal of Economics and Management Strategy* 8: 33-60.
- [23] Tangerås, Thomas P. (2001): "Collusion-proof yardstick competition," *Journal of Public Economics*, forthcoming.
- [24] Tirole, Jean (1988): *The Theory of Industrial Organization*, Cambridge, MA: MIT Press.