Racing to the Bottom: Competition and Quality[*]

Limor Golan[†]   Christine A. Parlour[‡]   Uday Rajan[§]

November 7, 2007

[†]`limor@andrew.cmu.edu`

[‡]`parlour@haas.berkeley.edu`

[§]`urajan@umich.edu`

Racing to the Bottom: Competition and Quality

Abstract

We model competition between risk-neutral principals who hire weakly risk-averse agents to produce a good of variable quality. The agent can increase the likelihood of producing a high-quality good by providing costly effort. We demonstrate that, when the agent is strictly risk-averse, the cost of providing incentives increases in the number of other firms in the industry. We characterize conditions under which the first-best outcome involves each firm inducing high effort. We then consider firms in competition, and identify parameter conditions under which (i) each firm induces high effort in the short run and low effort in the long run (ii) the first-best outcome has each firm inducing high effort, but long-run equilibrium results in each firm inducing low effort. Thus, in the long run, the average quality in the industry deteriorates, and increased competition leads to a "race to the bottom" in quality.

1 Introduction

Should social policy encourage competition in service industries? The answer appears to be "yes," judging from the deregulation of financial markets (the National Market System), competitive provision of directory assistance in the UK or the plethora of subprime mortgage brokers. However, is there an economic basis for this argument? This question is of immediate importance because the service industry is so large: In the U.S., it accounts for approximately two-thirds of domestic production.

Some recent experiences suggest that competition in service sectors has not been exemplary. For example, Propper, Burgess and Gossage (2003) examine the reforms of the National Health Service in the UK. Using quality measures (such as mortality) they find a negative relationship between quality and the degree of competition. Similarly, a National Audit Office Report on the privatization of Britain's directory enquiry services in November 2003 concluded that, initially, the proportion of accurately provided telephone numbers was only 62%. While this improved to 86% over a year, usage had fallen off dramatically, especially in the over-55 age group. Meanwhile, competition had increased substantially: A year after the privatization, as many as 217 directory enquiry numbers were in service.

The standard argument in favor of competition is that, fixing a production technology and factor prices, free entry drives firms to produce at the minimum of the long run cost curve, which is socially efficient. In service industries, the good being produced is typically intangible and depends on the interaction of agents. Indeed, there is no reason to view the "cost function" as invariant to market structure, an idea fundamental to the efficiency of competitive equilibrium. As we show, competition can change the cost of producing high quality services, rendering the notion of "the minimum of the long run average cost curve" specious.

We present a simple model of an industry in which risk-neutral principals offer incentive contracts to agents. If an agent puts in high effort, the good produced is more likely to be of high quality. A monopsonist consumer observes the quality offered by each firm. She then chooses a firm to patronize and buys a fixed amount. In our model there are therefore two contractible states. We derive the optimal contract between the principal and the agent, and show that, with a risk-averse agent, the cost of inducing high effort depends on the structure of the industry. In particular, it increases with both the number of high-effort firms and the number of low-effort firms in the industry.

We then consider first-best outcomes, in which firms can directly contract on effort and therefore are not subject to an incentive problem. We show that, if the monetary cost of high effort is sufficiently low, the first-best outcome has only high-effort firms in the

1

industry. Although quality is stochastic, an industry with only high-effort firms will have a high expected or average quality.

We consider two types of competitive equilibria — short run, in which the number of firms is fixed, and long run, with free entry of firms. In the latter, firms make close to zero profits (modulo an integer number of entrants). We focus on the case in which all firms optimally induce high effort in the short run. As the number of firms in the industry increases, the incentive (in terms of an increase in expected revenue) to incurring costly effort decreases. Further, if the agent is risk-averse, the cost to providing incentives increases. This is because competition can make the contracting problem between the principal and agent more noisy: An agent can fail to make a sale either because he shirked or because he faced an aggressive competitor. Both of these effects go in the same direction; for an identified set of parameter values, there is eventually a race to the bottom in long-run equilibrium, with each firm providing low effort. As a result, the average quality of services produced in the industry degenerates. The more risk-averse the agent (so the more severe the incentive problem), the sooner competitive forces drive the economy toward low effort.

In an example we show that a stark divergence between the short run and the long run can occur even when the agent is risk-neutral. With a risk-neutral agent, it is costless for the principal in the firm to align incentives. Thus, a wedge between first-best and long-run equilibrium is created solely due to excess entry in the latter. When the agent is risk-averse, the additional factor of compensating the agent for bearing risk leads to a quicker divergence from the first-best outcome.

Our model is more appropriate for service than manufacturing industries. First, we assume an immediate link between agent effort and product quality. The service sector is more consistent with flexible quality choice: It is more difficult to upgrade a car factory than it is to provide incentives for better service. Second, only the consumer knows the quality of the good. By nature, services are experience goods: It is easier to measure a car's attributes than to determine if a waiter was polite. Third, we do not explicitly consider price competition. Unlike the market for widgets, decreasing price in services industries does not automatically lead to increased market share, since the price may signal the quality of the product.

The previous literature on the effect of product market competition on managerial incentives largely takes the industry structure as exogenous, with the exception of Raith (2003), which we discuss later. Early work in this area includes Hart (1983), who provides a model in which competition reduces managerial slack by making it easier (i.e., cheaper) to provide the agent with incentives to put in high effort. Scharfstein (1988) demonstrates that Hart's result relies critically on a discontinuity in the utility function. With a continuous utility

function and a strictly risk-averse agent, competition exacerbates the incentive problem, as in our model.

In Hart (1983) and Scharfstein (1988) contracts only on an individual firm's absolute performance, rather than its relative performance in the industry. Holmstrom (1982) and Nalebuff and Stiglitz (1983) point out that, if output in an industry is correlated, then an important benefit of competition is the ability to base the agent's compensation on her relative performance. An immediate implication is that a monopolist would benefit from hiring multiple agents to generate more information. Unlike their frameworks, in our model, the industrial structure (i.e., one principal matched with one agent) is not inefficient per se: while a monopolist might internalize the market revenue externality we present, it could not arrange production more cheaply than a series of isolated principal-agent pairs.

Hermalin (1992) considers a manager offering a contract to shareholders (so the manager's participation constraint clearly does not bind). He demonstrates that competition has an ambiguous effect on managerial incentives: Competition may change the relative payoff of actions, and may induce the manager to consume different amounts of perquisites. He also identifies a "risk-adjustment" effect that arises because competition may change the informativeness of the agent's action. In our model, increased competition decreases the informativeness of an agent's action, which therefore requires a risk premium. This increased cost to the competing principals affects their equilibrium quality choice.

Schmidt (1997) considers the effect of competition on managerial incentives to reduce costs with a risk-neutral manager who incurs a utility cost if the firm goes bankrupt. Competition is modelled in reduced form, via a parameter in a firm's demand function. Increased competition increases the likelihood of a firm going under, so the manager works harder in an effort to stave off the personal cost of bankruptcy. Thus, increased competition unambiguously reduces the cost to implementing a higher level of effort. However, the marginal benefit of cost reduction (i.e., greater effort) is ambiguous in sign, and may decrease as competition increases. The tradeoff between these two effects implies that competition may sometimes lead to lower effort. Notably, in his model, if the participation constraint of a manager is binding (i.e., if managers in a competitive industry are not "scarce"), increased competition unambiguously leads to greater effort provision. In our model, the participation constraint always binds as employees are not scarce, yet we obtain the opposite result: with a risk-averse agent, competition increases the cost of providing incentives.

An important point of departure for our paper from the literature cited so far is that we endogenize the structure of the industry, by considering a Nash equilibrium in which each firm optimally chooses its effort level (i.e., its contract) in response to the choices of other firms in the market. This enables a direct comparison of the first-best structure with the

3

outcome of a long-run equilibrium with free entry.

A closely related paper is Raith (2003), which models entry and exit of firms on a circle. Each firm consists of a risk-neutral principal and a risk-averse agent. Production costs are decreasing in the unobservable effort exerted by an agent. However, as Raith assumes that realized costs are directly contractible, and any noise in the mapping from effort to cost is independent across firms, the cost of providing incentives is independent of the degree of competition. Rather, only the benefit to inducing a particular level of effort changes with changes in competition. By contrast, we find that the cost of inducing a particular quality level in the optimal contract depends on the market structure. Indeed, the cost of inducing high effort is increasing in the number of competing firms.

We briefly outline our model in Section 2. We present the optimal contract and expected cost of inducing high effort to the principal in Section 3. We demonstrate that, when the agent is risk-averse, the cost of inducing high effort is increasing in the number of other competing firms. We establish the first-best benchmark in Section 4, and identify conditions under which the first-best outcome consists of only high effort firms. Essentially, the cost of effort to an agent must be sufficiently low. The effects of competition, in both the short and long run, are examined in Section 5. We establish conditions under which long run competition (with free entry) deviates from the first-best. Broadly, these include the agent being sufficiently risk-averse. However, as we demonstrate via an example in Section 5.1, even with a risk-neutral agent, free entry can undermine a firm's willingness to induce high effort, due to a decreased benefit of effort.

2   Model

An industry with $n \geq 1$ firms provides a service that can be of variable quality. Each firm is owned by a risk-neutral principal who contracts with an agent (an employee) to produce the service. The employee chooses effort $e \in \{e_h, e_\ell\}$. The quality of the service offered depends on the effort level: high effort yields high quality with probability q, while low effort yields high quality with probability $1 - q$, where $q > \frac{1}{2}$. Agents are all identical. Each agent is weakly risk-averse with a utility function over wealth denoted by $u(\cdot)$. The agent has a reservation wage w_0 that gives rise to a reservation utility $u_0 = u(w_0)$. His utility is separable in wealth and the cost of effort. Let $g(\cdot)$ denote the inverse function of $u(\cdot)$.

Let e_0 denotes the monetary cost to the agent of high effort and e_u denotes the agent's disutility of exerting high effort. Then, $e_u = u(w_0 + e_0) - u(w_0)$. Low effort entails zero cost, and generates no disutility.

The service is purchased by a representative customer. She observes the quality of the

4

good offered by each firm, and spends y at one randomly chosen high-quality firm. (A constant revenue is consistent with a game in which the consumer has beliefs on quality given prices, coupled with arbitrary small costs of sequential searching. Her optimal strategy is as follows: If a price signals low quality, she moves on and buys from the first high quality firm she finds. Firms then have no incentive to reduce price.[1]) The principal cannot observe the quality of the service experienced by each customer, but does observe the sales revenue generated by her employee.

Thus, in our model, are two possible revenue states that the principal can observe: a high outcome, y, or a low one, which we normalize to zero.[2] Corresponding to the two revenue states, there are two wage levels she optimally offers, designated as w_h and w_ℓ respectively. Thus, the contract offered by the principal is defined by (w_h, w_ℓ). We assume that $w_0, e_0 > 0$, and $w_0 + e_0 \leq qy$, so that a monopolist facing no incentive problem has a non-negative expected profit if it induces high effort.

Let p_h denote the probability of achieving high revenue (y) when the agent puts in a high effort, and p_ℓ the corresponding probability when he chooses low effort. That is, $p_h = \Pr(y \mid e_h)$, and $p_\ell = \Pr(y \mid e_\ell)$. From the viewpoint of a single firm, p_h and p_ℓ are exogenous. However, the probabilities depend on the quality choices of the other firms.

To see this, consider an arbitrary firm i, and suppose that the industry consists of n_h other firms (i.e., not including firm i) that provide high effort, and n_ℓ other firms that produce low effort. Suppose that firm i provides high effort. The consumer will purchase from a firm only if it produces high quality. She randomizes with equal probability across all high-quality firms. If firm i has high effort, it produces high quality with probability q. Suppose exactly k of the other $n_h + n_\ell$ in the industry also produce high quality. The event that exactly k of the other $n_h + n_\ell$ firms have high quality has probability

$$
\begin{aligned}
\psi(k \mid n_h, n_\ell) &= \sum_{j=\max\{k-n_h,0\}}^{\min\{k,n_\ell\}} \binom{n_h}{k-j} q^{k-j}(1-q)^{n_h-(k-j)} \binom{n_\ell}{j}(1-q)^j q^{n_\ell-j} \\
&= \sum_{j=\max\{k-n_h,0\}}^{\min\{k,n_\ell\}} \binom{n_h}{k-j}\binom{n_\ell}{j} q^{n_\ell+k-2j}(1-q)^{n_h-k+2j}
\end{aligned}
$$

Further, if firm i generates high quality and exactly k of the remaining firms also generate high quality, the consumer buys from firm i with probability $\frac{1}{k+1}$. Thus, the overall

---

[1] For example, the 2005 NAO study on Britain's directory enquiry numbers found that 80% of the market share belonged to two numbers that charged on average 30% more than the previous regulated price, and price competition was limited by consumer's inability to find the best price.

[2] A generalization to a strictly positive low outcome is straightforward.

probability that firm i generates revenue y when it induces high effort is

$$p_h(n_h, n_\ell) = q \sum_{k=0}^{n_h+n_\ell} \psi(k \mid n_h, n_\ell) \frac{1}{k+1}$$

If, instead, firm i provides low effort, it generates high quality with a lower probability, $1 - q$. However, the probability distribution over the information that the consumer gets from the other $n_h + n_\ell$ firms is independent of the effort of firm i. Therefore,

$$\begin{aligned} p_\ell(n_h, n_\ell) &= (1-q) \sum_{k=0}^{n_h+n_\ell} \psi(k \mid n_h, n_\ell) \frac{1}{k+1} \\ &= \frac{1-q}{q} p_h(n_h, n_\ell). \end{aligned}$$

Notice that, if there are no other firms in the industry, $p_h(0,0) = q$ and $p_\ell(0,0) = 1 - q$.

The expected revenue of firm i is $p_h(n_h, n_\ell)y$ if it has high effort and $p_\ell(n_h, n_\ell)y$ if it has low effort. We show that (as is intuitive), the expected revenue declines as the industry grows more competitive via the entry of either high or low effort firms. That is, the probability that firm i earns a high revenue declines in both n_h and n_ℓ. Importantly, the difference between p_h and p_ℓ, which represents the marginal benefit to firm i of inducing high effort, also declines as the industry becomes more competitive.

Lemma 1 *The probabilities $p_h(n_h, n_\ell)$ and $p_\ell(n_h, n_\ell)$ are both strictly decreasing in n_h and n_ℓ, as is the difference $p_h(n_h, n_\ell) - p_\ell(n_h, n_\ell)$.*

Proofs of all results are relegated to the Appendix, in Section 7.

We turn to an examination of industry structure on the cost to a firm of inducing high effort.

3   Cost of High Effort and the Employee Risk Premium

Suppose that the principal wishes to induce low effort. Then, she must pay the agent enough for him to achieve the reservation utility u_0. Since the principal is risk-neutral, risk-sharing is not beneficial. Thus, an optimal low-effort contract has the agent earning the wage w_0 in both states. If the agent is strictly risk-averse, this is the unique optimal contract when low effort is induced.

By contrast, suppose the principal wishes to induce high effort. The optimal contract minimizes the cost of doing so, and specifies a wage in the high revenue state, w_h, and the low, w_ℓ, that are the solution to the problem:

$$\min_{w_h, w_\ell} \quad p_h w_h + (1 - p_h)w_\ell$$

$$\text{subject to}: \quad p_h u(w_h) + (1 - p_h)u(w_\ell) - e_u \quad \geq \quad p_\ell u(w_h) + (1 - p_\ell)u(w_\ell) \quad (1)$$

$$p_h u(w_h) + (1 - p_h)u(w_\ell) - e_u \quad \geq \quad u_0, \quad (2)$$

where (1) is the incentive compatibility constraint, and (2) is the participation constraint.

Define $u_h = u(w_h)$ and $u_\ell = u(w_\ell)$. That is, u_h is the utility induced by the high wage and u_ℓ is the that generated by the low wage. Recall that $g(\cdot)$ denotes $u^{-1}(\cdot)$. Then, we can write the problem with agent utilities u_h, u_ℓ rather than wages as the choice variables:

$$\min_{u_h, u_\ell} \quad p_h g(u_h) + (1 - p_h)g(u_\ell)$$

$$\text{subject to}: \quad (IC) \quad (p_h - p_\ell)(u_h - u_\ell) \geq e_u$$

$$(PC) \quad u_\ell + p_h(u_h - u_\ell) \geq u_0 + e_u.$$

It is immediate that the participation constraint (PC) must bind at the optimum. If not, u_ℓ and u_h can both be reduced by the same amount, leaving the incentive compatibility constraint (IC) unaffected. Further, if the agent is strictly risk-averse, the solution to the above problem is unique, with both the (IC) and (PC) constraints binding. If the agent is also risk-neutral, there is a continuum of optimal contracts each satisfying (PC) with equality. For ease of comparison to the risk-averse case, when the agent is risk-neutral, we select the contract which also satisfies (IC) with equality.

Given an optimal high-effort contract (u_h^*, u_ℓ^*), the cost to the firm of eliciting high effort is just the expected wage, $p_h \, g(u_h^*) + (1 - p_h) \, g(u_\ell^*)$. Since the firm takes p_h and p_ℓ as given, we write $c(p_h, p_\ell)$ to denote the cost to the firm when it wishes to induce high effort.

Lemma 2 *The optimal contract that elicits high effort satisfies $u_\ell = u_0 - \frac{p_\ell e_u}{(p_h - p_\ell)}$ and $u_h = u_0 + \frac{(1 - p_\ell)e_u}{(p_h - p_\ell)}$. The cost to the firm when high effort is elicited is therefore*

$$c(p_h, p_\ell) \quad = \quad p_h \, g(u_0 + \frac{(1 - p_\ell)e_u}{(p_h - p_\ell)}) + (1 - p_h) \, g(u_0 - \frac{p_\ell e_u}{(p_h - p_\ell)}).$$

In the special case of a risk-neutral agent (i.e., with $g(u) = u$) the cost of inducing high effort is readily obtained to be $w_0 + e_0$. In particular, the cost is independent of p_h and p_ℓ, and is therefore invariant to the structure of the industry.

When the agent is risk-averse, however, the cost of inducing high effort depends on p_h and p_ℓ. In turn, p_h and p_ℓ depend on the number of other firms in the market. Thus, the cost of inducing high effort if the agent is risk-averse depends critically on the industrial structure. With a slight abuse of notation, we refer to $c(n_h, n_\ell)$ as the cost to an arbitrary

firm i of inducing high effort, when there are n_h other firms in the industry providing high effort and n_ℓ firms in the industry providing low effort.

We show that the cost of effort is increasing in the number of both high and low quality firms in the industry. The intuition for this result is as follows: As the number of firms in the industry increases, the likelihood that the agent loses the sale is higher even if he puts in high effort. As a result, his wage when revenue is high must increase, else he will earn less than his reservation utility. However, the increased difference in the wages in the high and low revenue states exposes the agent to increased risk. If he is risk-neutral, this increase in risk is irrelevant. However, if he is strictly risk-averse, he must be further compensated for bearing extra risk. In this way, the "production" of effort becomes more costly as the industry grows more competitive.

Proposition 1 *Suppose the agent is strictly risk-averse; that is, $u'' < 0$. Then, the cost of inducing high effort, $c(\cdot)$, strictly increases in each of n_h and n_ℓ.*

The excess cost of inducing high effort, $c(n_h, n_\ell) - [w_0 + e_0]$, may also be interpreted as the risk-premium that must be provided to the agent to induce him to participate in the contract. Since quality is stochastic, high effort necessarily leads to uncertainty in outcomes. A risk-averse agent must be compensated for the extra uncertainty. Of particular interest to us going forward is the risk-premium that a monopolist must provide, or $c(0, 0) - [w_0 + e_0]$.

Suppose, for example, the agent has CARA utility, so that $u(w) = -exp(-\rho w)$, where ρ is the coefficient of absolute risk aversion. Then, $g(u) = \frac{1}{\rho}(-\ln(-u))$. Suppose w_0 and e_0 are held constant, with $e_u = u(w_0 + e_0) - u(w_0)$ varying as ρ changes. In this case, $c(0, 0)$ is increasing in the risk aversion coefficient ρ, as is the marginal cost of effort, $c(0, 0) - w_0$. Finally, as observed earlier, with a risk-neutral agent, $c(\cdot) = w_0 + e_0$ in all cases.

The overall decision to provide high effort, of course, depends on both the benefit and cost of high effort. In the next section, we consider the first-best structure of the industry, in which a planner who does not face incentive issues chooses the optimal industry structure, and turn to the competitive structure in Section 5. In the latter, each firm decides independently whether to enter the industry, and, if so, which effort level to provide.

4   First-Best Industry Structure

In the first-best structure, there is no incentive problem: a planner wishing to induce high effort merely needs to compensate agents for the marginal disutility of high effort, e_u (the cash equivalent of which is e_0). In addition, the planner chooses the number of high and low effort firms, n_h and n_ℓ to maximize social welfare. For convenience, we assume consumer

surplus to be zero, so that overall welfare is zero if no sale occurs, and equal to the aggregate industry profit if a sale occurs.[3]

The consumer buys if she believes that the service will be of high quality, therefore a sale occurs as long as at least one firm produces high quality. If there are n_h firms inducing high quality and n_ℓ firms inducing low quality, then the event that every firm produces low quality has probability $(1-q)^{n_h} q^{n_\ell}$. Thus, with probability $1 - (1-q)^{n_h} q^{n_\ell}$, at least one firm produces high quality. Further, the cost to the planner is w_0 for each low-effort firm and $w_0 + e_0$ for each high-effort firm. Thus, the aggregate industry profit is given by

$$\Pi(n_h, n_\ell) = \left[1 - q^{n_\ell}(1-q)^{n_h}\right] y - (n_h + n_\ell) w_0 - n_h e_0.$$

By definition, an industry structure with n_h firms eliciting high effort and n_ℓ firms eliciting low effort is first-best if aggregate industry profit cannot be improved by changing either the number of high-effort or the number of low-effort firms; that is, if

$$\Pi(n_h, n_\ell) \geq \Pi(\tilde{n}_h, \tilde{n}_\ell) \text{ for any } \tilde{n}_h, \tilde{n}_\ell \geq 0.$$

It is intuitive that, if e_0 is sufficiently high (as an extreme case, consider $e_0 \geq y$), the first-best allocation would have no high-effort firms. This case is uninteresting, since when e_0 is sufficiently high, no firm in competition will choose to supply high effort. Therefore, to highlight the difference between the competitive and first-best structures, we consider a situation under which e_0 is sufficiently low so that the first-best structure entails only high-effort firms.

Consider a planner who is offering high quality services in an $n-1$ firm industry. If he adds one more agent, he is faced with an increased wage bill of $w_0 + e_0$. The chance that the consumer buys the good (but did not before) is $q(1-q)^{n-1}$, namely, the chance that the firm produces high quality when all else fails. Define $\hat{n}^f$ as the largest integer n for which

$$q(1-q)^{n-1} \geq \frac{w_0 + e_0}{y}. \tag{3}$$

That is, if there are $\hat{n}^f$ high-effort firms in the market, adding an extra high-effort firm results in all $\hat{n}^f + 1$ firms making a loss. Clearly, a planner will never wish to have more than $\hat{n}^f$ high-effort firms. In Proposition 2 below, we identify a condition under which it is optimal for a planner to have exactly $\hat{n}^f$ firms in the industry, each incurring high effort.

Given our assumption that $q \geq \frac{w_0 + e_0}{y}$, inequality (3) that defines $\hat{n}^f$ is satisfied for $n = 1$. Further, the left-hand side decreases to zero as n gets large. Thus, a unique $\hat{n}^f$ exists, and is at least 1. Notice, also that $\hat{n}^f$ is increasing in y the per customer revenue.

---

[3]It is straightforward to generalize to the case in which the consumer obtains a utility $\hat{y} > y$ from consuming the good, with a consumer surplus $\hat{y} - y$.

Ensuring that it is optimal to supply effort will frequently result in comparing the payoff to a firm that induces high effort to a firm that induces low. Define

$$B = \frac{2q - 1}{1 - q}.$$

Since $2q - 1 = q - (1 - q)$, B is the probability that high effort generates high quality minus the probability that low effort generates high quality, divided by the probability of high quality given low effort. It summarizes the benefit to a firm (in terms of increased likelihood of high quality) of inducing high effort.

Proposition 2 *Suppose that the agent's cost of effort is sufficiently low so that $e_0 \leq Bw_0$. Then, the first-best allocation has $\hat{n}^f$ firms in the industry, and each firm provides high effort.*

The condition on e_0 is intuitive: To ensure that all firms producing high quality is first best, it must be the case that for an industry of size $\hat{n}^f$, it is not in the central planner's best interest to convert a worker's contract to one that induces low effort. If the central planner does so, then he saves the effort cost, e_0, but reduces the probability of success. This effect is captured by the B term, which relates to the marginal benefit (in terms of higher probability) of inducing high effort. The condition in the proposition is conservative; a tighter sufficient condition is $\frac{e_0}{y} \leq (2q - 1)(1 - q)^{\hat{n}^f}$. From condition (3), $\hat{n}^f$ depends on w_0, e_0, and y. We show in the proof of Proposition 2 that the requirement that $\frac{e_0}{w_0} \leq B$ is sufficient to ensure that the cost of effort is low compared to the revenue potential y.

5   Racing to the Bottom

Do market forces "work" in service industries? If the first-best industry structure entails only high-effort firms, will competition lead to the same outcome? There are two reasons why competition may yield an outcome divergent from the first-best one. First, if the agent is risk-averse, the incentive problem has bite, and the cost to a firm in equilibrium of inducing high effort is higher than the corresponding cost to a central planner. Second, even controlling for the incentive cost, too many firms may wish to enter the industry in equilibrium.

Our focus here is on the latter effect, which represents a race to the bottom in the industry.[4] To distinguish between the two effects, we consider equilibria in both the short

---

[4]It is straightforward that, if the agent is sufficiently risk-averse, a firm subject to an incentive problem may wish to induce low effort whereas a firm which does not face such a problem is willing to induce high effort.

and long run. In the short-run, every firm in the industry must earn a non-negative profit, and must be satisfied with its own effort (i.e., it should not be able to increase its expected profit by switching effort). However, new firms cannot enter the market, so that firms in the industry may earn positive profits. In the long-run, no firm must be willing to enter the market at either effort level. In other words, existing firms must make close to zero profits.[5] Since firms in the short-run face the same incentive problem as firms in the long-run, any divergence between a short-run outcome and a long-run one must be due to excess entry.

Recall that the profit of an individual firm in an industry with $n_h + n_\ell + 1$ firms is

$$\pi_i(n_h, n_\ell) = \begin{cases} p_h(n_h, n_\ell)y - c(n_h, n_\ell) & \text{if firm } i \text{ provides high effort} \\ p_\ell(n_h, n_\ell)y - w_0 & \text{if firm } i \text{ provides low effort.} \end{cases}$$

We first define short-run equilibrium.

Definition 1 In a short-run industry equilibrium with n_h firms inducing high effort and n_ℓ firms inducing low effort then
(i) No high effort firm wishes to either switch to low effort or to exit: If $n_h \geq 1$, then

$$p_h(n_h - 1, n_\ell)\, y - c(n_h - 1, n_\ell) \;\geq\; \max\{p_\ell(n_h - 1, n_\ell)\, y - w_0 \,,\, 0\}.$$

(ii) No low quality firm wishes to either switch to high quality or to exit: If $n_\ell \geq 1$, then

$$p_\ell(n_h, n_\ell - 1)\, y - w_0 \;\geq\; \max\{p_h(n_h 1, n_\ell - 1)\, y - c(n_h, n_\ell - 1) \,,\, 0\}.$$

Next, consider long-run equilibrium. In the long run, there are no barriers to entry, and firms can continue to enter as long as positive profits are available.

Definition 2 A long-run equilibrium with n_h firms inducing high effort and n_ℓ firms inducing low is a short-run equilibrium with the additional provision that no new firm wishes to enter the market at either effort level:

$$p_h(n_h, n_\ell)y - c(n_h, n_\ell) \leq 0 \quad \text{and} \quad p_\ell(n_h, n_\ell)y - w_0 \leq 0.$$

Our goal here is not to characterize all equilibria of the game, but rather to consider situations under which competitive equilibria do not implement the first best. We consider an extreme divergence between outcomes in the short and long run, under which competition leads to all firms supplying high effort in the short run, but following the entry of new firms,

---

[5]An integer number of firms may preclude exactly zero profits for industry participants.

each firm in the industry provides low effort in the long run. In this case, firm entry directly reduces the average quality of the industry. More broadly, of course, there will be parameter values for which the industry in either the short or long run may have a mix of high and low effort firms. We focus on the extremes to make our point more cleanly.

First, consider the short run. If competing firms produce high quality, then the industry size must be sufficiently small. We identify a threshold number of firms such that each firm in the industry supplies high effort if the short-run number of firms is less than the threshold. Define $\hat{n}_s^c$ as the largest integer n to satisfy

$$p_h(n-1,0) \geq \max\left\{\frac{q}{(1-q)B}\left(\frac{c(n-1,0)-w_0}{y}\right), \frac{c(n-1,0)}{y}\right\} \tag{4}$$

The two parts of the RHS come from the joint requirements that firms producing high quality do not switch to low quality and secondly, do not exit. From Lemma 1, we know that the LHS of (4) is declining in n, and it goes to zero as n gets large. From Proposition 1, both terms on the RHS are increasing in n. Thus, if the condition is satisfied at $n = 1$, there exists a unique $\hat{n}_s^c$ greater than or equal to one. Further, the condition is also satisfied by all $n < \hat{n}_s^c$. Notice also that the more risk-averse the agent, the faster the RHS of (4) increases. Thus, increasing risk aversion on the part of the agent leads to a weakly lower $\hat{n}_s^c$.

The maximum number of firms that the industry can sustain is determined by a zero profit condition. In the long-run, if the industry has $\hat{n}_\ell^c$ firms and each firm supplies low effort, and no new firms wish to enter then $\hat{n}_\ell^c$ is the largest integer that satisfies

$$p_\ell(0,n-1) \geq \frac{w_0}{y}. \tag{5}$$

Note that the left-hand side is declining in n. Thus, if there are $\hat{n}_\ell^c$ low-effort firms in the industry, and an extra low effort firm were to enter, each firm would make a loss, since $p_\ell(0,n)y < w_0$.

We next identify sufficient conditions for there to be a stark divergence between short-run and long-run equilibrium outcomes: in the short-run, each firm induces high effort, whereas in the long-run, each firm supplies low effort. Essentially, high effort in the short-run entails a sufficiently low cost of effort; conversely, low effort in the long-run entails a sufficiently high cost of effort. The next proposition formalizes these arguments by identifying bounds for the cost of effort under which both phenomena obtain.

Proposition 3 *Suppose that (i)* $\frac{c(0,0)-w_0}{y} \in \left[\frac{(1-q^2)B}{2}, (1-q)B\right]$, *and (ii)* $\frac{c(0,0)}{y} \leq q$. *Then, if the number of firms in short run equilibrium is no greater than* $\hat{n}_s^c$ *each firm supplies high*

effort. However, there is a long-run equilibrium with $\hat{n}_\ell^c > \hat{n}_s^c$ firms, each supplying low effort.

In the proposition, condition (ii) merely ensures that a monopolist earns a non-negative profit by inducing high quality. The expression $\frac{c(0,0)-w_0}{y}$ in condition (i) measures the extra cost to a monopolist of providing incentives per unit revenue. The upper bound of (i) ensures that a monopolist earns a higher profit from high quality, as compared to low quality. The lower bound in condition (i) ensures that, in the long-run, a firm would rather supply low quality. This lower bound is conservative. As we show in the proof of the Proposition, the exact condition is $\frac{c(\hat{n}_\ell^c - 1, 0) - w_0}{y} \geq B \frac{1 - q^{\hat{n}_\ell^c}}{\hat{n}_\ell^c}$. A sufficient condition is found by requiring the inequality to hold at $\hat{n}_\ell^c = 2$, and considering $c(0,0)$ rather than $c(\hat{n}_\ell^c - 1, 0)$ on the left-hand side.

Suppose the agent is risk-neutral, so that $c(0,0) = w_0 + e_0$. Then, condition (i) reduces to $\frac{e_0}{y} \in \left[\frac{(2q-1)(1+q)}{2}, (2q-1) \right]$. Fix y and pick a value of e_0 to satisfy this condition; condition (ii) then merely requires that $w_0 + e_0 \leq qy$, which is easily arranged by choosing w_0 to be sufficiently low. Thus, even when the agent is risk-neutral (so that the incentive problem is irrelevant), there can be a divergence between a short-run equilibrium and a long-run equilibrium. This divergence is highlighted via an example in Section 5.1.

Under which conditions do the first-best solution and the competitive outcome diverge? Again, we focus on extreme outcomes, in which the first-best outcome entails all firms supplying high effort and yet, in long-run equilibrium all firms supply low effort. Putting together the conditions from Propositions 2 and 3 provides the requirements for such a divergence. As with Proposition 3, condition (i) in the proposition below is conservative. The exact condition is $\frac{c(0,0)-w_0}{y} \geq \frac{(1-q^{\hat{n}_\ell^c})B}{\hat{n}_\ell^c}$.

Proposition 4 *Suppose (i) $\frac{c(0,0)-w_0}{y} \geq \frac{(1-q^2)B}{2}$, and (ii) $e_0 \leq w_0 B$. Then, the first-best solution has $\hat{n}^f$ firms, each providing high effort. However, there exists a long-run equilibrium that has $\hat{n}_\ell^c$ firms, each providing low effort.*

There are two effects that cause competition to diverge from the first best. First, suppose the agent is risk-neutral, so that the principal can costlessly align the agent's incentives. Then, condition (i) in the Proposition reduces to $\frac{e_0}{y} \geq \frac{(1-q^2)B}{2}$ (as mentioned after Proposition 3, this condition is conservative). With a risk-neutral agent, the race to lower effort occurs entirely due to a reduction in the benefit of high effort, analogous to the "business stealing" effect identified by Raith (2003). As the number of firms increases, the benefit accruing to a firm of inducing effort decreases. If all other firms produce high quality then each firm's revenue will be proportional to $\frac{1}{n}$. Therefore, for fixed costs, the

benefit of high effort declines and free entry drives firms to produce low quality because of declining expected benefit.

However, in addition to this revenue effect, we identify a cost externality: With a risk-averse agent, increasing the number of competitors increases the cost to each firm of providing incentives, by reducing the likelihood of achieving a high revenue. This cost is important, because it determines how "quickly" the industry diverges from first best.

To emphasize the effect of agent risk-aversion, we explicitly consider a situation in which, with a risk-neutral agent, there exists a long-run equilibrium with all firms providing high effort. However, with sufficiently risk-averse agents, there is a long-run equilibrium with each firm providing low effort.

Define $\hat{n}^h$ to be the largest integer that satisfies

$$\left[\frac{1 - (1 - q)^n}{n}\right] y \geq w_0 + e_0. \tag{6}$$

Since $qy > w_0 + e_0$ by assumption, $\hat{n}^h$ exists and weakly exceeds one.

If e_0 is sufficiently low, there exists a long-run equilibrium with $\hat{n}^h$ firms in the market, each supplying high effort. However, even at the low cost of effort in a world with certainty, if the agent is sufficiently risk-averse, there is a long-run equilibrium in which each firm supplies low effort.

Proposition 5 *There exists an $\bar{e} > 0$ such that, if $\frac{e_0}{y} \leq \bar{e}$ and the agent is risk-neutral, it is a long-run equilibrium for there to be $\hat{n}^h$ firms in the market, each supplying high effort. However, with risk averse agents, if $\frac{c(0,0) - w_0}{y} \geq \frac{(1 - q^2)B}{2}$, it is a long-run equilibrium for there to be $\hat{n}_\ell^c$ firms in the market, each supplying low effort.*

It is intuitive that when e_0 (the cost of effort to a firm employing a risk-neutral agent) is low, an equilibrium will entail high-effort firms. As mentioned in Section 3, the quantity $[c(0,0) - w_0]$ represents the marginal cost to a monopolist of inducing high effort, and increases in the risk-aversion of the agent. To see this, suppose that the agent has CARA utility. For any fixed e_0 that satisfies the first part of the proposition, if the coefficient of risk aversion is sufficiently high, there will be a long-run equilibrium with only low-effort firms.

5.1 Example

To illustrate our results, we consider the following numeric example. Let $q = 0.7$, $y = 1$, $w_0 = 0.08$, and $e_0 = 0.11$. At these parameter values, $\hat{n}^f = 2$ and $\hat{n}_\ell^c = 12$. We demonstrate that even with risk-neutral agents, there exists a long-run equilibrium with low-effort firms,

along with a long-run equilibrium with high-effort firms. When the agent is sufficiently risk-averse, the high-effort equilibrium no longer exists.

Notice that $e_0 = 0.11 > w_0 B = 0.1067$, so the condition of Proposition 2 is not satisfied. However, as mentioned, this condition represents a loose upper bound; the exact condition is $\frac{e_0}{y} \leq (2q - 1)(1 - q)^{\hat{n}^f - 1} = 0.12$, which is satisfied. Therefore, the first-best outcome entails two high-effort firms in the industry, and no low-effort firms.

The optimality of two high-effort firms when there is no incentive problem may be checked directly. The industry profit in the first-best case is given by $[1 - (1 - q)^2]y - 2(w_0 + e_0) = 0.53$. Switching one firm to low-effort yields a profit $[1 - q(1 - q)]y - 2w_0 - e_0 = 0.52$. Removing one firm entirely yields a profit $qy - (w_0 + e_0) = 0.51$. Adding a third high-effort firm yields $[1 - (1 - q)^3]y - 3(w_0 + e_0) = 0.40$. Finally, adding a third firm that provides low-effort yields a profit $[1 - q(1 - q)^2]y - 3w_0 - 2e_0 = 0.48$. Thus, it is optimal to have two high-effort firms, and no low-effort firms.

Now, suppose the agent is risk-neutral. Then, from condition (4), $\hat{n}_s^c = 5$. Note that condition (i) of Proposition 3 is violated, since $\frac{e_0}{y} = 0.11 < \frac{(1 - q^2)B}{2} = 0.34$. However, as mentioned in the text, the right-hand side of the condition represents a loose lower bound for e_0. The exact condition in the risk-neutral case is $\frac{e_0}{y} \geq \frac{1 - q^{\hat{n}_\ell^c}}{\hat{n}_\ell^c} B = 0.1096$. Thus, there is a long-run equilibrium with 12 low-effort firms and no high-effort firms in the market.

Note that, when the agent is risk-neutral, a firm in equilibrium incurs the same effort cost as the central planner. That is, the incentive problem has no bite. Nevertheless, the example shows that there is a divergence between the first-best outcome and long-run equilibrium. In the long-run, too many firms enter the industry, causing all firms to switch to low effort.

Next, suppose the agent has CARA utility with risk-aversion coefficient 1,that is, her utility over consumption is defined by $u(w) = -e^{-w}$. Now, the marginal cost of high effort exceeds e_0 at each industry configuration, reducing the profit from providing high effort. Thus, $\hat{n}_s^c$ falls to 3. The long-run equilibrium in which all firms provide low-effort is unaffected.

These results are illustrated in Figure 1, which plots the profit of firm i when all other firms in the industry are providing high effort (left-hand panel) and low effort (right-hand panel).

Consider the left-hand panel of Figure 1 first. If all other firms in the industry are providing high effort, and the agent is risk-neutral, it is a best response for firm i to also provide high effort as long as the number of other firms in the industry is 4 or less. When there are 5 other firms in the industry, firm i makes a loss regardless of the effort level it provides. Thus, it is a long-run equilibrium for there to be exactly 5 high-effort firms in
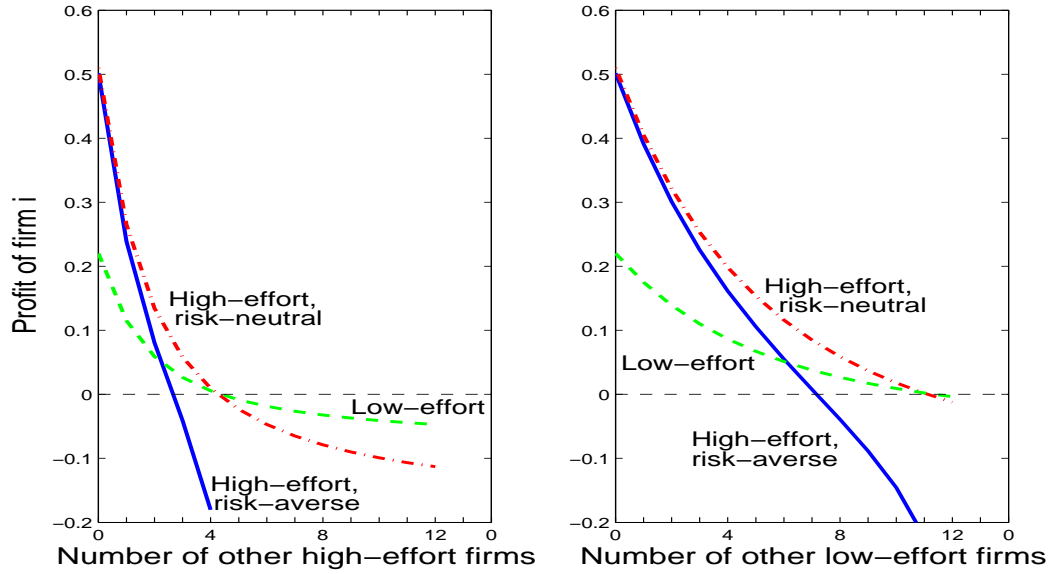
Figure 1: Firm profits in short-run and long-run equilibrium

the industry. Conversely, if the agent is risk-averse as described above, the profit from high effort is lower: now, high effort is a best response only if the number of other firms in the industry is 2 or lower.

Next, consider the right-hand panel. If all other firms in the industry provide low effort, and the agent is risk-neutral, it is a best response for firm i to provide high effort if there are 10 or fewer other firms, but to switch to low effort when there are 11 other firms. Thus, it is a long-run equilibrium for there to be 12 low-effort firms in the market. Therefore, when the agent is risk-neutral, there exist multiple long-run equilibria: one in which all firms provide high effort, and another in which all firms provide low effort. In the short-run, if 6 or fewer other firms provide low effort, it is a best response to provide high effort.

Therefore, in this example:

- The first-best outcome has 2 high-effort firms.

- There is a long-run equilibrium with 12 low-effort firms, regardless of the agent's preferences.

- When the agent is risk-neutral, there is a long-run equilibrium with 5 high-effort firms, and short-run equilibria with 5 or fewer high-effort firms.

16

- When the agent is risk-averse with CARA coefficient 1, there are short-run equilibria with 3 or fewer high-effort firms.

6 Conclusion

Competition in this environment can induce a race to the bottom, and reduce welfare. At some point competitive forces encourage too much entry which leads to a lower average industry quality. As firms enter the market, they reduce the incentives of existing firms to produce high quality and also are more likely to introduce low quality goods.

That competition in service industries has pernicious effects has a number of policy implications. First, barriers to entry may be a good thing. Specifically trade restrictions or professional accreditations that prevent entry or reduce supply may result in higher quality. Since effort is positively correlated with quality, the implication of Proposition 3 is that, if the marginal cost of high effort is sufficiently high (but not high enough to entirely preclude provision of effort), the average quality of the service offered will be low in long-run equilibrium, but may be high in the short-run. Thus, in sectors identifiable with costly effort, prevention of entry may preserve a high average quality across firms. For example, industry associations in professions such as accountancy, law, and medicine, are often accused of erecting barriers to entry in an effort to keep supply low. Our model provides a justification for such barriers: in their absence, the average quality of the service would plummet.

Second, the proponents of government "outsourcing" argue that opening up competition in the provision of services generates service provision at a lower cost. This model suggests that there is another externality at work, namely that an increase in competition in a service industry may actually decrease the quality of the good. If quality is difficult to measure it suggests that monopoly or restricted provision may be optimal if quality is an issue.

Finally, the health care market is one service industry that has befuddled most high income countries. Aside from anecdotal evidence, there is some indication that increased competition has generated worse outcomes: Propper, Burgess and Gossage (2003) conclude that competition has led to lower quality in the provision of health services by the National Health Service in the UK.

This is a very stylized model, but the intuition that competing firms can make it more expensive to elicit high effort in service industries is robust. Whether that results in equilibrium lower effort levels depends, of course on the benefit to the principal in so doing. Unreported in this paper, we have also considered a model in which the revenue generated by a transaction depends on the number of firms in the industry. The results go through

(albeit with more parameter restrictions). In sum, increased competition in any industry with risk-averse employees offering services may be susceptible to a race to the bottom.

7 Appendix: Proofs

Proof of Lemma 1

Recall that p_h is defined as $p_h(n_h, n_\ell) = q \sum_{k=0}^{n_h+n_\ell} \psi(k \mid n_h, n_\ell) \frac{1}{k+1}$. Therefore, $p_h(\cdot)$ is the expectation of a random variable which takes on values in the set with $\{\frac{1}{n_h+n_\ell+1}, \frac{1}{n_h+n_\ell}, \ldots, 1\}$, and the density at $\frac{1}{k+1}$ is given by $\psi(k \mid n_h, n_\ell)$.

We will show that the distribution generated by $\psi(\cdot \mid \tilde{n}_h, \tilde{n}_\ell)$ first-order stochastically dominates the distribution generated by $\psi(\cdot \mid n_h, n_\ell)$ whenever either $\tilde{n}_h < n_h$ or $\tilde{n}_\ell < n_\ell$, or both.

First, consider the binomial distribution over $\{0, \ldots, n\}$, with success probability p. Let $\phi(k \mid n, p) = \binom{n}{k} p^k (1-p)^{n-k}$ denote the density at k and let $\Phi(\cdot \mid n, p)$ denote the associated distribution.

Consider any $k \in \{0, 1, 2, \ldots, n\}$. Then, $\Phi(k \mid n, p) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{k-i}$. Hence, $\Phi(k \mid n+1, p) = \sum_{i=0}^{k} \binom{n+1}{i} p^i (1-p)^{n-i+1}$. Now, $\binom{n}{i} + \binom{n}{i+1} = \binom{n+1}{i+1}$ whenever $n \geq i+1$. Therefore,

$$\Phi(k \mid n+1, p) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i+1} + \sum_{i=0}^{k-1} \binom{n}{i+1} p^i (1-p)^{n-i+1}.$$

We can now write

$$
\begin{aligned}
\Phi(k \mid n, p) - \Phi(k \mid n+1, p) &= \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}[1 - (1-p)] - \sum_{i=0}^{k-1} \binom{n}{i+1} p^i (1-p)^{n-i+1} \\
&= \sum_{i=0}^{k} \binom{n}{i} p^{i+1} (1-p)^{n-i} - \sum_{i=1}^{k} \binom{n}{i} p^{i+1} (1-p)^{n-i} \\
&= p(1-p)^n > 0.
\end{aligned}
$$

Therefore, the distribution $\Phi(\cdot \mid n+1, p)$ first-order stochastically dominates by $\Phi(\cdot \mid n, p)$.

Now, notice that $\psi(k \mid n_h, n_\ell) = \sum_{j=\max\{k-n_h, 0\}}^{\min\{k, n_\ell\}} \phi(k - j \mid n_h, q) \phi(j \mid n_\ell, (1-q))$. Let $\Psi(k \mid n_h, n_\ell) = \sum_{i=0}^{k} \psi(i \mid n_h, n_\ell)$ be the probability that at least k of the remaining firms are high quality.

Suppose one additional high-effort firm enters the market, so that the number of high-effort firms (other than firm i, which we are interested in) increases from n_h to $n_h + 1$, while the number of other low quality firms remains fixed at n_ℓ. Note that $\phi(\cdot \mid n_\ell, (1-q))$ is unchanged, whereas $\Phi(k - j \mid n_h + 1, q) < \Phi(k - j, \mid n_h, q)$ for all $(k - j) \in \{0, \cdots, n_h\}$. Since the last inequality holds for all $j \in \{\max\{0, k - n_h\}, \cdots, \min\{k, n_\ell\}\}$, it follows that

18

$\Psi(k \mid n_h + 1, n_\ell) < \Psi(k \mid n_h, n_\ell)$ for all $k \in \{0, \cdots, n_h + n_\ell\}$. It follows immediately that $p_h(\cdot)$ is strictly decreasing in n_h.

A similar argument shows that $p_h(\cdot)$ is strictly decreasing in n_ℓ.

Now, $p_\ell(n_h, n_\ell) = \frac{1-q}{q} p_h(n_h, n_\ell)$, and $p_h(n_h, n_\ell) - p_\ell(n_h, n_\ell) = (2 - \frac{1}{q}) p_h(n_h, n_\ell)$. Since $p_h(\cdot)$ is strictly decreasing in n_h and n_ℓ, so are p_ℓ and $p_h - p_\ell$. ∎

Proof of Lemma 2

First, suppose the agent is strictly risk-averse. Then, the objective function $p_h g(u_h) + (1-p_h) g(u_\ell)$ is strictly convex. From the arguments in the text, the participation constraint (PC) must bind. If the IC does not bind, the first-order conditions imply that $g'(u_h) = g'(u_\ell)$, or $u_h = u_\ell$. However, a constant wage will violate the IC constraint. Therefore, the IC constraint must bind.

Next, suppose the agent is risk-neutral. Then, any u_h, u_ℓ which Then, any u_ℓ, u_h that exactly satisfy the PC result in the same cost (u_0). Thus, any pair (u_h, u_ℓ) that exactly satisfy the PC and weakly satisfy the IC represent an optimal contract. In particular, the solution obtained when both constraints bind is optimal.

When both IC and PC bind, solving the two equations simultaneously yields the utilities exhibited in the statement of the Lemma. The expected cost to the firm is the expected wage, and is given by

$$c(p_h, p_\ell) = p_h\, g(u_0 + \frac{(1 - p_\ell)e_u}{(p_h - p_\ell)}) + (1 - p_h)\, g(u_0 - \frac{p_\ell e_u}{(p_h - p_\ell)}).$$

∎

Proof of Proposition 1

The cost of inducing high effort is

$$c(n_h, n_\ell) = p_h g(u_0 + \frac{(1 - p_\ell)e_u}{p_h - p_\ell}) + (1 - p_h) g(u_0 - \frac{p_\ell e_u}{p_h - p_\ell}), \qquad (7)$$

where p_h, p_ℓ are functions of n_h and n_ℓ.

Since $p_\ell(n_h, n_\ell) = \frac{1-q}{q} p_h(n_h, n_\ell)$ for every pair (n_h, n_ℓ) it follows that $p_h - p_\ell = \frac{2q-1}{q} p_h$ and $(1 - p_\ell) = \frac{q - (1-q)p_h}{q}$, where we suppress the dependence on n_h, n_ℓ for brevity. Substituting these expressions into (7),

$$c(n_h, n_\ell) = p_h\, g\left(u_0 + \frac{(q - (1-q)p_h)e_u}{(2q-1)p_h}\right) + (1 - p_h)\, g\left(u_0 - \frac{(1-q)e_u}{2q-1}\right).$$

Let $u_h = u_0 + \frac{(q - (1-q)p_h)e_u}{(2q-1)p_h}$ and $u_\ell = u_0 - \frac{(1-q)e_u}{2q-1}$. Then,

$$\frac{\partial c(n_h, n_\ell)}{\partial p_h} = g(u_h) - g(u_\ell) + p_h g'(u_h)\frac{\partial u_h}{\partial p_h} = g(u_h) - g(u_\ell) - g'(u_h)\left(\frac{q e_u}{(2q-1)p_h}\right)$$
$$= g(u_h) - g(u_\ell) - g'(u_h)\frac{e_u}{p_h - p_\ell}.$$

19

Further, $u_h - u_\ell = \frac{e_u}{p_h - p_\ell}$, so that

$$\frac{\partial c(n_h, n_\ell)}{\partial p_h} = g(u_h) - g(u_\ell) - g'(u_h)(u_h - u_\ell).$$

Now, given $g(u(w)) = w$ for every w, it follows that $g'(u(w))u'(w) = 1$. Since $u'(\cdot) > 0$, it further follows that $g'(\cdot) > 0$.

Further, $g''(u')^2 + g'u'' = 0$, or $g''(u(w)) = -\frac{g'(u(w))u''(w)}{(u'(w))^2}$ for every w. Now, since $g'(\cdot) > 0$, $u''(\cdot) < 0$, and $u'(\cdot) > 0$, it must be that $g''(\cdot) > 0$, or g is strictly convex. Therefore,

$$g'(u_h) > \frac{g(u_h) - g(u_\ell)}{u_h - u_\ell}, \quad \text{or} \quad g(u_\ell) > g(u_h) - g'(u_h)(u_h - u_\ell).$$

Therefore, $\frac{\partial c(n_h, n_\ell)}{\partial p_h} < 0$.

Now, from Lemma 1, p_h is strictly decreasing in n_h and n_ℓ. Thus, $c(\cdot)$ is strictly increasing in each of n_h and n_ℓ. ∎

Proof of Proposition 2

Notice that the aggregate industry profit is concave in n_h and n_ℓ. Thus, it is sufficient to consider a situation in which all firms are producing high effort, and show that local changes in the number of high and low effort firms do not increase profit.

Suppose, therefore, that the first-best industry structure consists of n firms, each producing high effort. Then, the following four conditions must hold:

(i) Increasing the number of high effort firms to $n + 1$ must not increase profit. That is,

$$[1 - (1 - q)^n]y - n(w_0 + e_0) \geq [1 - (1 - q)^{n+1}]y - (n + 1)(w_0 + e_0),$$
$$\text{or,} \qquad q(1 - q)^n y \leq w_0 + e_0. \tag{8}$$

Now, $\hat{n}^f$ is defined as the largest integer n for which $q(1 - q)^{n-1} \geq \frac{w_0 + e_0}{y}$. Thus, if there are $\hat{n}^f$ firms in the industry, $q(1 - q)^n y < w_0 + e_0$, so that (8) holds.

(ii) Reducing the number of high effort firms to $n - 1$ must not increase profit. That is,

$$[1 - (1 - q)^n]y - n(w_0 + e_0) \geq [1 - (1 - q)^{n-1}]y - (n - 1)(w_0 + e_0),$$
$$\text{or,} \qquad q(1 - q)^{n-1} y \geq w_0 + e_0, \tag{9}$$

which holds at $n = \hat{n}^f$, given the definition of $\hat{n}^f$.

20

(iii) Switching a high effort firm to a low effort firm does not improve profit. That is,

$$[1 - (1 - q)^n]y - n(w_0 + e_0) \geq [1 - q(1 - q)^{n-1}]y - (n - 1)(w_0 + e_0) - w_0,$$

$$\text{or,} \qquad (2q - 1)(1 - q)^{n-1}y \geq e_0, \tag{10}$$

We have assumed that $e_0 \leq \frac{2q-1}{1-q}w_0$, which is equivalent to the inequality $\frac{w_0}{e_0} \geq \frac{1-q}{2q-1}$, and hence to $\frac{e_0+w_0}{e_0} \geq \frac{q}{2q-1}$, or $\frac{e_0+w_0}{q} \geq \frac{e_0}{2q-1}$.

Further, by definition of $\hat{n}^f$,

$$q(1 - q)^{\hat{n}^f - 1}y \geq e_0 + w_0,$$

or $(1 - q)^{\hat{n}^f - 1}y \geq \frac{e_0+w_0}{q}$. Since $\frac{e_0+w_0}{q} \geq \frac{e_0}{2q-1}$, it follows that $(1 - q)^{\hat{n}^f - 1}y \geq \frac{e_0}{2q-1}$, so that condition (10) holds.

(iv) Adding a low effort firm does not improve profit. That is,

$$[1 - (1 - q)^n]y - n(w_0 + e_0) \geq [1 - q(1 - q)^n]y - n(w_0 + e_0) - w_0,$$

$$\text{or,} \qquad (1 - q)^{n+1}y \leq w_0. \tag{11}$$

We show that (11) is satisfied, given (8) and (10). Multiply (10) by -1 and add to (8). This directly yields (11), which is therefore satisfied.

Thus, if there are exactly $\hat{n}^f$ high-effort firms and zero low-effort firms in the industry and $e_0 \leq Bw_0$, a local change in either the number of high or low effort firms does not increase aggregate profit. Since the aggregate industry profit is concave in n_h, n_ℓ, the first-best solution has $\hat{n}^f$ firms in the industry, each producing high effort. ∎

Proof of Proposition 3

First, we show that it is a long-run equilibrium for $\hat{n}_\ell^c$ firms in the market to each supply low effort. By definition of $\hat{n}_\ell^c$, no new firm wishes to enter the market and supply low effort. Also, by definition of $\hat{n}_\ell^c$, no firm wishes to exit the market, since each firm earns a weakly positive profit.

Thus, we need to check the following two conditions:

(i) No firm in the industry wishes to switch from low to high effort. No firm will switch in this manner if

$$p_h(0, \hat{n}_\ell^c - 1)y - c(0, \hat{n}_\ell^c - 1) \leq p_\ell(0, \hat{n}_\ell^c - 1)y - w_0. \tag{12}$$

Since $p_h(0, \hat{n}_\ell^c - 1) = \frac{q}{1-q}p_\ell(0, \hat{n}_\ell^c - 1)$, the above inequality can be re-written as $\frac{2q-1}{1-q}p_\ell(0, \hat{n}_\ell^c - 1)y \leq c(0, \hat{n}_\ell^c - 1) - w_0$. Now, note that $p_\ell(0, \hat{n}_\ell^c - 1) = \frac{1-q^{\hat{n}_\ell^c}}{n}$, and (from Proposition 1)

$c(0, \hat{n}_\ell^c - 1) \geq c(0,0)$. Thus, (12) is satisfied if

$$c(0,0) - w_0 \quad \geq \quad \frac{2q-1}{1-q} \left(\frac{1 - q^{\hat{n}_\ell^c}}{\hat{n}_\ell^c} \right) y, \tag{13}$$

It is straightforward to see that, for $n \geq 1$, the expression $\frac{1-q^n}{n}$ is decreasing in n. Hence, if condition (13) is satisfied for $\hat{n}_\ell^c = 2$, it is satisfied for all higher $\hat{n}_\ell^c$. When $\hat{n}_\ell^c = 2$, the condition reduces to $c(0,0) - w_0 \geq \frac{(2q-1)(1+q)}{2} y$, or $\frac{c(0,0)-w_0}{y} \geq \frac{(1-q^2)B}{2}$, which is assumed in the statement of the Proposition.

(ii) No firm wishes to enter and supply high effort. That is,

$$p_h(0, \hat{n}_\ell^c)y - c(0, \hat{n}_\ell^c) \quad \leq \quad 0. \tag{14}$$

Note that (12) can be written as

$$[p_h(0, \hat{n}_\ell^c - 1) - p_\ell(0, \hat{n}_\ell^c - 1)]y \quad \leq \quad c(0, \hat{n}_\ell^c - 1) - w_0.$$

Since $[p_h(\cdot) - p_\ell(\cdot)]$ is decreasing in both arguments and $c(\cdot)$ is increasing in both arguments, it follows that

$$[p_h(0, \hat{n}_\ell^c) - p_\ell(0, \hat{n}_\ell^c)]y \quad \leq \quad c(0, \hat{n}_\ell^c) - w_0. \tag{15}$$

Further, $p_\ell(0, \hat{n}_\ell^c)y < w_0$ by definition of $\hat{n}_\ell^c$. Add this last inequality to (15), and we obtain (14). Therefore, no firm wishes to enter and supply high effort.

Hence, there exists a long-run equilibrium with $\hat{n}_\ell^c$ firms in the market, each supplying low effort. Notice that the argument above implies that a sufficient condition for there to exist a long-run equilibrium with $\hat{n}_\ell^c$ firms, each supplying low effort, is $\frac{c(0,0)-w_0}{y} \leq \frac{(1-q^2)B}{2}$.

Next, we show that if there are $n \leq \hat{n}_s^c$ in the market in the short-run, each firm will supply high effort. There are two conditions we need to check:

(i) No firm wishes to switch from high to low quality; that is,

$$p_h(n - 1, 0)y - c(n - 1, 0) \quad \geq \quad p_\ell(n - 1, 0)y - w_0.$$

This last condition reduces to $\frac{2q-1}{q} p_h(n - 1, 0)y \geq c(n - 1, 0) - w_0$, which is satisfied by definition of $\hat{n}_s^c$. For $\hat{n}_s^c$ to weakly exceed one, the condition must be satisfied at $n = 1$. Since $p_h(0,0) = q$, this requires $c(0,0) - w_0 \leq (2q-1)y$, or $\frac{c(0,0)-w_0}{y} \leq (1-q)B$, which has been assumed.

(ii) No firm wishes to exit the market; that is,

$$p_h(n - 1, 0)y \quad \geq \quad c(n - 1, 0) \tag{16}$$

22

This condition is again satisfied by definition of $\hat{n}_s^c$. For $\hat{n}_s^c$ to weakly exceed one, the condition must be satisfied at $n = 1$, which implies that $c(0,0) \leq qy$.

Finally, we show that $\hat{n}_s^c < \hat{n}_\ell^c$. For this to hold, it is sufficient that, if there are $\hat{n}_\ell^c$ firms in the industry each supplying high effort, each firm makes a loss; that is, $p_h(\hat{n}_\ell^c - 1, 0) < \frac{c(\hat{n}_\ell^c - 1, 0)}{y}$. Now, $p(\hat{n}_\ell^c - 1, 0) = \frac{1 - (1-q)^{\hat{n}_\ell^c}}{\hat{n}_\ell^c}$. Thus, a sufficient condition for $\hat{n}_s^c$ to be strictly less than $\hat{n}_\ell^c$ is $\frac{1 - (1-q)^{\hat{n}_\ell^c}}{\hat{n}_\ell^c} < \frac{c(\hat{n}_\ell^c - 1, 0)}{y}$. Since $c(\cdot)$ is strictly increasing in both arguments (from Proposition 1), it is sufficient that

$$\frac{1 - (1-q)^{\hat{n}_\ell^c}}{\hat{n}_\ell^c} \quad < \quad \frac{c(0,0)}{y} \tag{17}$$

Further, from the definition of $\hat{n}_\ell^c$, it must be that $p_\ell(0, \hat{n}_\ell^c - 1) \geq \frac{w_0}{y}$, or $\frac{1 - q^{\hat{n}_\ell^c}}{\hat{n}_\ell^c} \geq \frac{w_0}{y}$. Multiply both sides by -1 and add to (17). This yields

$$\frac{c(0,0) - w_0}{y} > \frac{q^{\hat{n}_\ell^c} - (1-q)^{\hat{n}_\ell^c}}{\hat{n}_\ell^c}. \tag{18}$$

The RHS is declining in $\hat{n}_\ell^c$, so it is sufficient if the RHS holds at $\hat{n}_\ell^c = 2$, which yields $\frac{c(0,0) - w_0}{y} > \frac{2q-1}{2}$. However, by assumption $\frac{c(0,0) - w_0}{y} > (1+q)\frac{2q-1}{2} > \frac{2q-1}{2}$. Hence, $\hat{n}_s^c < \hat{n}_\ell^c$. ∎

Proof of Proposition 4

As shown in the proof of Proposition 2, the condition $\frac{e_0}{y} \leq w_0 B$ is sufficient for the first-best to have $\hat{n}^f$ high-effort firms and no low-effort firms.

Further, as shown in the proof of Proposition 3, a sufficient condition for there to exist a long-run equilibrium with only low-effort firms is $\frac{c(0,0) - w_0}{y} \geq \frac{(1 - q^{\hat{n}_\ell^c})B}{\hat{n}_\ell^c}$. Since the right-hand side of this last inequality is decreasing in $\hat{n}_\ell^c$, it is satisfied when $\frac{c(0,0) - w_0}{y} \geq \frac{(1-q^2)B}{2}$, provided that $\hat{n}_\ell^c \geq 2$. Condition (ii) in the statement of the Proposition ensures that $\hat{n}_\ell^c \geq 2$: Since $\frac{1-q^2}{2} = p_\ell(0,1)$, the condition implies that $p_\ell(0,1)y \geq w_0$. ∎

Proof of Proposition 5

In the proof of Proposition 3, we show that the condition $\frac{c(0,0) - w_0}{y} \geq \frac{(1-q^2)B}{2}$ is sufficient for there to exist a long-run equilibrium with $\hat{n}_\ell^c$ firms, each supplying low effort. Thus, what remains to be proved is the first part of the Proposition.

Suppose the agent is risk-neutral. Note that $\hat{n}^h$ may also be defined as the largest integer n for which

$$p_h(n - 1, 0)y \quad \geq \quad w_0 + e_0.$$

Thus, it follows that

$$p_h(\hat{n}^h - 1, 0)y - (w_0 + e_0) \geq 0 \geq p_h(\hat{n}^h, 0)y - (w_0 + e_0).$$

Therefore, if there are $\hat{n}^h$ firms in the industry, each supplying high effort, no existing firm has an incentive to leave the industry, and no new firm has an incentive to enter and supply high effort.

We therefore need to check that no existing firm wishes to switch to low effort, and no new firm wishes to enter and supply low effort. We check each of these in turn.

(i) If an existing firm switches to low effort, its profit is $p_\ell(\hat{n}^h - 1, 0)y - w_0 = \frac{1-q}{q} p_h(\hat{n}^h - 1, 0)y - w_0$. For such a deviation to be ruled out, we need

$$
\begin{aligned}
p_\ell(\hat{n}^h - 1, 0)y - w_0 &\leq p_h(\hat{n}^h - 1, 0)y - (w_0 + e_0), \\
\text{or,} \qquad \frac{e_0}{y} &\leq p_h(\hat{n}^h - 1, 0) - p_\ell(\hat{n}^h - 1, 0).
\end{aligned} \tag{19}
$$

Define

$$
\bar{e} = [p_h(\hat{n}^h, 0) - p_\ell(\hat{n}^h, 0)] = \frac{2q-1}{q} \left[\frac{1 - (1-q)^{\hat{n}^h+1}}{\hat{n}^h + 1} \right]
$$

Since the term $[p_h(\cdot) - p_\ell(\cdot)]$ is strictly decreasing in both arguments (Lemma 1), it follows that $\bar{e} < p_h(\hat{n}^h - 1, 0) - p_\ell(\hat{n}^h - 1, 0)$. Thus, if $\frac{e_0}{y} \leq \bar{e}$, condition (19) is satisfied, and no existing firm has an incentive to switch to low effort.

(ii) Suppose $\frac{e_0}{y} \leq \bar{e}$; that is,

$$
e_0 \leq [p_h(\hat{n}^h, 0) - p_\ell(\hat{n}^h, 0)]y. \tag{20}
$$

In addition, from the definition of $\hat{n}^h$,

$$
w_0 + e_0 > p_h(\hat{n}^h, 0)y. \tag{21}
$$

Multiply (20) by -1 and add to (21), to obtain

$$
w_0 > p_\ell(\hat{n}^h, 0)y.
$$

That is, no new firm can enter with low effort and earn a non-negative profit.

Therefore, if $\frac{e_0}{y} < \bar{e}$ and the agent is risk-neutral, it is a long-run equilibrium for there to be $\hat{n}^h$ firms in the industry, each supplying high effort. ∎

References

[1] Hart, Oliver (1983), "The Market as an Incentive Mechanism," *Bell Journal of Economics*, 14, 366–382

[2] Hermalin, Ben (1992) "The Effects of Competition on Executive Behavior," *RAND Journal of Economics* 23, 350–365.

[3] Holmstrom, Bengt (1982), "Moral Hazard in Teams," *Bell Journal of Economics* Vol. 13, 323–340.

[4] Nalebuff, Barry and Joseph Stiglitz (1983), "Information, Competition and Markets," *American Economic Review* Papers and Proceedings, 73, 278–283.

[5] National Audit Office Report (2005), "Directory Enquiries — From 192 to 118," Report by the Comptroller and Auditor General, U.K. Available at `http://www.nao.org.uk/publications/nao_reports/04-05/0405211.pdf`.

[6] Propper, Carol, S. Burgess and D. Gossage (2003), "Competition and Quality: Evidence from the NHS Internal Market 1991–1999," CMPO working paper series.

[7] Raith, Michael, (2003) "Competition, Risk and Managerial Incentives," *American Economic Review* Vol. 93, 1425–1436

[8] Schmidt, Klaus M. (1997), "Managerial Incentives and Product Market Competition." *Review of Economic Studies* Vol. 64(2), 191–213

[9] Scharfstein, David (1988) "Product Market Competition and Managerial Slack," *The Rand Journal of Economics* Vol 19(1), 147–155.