Performance Measures and Optimal Organization

Amir Ziv

Journal of Law, Economics, & Organization, Vol. 9, No. 1 (Apr., 1993), 30-50.

Stable URL:
http://links.jstor.org/sici?sici=8756-6222%28199304%299%3A1%3C30%3APMAOO%3E2.0.CO%3B2-8

Journal of Law, Economics, & Organization is currently published by Oxford University Press.

Performance Measures and Optimal Organization

Amir Ziv
Yale University

1. Introduction

Why are firms organized the way they are? What are the underlying factors that give rise to different organizational structures? "Why is not all production carried on in one big firm?" (Coase, 1937). The desire to understand the nature of the firm is a fundamental objective of organization theory, as well as a necessary building block for an analytical framework in managerial accounting. Much of the discussion in organization theory has centered on "transaction-cost economics" arguments, such as bounded rationality, asset specificity, opportunism, and uncertainty (see Williamson, 1985). The accounting literature has focused on a subset of these: strategic behavior in the presence of asymmetric information.

In this article, I explore the following issue: *Can differences in the way firms are organized be explained in terms of the underlying information structures?* Alternatively stated, what are the consequences of changes in the nature and precision of performance measures for the optimal firm structure, that is, the optimal number of workers and the resulting workers' contributions to the production process? Obviously, questions of organization structure

This article is based on Chapter 2 of my Ph.D. dissertation at Stanford University and was originally titled "Performance Measures and Optimal Firm Size." I would like to thank Joel Demski, Jonathan Feinstein, Joseph Greenberg, Stefan Reichelstein, Mark Wolfson, and, above all, Nahum Melumad for their suggestions, comments, and encouragement. Helpful comments from seminar participants at various universities are gratefully acknowledged. All errors are my own responsibility.

 8756-6222/93/$5.00

are more complicated; hierarchical structure, decentralization issues, business integration questions, and more should be considered as well. Further, information structure is not the only variable that affects organization structure. It seems, however, that answering the above questions may improve our understanding of the nature of the firm, by demonstrating the incremental role that information technologies play in determining firm size.

The tool I use to answer these questions is a principal–agent model. Such models frequently appear in the literature to represent intrafirm relationships and are used to analyze conflicts of interest under conditions of asymmetric information.[1] In these models, a principal contracts with an agent who possesses better information about the task to be performed. The literature analyzes two forms of asymmetric information: *moral hazard* (hidden action) and *adverse selection* (hidden information). These compound with differences in goals between the principal and the agent (each maximizes his own utility) to create a contracting problem.[2]

Under moral hazard, the input of the agent is unobservable by the principal, so the agent's compensation cannot depend *directly* on his input. If a risk-neutral principal could observe the agent's action, the risk-averse agent's compensation would be a fixed wage (with no risk imposed on the agent). This contract is not feasible under moral hazard. Given a fixed salary, the (work-averse) agent would not exert the required level of effort, since the agent is not held responsible for the outcome. The optimal contract under moral hazard results in a compensation scheme conditioned on the outcome, where the principal infers the agent's action from the observed outcome (signal). This contract implies nonoptimal risk sharing and efficiency reduction (see, e.g. Holmstrom, 1979). The intuition is that when the signal conveys less than perfect information, the principal may "mistakenly" punish the agent even though the agent had acted according to the principal's recommendation. Since the agent is risk averse, this risk is costly to the principal, who in turn reduces the required level of agent's effort. (By doing so, the principal increases the marginal productivity of the agent and equates it to the higher marginal costs.)

Further contracting problems arise when the principal employs more than one agent. In this case, one may have additional observation problems (e.g., the principal cannot distinguish among agents' contributions, as in Holmstrom, 1982), and the intercorrelation among agents may influence the agents' behavior, requiring changes in the optimal contract. See Demski and Sappington (1983); Holmstrom and Milgrom (1990); and Ma, Moore, and Turnbull (1988).

The literature to date has not dealt formally with the important question of

1. For a survey of the agency literature, see Baiman (1990).

2. Moral hazard is discussed below. Adverse selection occurs when the agent possesses superior information that the principal would like to incorporate into the optimal decision. Since the agent will use this information strategically, the principal should design the contract accordingly. Usually, information rent is included in the contract, above the agent's reservation price. See, for example, Rothschild and Stiglitz (1976), Myerson (1982), and Melumad and Reichelstein (1989).

optimal *firm size* or the number of agents the principal should employ.[3] In other words, the number of agents has been considered as exogenously given rather than treated as a decision variable.

In this article, I allow the principal to decide on the firm's optimal size in a moral hazard setting. Introducing size as a decision variable should improve our understanding of firm structure by revealing trade-offs between the number of agents employed and individual performance induced by an optimal contracting arrangement. If the number of agents is given exogenously, the reaction of the principal to changes in market parameters is constrained, and the solution is not optimal in general. In this article, the principal may vary the number of employees and/or choose different employment contracts as a function of the underlying information structure and market parameters. I identify the factors that influence the principal's decision in a model where the choice parameters are *contracts* and the *number of agents*. I provide intuition for how changes in information influence the firm's decisions, and explain when and why including the number of agents as a decision variable overturns some of the traditional agency results. This approach also enables me to deal *explicitly* with a market setting where questions of how much to produce and how to do so efficiently are relevant.

As will become apparent from the discussion of alternative information technologies, the dimensions analyzed cannot be under the control of the firm. Thus, one should interpret the results as the effect of available information technology on various measures of interest regarding the firm, rather than what is the optimal information system for the firm.

I distinguish among different information structures along the following dimensions: (i) level of detail about individual agents (individual versus aggregate information), and (ii) information accuracy.[4] I analyze the optimal solution to the principal's problem for several cases and present comparisons among different information structures.

After introducing the general model, I consider specific production functions, utility functions, and signal distributions. These specifications lead to cases where the first-best firm size is larger than the second-best firm size (Proposition 2), and to other cases where the reverse is true (Proposition 3). Dealing with cases involving loss of control (i.e., where the accuracy of each agent's signal decreases with the number of agents employed), I show that the optimal effort level is identical to that of the first-best firm, irrespective of the primitive level of inaccuracy (Proposition 4). The aggregation of individual signals (before being observable by the principal) may either cause no change in the optimal solution, if the aggregated signal is a sufficient statistic (Observation 1), or may increase costs (Propositions 5 and 6).

In the next section, I present the model and introduce the decision variables the firm faces. In Section 3, I show how different information structures affect the trade-off between individual effort and firm size, using two probability

3. Mirrlees (1976) deals qualitatively with these issues.

4. A third important dimension is correlation among signals regarding individual agents. Analysis of the consequences of correlation appears in Ziv (1990).

distributions to illustrate how and why the change in the distribution function may overturn some results. Closing remarks are provided in Section 4.

2. The Model

Consider a risk-neutral firm in a competitive market. The market price for its product is P, and its goal is to maximize expected profits Π. The firm has a production function $g_n(\mathbf{a})$, where $\mathbf{a}$ is the vector of the firm's employees' inputs and n is the number of workers in the firm.[5] The production function sets the mean for the firm's output Y, which is randomly distributed. Formally, $E(Y) \equiv \bar{Y} = g_n(\mathbf{a})$.

The firm hires employees from a large pool of (potential) identical risk- and work-averse workers, each with a utility function $u(w) - v(a)$, where w is the monetary compensation and a is the effort exerted by the agent.[6] Each agent has a market alternative of $\bar{u}$ with a strictly positive certainty equivalent.

Agents are subject to moral hazard: The agents' effort is not observable, and therefore cannot be contracted upon. Thus, compensation schemes based on performance measures are used to motivate agents.

The time line is as follows. The firm decides on the number of agents, n, it wants to employ and offers each a contract $s_i(\mathbf{x})$, where $\mathbf{x}$ is the set of available information (signals). I assume that $\mathbf{x} = \Theta(\bar{\mathbf{x}})$, where $\bar{\mathbf{x}} \in R^n$ is the set of individual signals and Θ is the information technology available to the firm. The primitive signals $\bar{\mathbf{x}}$ are transformed by Θ into an observable vector.[7]

Each agent, after accepting his contract, decides on the effort he wants to exert simultaneously with all other agents. Next, the accounting system produces its report (the set of signals $\mathbf{x}$), which is observed by the principal (the firm) and all agents. The compensation paid to the employees is dependent upon these signals.[8]

Given this contractual environment, the firm solves the following two problems: (i) how much to produce (or what is the optimal level of $\bar{Y}$ given P and the production function), and (ii) how to organize the production for each given $\bar{Y}$ (this becomes an interesting question only after I introduce the trade-off between n and a). These are two of the three basic neoclassical questions

5. The variable n in $g_n(\mathbf{a})$ is redundant, as it is included in the vector $\mathbf{a}$. It is included to emphasize the size parameter.

6. The assumption that agents are risk averse is crucial to my model. If agents are risk neutral (but still work averse), the principal can implement the first-best outcome, for any possible (informative) information system, by offering properly chosen linear contracts.

7. If an element of $\bar{\mathbf{x}}$ is $(x_1, x_2, \ldots, x_n)$, then $\Theta(\bar{\mathbf{x}})$ may be, for example, the summation operator $\mathbf{x} = \Theta(x_1, x_2, \ldots, x_n) = \Sigma x_i$.

8. Another source of information for the principal may be the realization of the output, Y. Inclusion of Y in contracts may provide further information on the agent's performance and reduces the information cost for the firm. This complicates the discussion but has no effect on the qualitative results. Any of the following assumptions will eliminate these complications: (i) The realization of output Y conveys no new information beyond the information in the signals (Y is not informative in the sense of Holmstrom 1979); this is the case, for example, if Y is the sum of all observed individual signals. (ii) The firm cannot contract on Y with the agents. This may happen if Y's realization is after the payment of compensation or if the firm itself is subject to moral hazard on the reporting of Y, and Y is unobservable to individual workers. [See Williamson (1985: 139) for a related case.]

for the firm (see, e.g., Lipsey, 1966). The third question, *what* to produce, is not addressed here: I model a single-product firm.

I ask how the availability and accuracy of different information technologies affects the firm's decision in terms of production level, number of agents, and induced effort. I am also interested in studying the effect of changes in the market price of output on the optimal levels of these variables. Formally, I am looking for changes in a^*, n^*, $\bar{Y}^*$, and $s^*(\mathbf{x})$ (where superscript * denotes optimal level) across different information systems and within the same system for different levels of accuracy.

I solve the firm's problem in reverse order. First, I deal with the problem of efficient production of a given expected output $\bar{Y}$. Next, I find the (marginal) cost function associated with each level of $\bar{Y}$ and then use it to determine the optimal output.

For efficient production of a given expected output $\bar{Y}$, the firm minimizes its costs under technology, information, and hiring constraints. Formally,

$$\underset{s_i(\mathbf{x}),\mathbf{a},n}{\text{Min}} \int \sum_{i=1}^{n} s_i(\mathbf{x}) f(\mathbf{x},\mathbf{a}) d\mathbf{x},$$

subject to, $\forall i = 1,\ldots,n$,

Individual Rationality (IR):

$$\int u_i(s_i(\mathbf{x})) f(\mathbf{x},\mathbf{a}) d\mathbf{x} - v_i(a_i) \geq \bar{u},$$

Incentive Compatibility (IC):

$$a_i \in \underset{\tilde{a}_i|a-i}{\text{argmax}} \int u_i(s_i(\mathbf{x})) f(\mathbf{x},(\tilde{a}_i,a_{-i})) d\mathbf{x} - v_i(\tilde{a}_i),$$

Production: $\quad g_n(\mathbf{a}) \geq \bar{Y}.$

The first constraint is the individual rationality (participation) constraint. The firm must provide agents with expected utility at least as high as their outside opportunities; otherwise it will not be able to hire any agent.

The second is the incentive compatibility constraint. Because the agent privately chooses his own level of effort, it must be the case that the effort level designed by the principal will be part of the agent's best response set; otherwise the agent will choose a different action.

To make the problem tractable, I impose additional structure on the model, and adopt three assumptions. The first pertains to the agents' utility functions.[9]

Assumption 1 (Utility Function). $u(w) - v(a) = 2\sqrt{w} - a^2$.

This simplification was used by Holmstrom (1979), among others; it has the major advantage of using the first two moments in the derivation of the

9. Alternatively, one can use the exponential utility function and linear contracts as in Holmstrom and Milgrom (1987). For all applicable cases, the results show no qualitative change when this alternative is used.

optimal contract. I will adopt this assumption throughout the rest of this article.

The next assumption relates to the production technology. In general, inputs can be characterized as complements or substitutes. When inputs are complements, one cannot replace the effort of an agent by effort from other agents. In this case, the number of agents is not a decision variable, and the questions that I address are not meaningful. When the inputs are substitutes, the firm can replace the effort of one agent with the effort of another. In this article, I focus on a case of substitutable efforts.[10,11] Specifically,

Assumption 2 (Production Function). $g_n(\mathbf{a}) = (\Sigma_1^n a_i)^\beta$, where $\beta \in (0,1)$.[12]

For most cases I use the normal distribution to illustrate alternative information structures.[13] I also use the exponential distribution to demonstrate additional properties of the firm size problem.

As a benchmark, I characterize the *first-best* case, where individual agents' inputs are either observed or can be deduced perfectly from the output—that is, the case in which the incentive constraint is not present. The Lagrangian becomes

$$\mathcal{L} = \int \left\{ \sum_{i=1}^{n} [s_i(\mathbf{x}) - \lambda_i(u_i(s_i(\mathbf{x})) - v_i(a_i) - \bar{u})]f(\mathbf{x},\mathbf{a}) \right\} d\mathbf{x}$$

$$- \mu(g_n(\mathbf{a}) - \bar{Y}).$$

The solution is

$$s_i(\mathbf{x}) = \lambda_i^2 = \left(\frac{\bar{u} + a_i^2}{2} \right)^2, \qquad a_f^* = \left(\frac{\bar{u}}{3} \right)^{1/2}, \qquad n_f = \left(\frac{3}{\bar{u}} \right)^{1/2} \cdot \bar{Y}^{1/\beta},$$

where subscript f denotes first best and subscript s denotes second best. Observe that the compensation is independent of the signal's realization (as in Holmstrom, 1979). This is so because the principal is interested in the agent's input and not in the realization of $\mathbf{x}$. Optimal risk sharing imposes all the risk on the (risk-neutral) principal with constant compensation for the agent.

The optimal effort level and contract—a_f^* and $s_f^*(\mathbf{x})$—are independent of the expected output $\bar{Y}$. There is a simple economic intuition behind this result.

10. Having capital as an additional input can demonstrate the trade-off between inputs as the information structure is changed. The relative changes within the labor input will remain the same.

11. This assumption precludes multitask consideration in contracting as in Holmstrom and Milgrom (1991).

12. Some of my results can be replicated for the more general Cobb–Douglas family $g_n(\mathbf{a}) = (\Sigma a_i)^\beta n^\alpha$. This family with $\alpha < 0$ represents the mutual disturbance among agents: The more agents there are, the less productive they are, so the same level of total effort produces less.

13. The normal distribution is commonly used in many applied models; it often makes the analysis more tractable, because its first two moments are independent of each other. Further, the multivariate normal distribution is also well defined, and we can use it in multiagent setting. More on the technical aspects of the use of the normal distribution in this setting appears in Ziv (1992).

Since a and n are perfect substitutes, the firm may change each one of them when it wants to change its production. Changing n has the expected fixed cost $E(s_i(\mathbf{x}))$ per agent, while the cost of changing a is increasing $[v''(a) > 0]$. Hence, once an agent reaches the optimal level of effort; it is less costly to increase the number of agents than to increase his optimal level of effort.

Next I move to the second decision the firm should make: how much to produce. The general expected profit function is $E(\Pi) = P \cdot E(Y) - E[TC(E(Y))]$—where TC and MC denote, respectively, total and marginal costs—with the familiar first-order condition: $MC = P$.

Using the fact that the Lagrangian multiplier, μ, represents the marginal cost of production, I can calculate

$$\mu_f = \frac{4\ \bar{u}^{3/2}}{\beta\ 3^{3/2}} \cdot \bar{Y}^{(1-\beta)/\beta}.$$

Equating it to P,

$$Y_f^* = \left(\frac{\beta\ 3^{3/2}}{4\ \bar{u}^{3/2}}\right)^{\beta/(1-\beta)} \cdot P^{\beta/(1-\beta)},$$

$$n_f^* = \left(\frac{3}{\bar{u}}\right)^{1/2} \cdot \left(\frac{\beta\ 3^{3/2}}{4\ \bar{u}^{3/2}}\right)^{1/(1-\beta)} \cdot P^{1/(1-\beta)}.$$

This formulation of the first best will serve as a benchmark for all second-best cases that follow.

3. Firm Size and Alternative Information Structures

Once actions of the agents are not observable, the principal should rely on other (available) information, represented here by realization of signals. In this section I solve the principal's problem under alternative information structures. I consider the following dimensions: (i) level of detail about individual agents (individual versus aggregate information), and (ii) information accuracy.

First I write the basic Lagrangian:[14]

$$\mathcal{L} = \int \left\{ \sum_{i=1}^{n} [s_i(\mathbf{x}) - \lambda_i(u_i(s_i(\mathbf{x})) - v_i(a_i))f(\mathbf{x},\mathbf{a}) - \eta_i(u_i(s_i(\mathbf{x}))f_{a_i}(\mathbf{x},\mathbf{a})) \right.$$

$$\left. + \lambda_i \bar{u} + \eta_i v_i'(a_i)] \right\} d\mathbf{x} - \mu(g_n(\mathbf{a}) - \bar{Y}).$$

Using the first-order condition with respect to $s_i(\mathbf{x})$, I conclude

$$(s_i(\mathbf{x}))^{1/2} = \lambda_i + \eta_i \frac{f_{a_i}(\mathbf{x},\mathbf{a})}{f(\mathbf{x},\mathbf{a})}. \qquad (1)$$

14. Observe that I replace the IC constraint with the local constraint $\int u_i(s_i(\mathbf{x}))f_{a_i}(\mathbf{x},\mathbf{a})d\mathbf{x} = v_i'(a_i)$. As pointed out by Mirrlees (1974), this is a weaker constraint and may be incorrect.

Observe that in Equation (1) compensation depends on the realization of the signal x. Because the principal cannot directly observe agents' actions, the principal uses the information available in the signal observed and imposes some of the risk in the signal on the agent to motivate the agent to exert effort. Equation (1) represents the optimal structure of the compensation function. The term $f_{a_i}(\mathbf{x},\mathbf{a})/f(\mathbf{x},\mathbf{a})$ represents the informativeness of the signal the principal observes, which depends on the information structure.

3.1 Individual Signals, Constant Variance (IC)

This case is the traditional principal–agent setting, modified only by endogenizing the number of agents. With the normal distribution assumption, $x_i \sim N(a_i,\sigma_i^2)$, the structure of $s_i(\mathbf{x})$ in Equation (1) becomes

$$(s_i(\mathbf{x}))^{1/2} = \left(\alpha_i + \gamma_i \frac{x_i - a_i}{\sigma_i^2}\right). \tag{2}$$

The optimal solution is

$$\begin{aligned}
a^*_{\mathrm{IC}} &= \left(\frac{-(\bar{u} + 2\sigma_i^2) + 2\sqrt{\bar{u}^2 + \sigma_i^4 + \bar{u}\sigma_i^2}}{3}\right)^{1/2}, \\
n^*_{\mathrm{IC}} &= \left[\frac{\beta a^{*\beta}_{\mathrm{IC}}}{a^{*2}_{\mathrm{IC}}(a^{*2}_{\mathrm{IC}} + \bar{u} + 2\sigma_i^2)}\right]^{1/(1-\beta)} \cdot p^{1/(1-\beta)}, \\
Y^*_{\mathrm{IC}} &= \left[\frac{\beta}{a^*_{\mathrm{IC}}(a^{*2}_{\mathrm{IC}} + \bar{u} + 2\sigma_i^2)}\right]^{\beta/(1-\beta)} \cdot p^{\beta/(1-\beta)}, \\
\mu_{\mathrm{IC}} &= \frac{a^*_{\mathrm{IC}}(a^{*2}_{\mathrm{IC}} + \bar{u} + 2\sigma_i^2)}{\beta} \cdot Y^{(1-\beta)/\beta}.
\end{aligned} \tag{3}$$

As in the first-best case, the optimal effort level is independent of the output; it depends on the reservation utility and on the level of precision of the signals. The economic intuition resembles the one discussed in the context of the first-best case. As before, the prediction of this model with respect to adjustments made by the firm when market conditions change—change the number of workers and keep effort level of each employee constant—is different from predictions made in models that assume a given number of

Rogerson (1985) found sufficient conditions for this approach to be valid. However, his conditions (the monotone likelihood ratio condition and convexity of the distribution function) are not satisfied by most of the commonly used distribution functions. Recently, Jewitt (1988) provided a different set of sufficient conditions for the approach to be valid in the single-agent setting. These conditions are satisfied for a broader set of distributions—in particular, the exponential family. Jewitt's proof holds for the case of individual independent signals in a multiple-agent setting. When the principal observes an aggregated signal, Jewitt's general proof is not applicable. For all of the cases discussed in this section, it is possible to show (directly) that the first-order approach is valid.

agents—change effort level. Also, observe that when $\sigma_i^2 = 0$ (perfect monitoring), the solution is the first-best one.

Proposition 1 shows that the effort under the IC case is lower than the effort under the first-best case, for sufficiently high reservation utility levels for the agents.[15] It also examines the effect of moral hazard on marginal costs and production level.

Proposition 1. Consider the case of individual signals and constant variance (IC). Then, (a) (i) the higher the variance of the signal, the smaller is the optimal level of effort, and (ii) for every $\sigma_i^2 > 0$, optimal effort level is below first-best effort level ($a_f^* > a_{\mathrm{IC}}^*$); (b) for any output level the first-best marginal cost is lower than the marginal cost in the IC case ($\mu_f < \mu_{\mathrm{IC}}$); and (c) the first-best production level is higher than the production level in the IC case ($Y_f > Y_{\mathrm{IC}}$).

Proof. All proofs are provided in the Appendix.

The intuition is that when the signal conveys less information, the principal may "mistakenly" punish the agent even though the agent had acted according to the principal's recommendation. Since the agent is risk averse, this risk is costly to the principal, who in turn reduces the level of agent's effort. (By doing so, the principal increases the marginal productivity of the agent and equates it to the higher marginal costs.) As will become apparent, this intuition does not apply to some of the cases presented below, where different agents' signals are related.

The optimal compensation scheme can be rewritten as $s_i(\mathrm{x}) = z_1 + z_2x + z_3x^2$, where $z_1 = 0.25(\bar{u} - a^{*2})^2$, $z_2 = (\bar{u} - a^{*2})a^2$, and $z_3 = a^{*2}$. Using the results of Proposition 1, I predict that $\partial z_1/\partial\sigma_i^2 > 0$, $\partial z_2/\partial\sigma_i^2 < 0$, and $\partial z_3/\partial\sigma_i^2 < 0$.[16] As information becomes less accurate, the principal increases the fixed component and decreases the slopes in the agents' contracts. The larger signals' variance implies an increase in the risk imposed on the agents; hence, the risk premium to the agents should increase. The principal trades off some of this additional risk by reducing the incentive component of the contract.

Next I discuss the relationship between first- and second-best firm size (in terms of number of workers). In general, it is not clear whether the first-best firm is larger or smaller than the second-best firm. The reason is the existence of two possibly opposite forces with no general dominance.

The first force is the difference in output. Since $Y_s < Y_f$, it seems the first-best firm will employ more agents.[17] But at the same time it is often the case, as in Proposition 1, that the required input from each agent is lower under the second-best. This, in turn, implies that *for a given level of output,* the second-best firm hires more employees. As I demonstrate later, the net effect is not uniquely determined.

15. See Ziv (1992) for a discussion of this requirement.

16. If one could observe σ_i^2, one would be able to test these predictions empirically.

17. The results of Proposition 1 (b) and (c) hold for *all* information structures; hence, the discussion below uses the general term of second best.

A graphic illustration of these forces is provided in Figure 1. Suppose that given P, the second-best firm chooses to produce Y_s^*. At this point $\mu_f(Y_s^*) < P$. But μ_f may have different structures. If μ_f is flat (large β), Y_f^* is much greater than Y_s^* and probably $n_f > n_s$; but if μ_f is steep (small β), Y_f^* is very close to Y_s^* and n_s may be bigger than n_f.

While in the general case one cannot predict the effect of imperfect information on the optimal size of the firm, in the current setting (IC), when the signals are drawn from a normal distribution, I show that the first-best firm is larger than the second-best firm. In this case, the increase in production for the first best dominates the decrease in effort.

Proposition 2. Consider the case of individual signals and constant variance (IC). Then, (a) the optimal size of the firm is decreasing in the variance of the signals, σ_i^2, for all $0 < \beta < 1$; and (b) for any $\sigma_i^2 > 0$, the optimal size of the first-best firm is larger than the optimal size of the second-best firm (i.e., $n_f^* > n_{IC}^*$).

Combining the results of Propositions 1 and 2, I predict that agents' effort and the number of agents are positively correlated, such that when the noise in the system increases, the principal decreases both the number of agents and the optimal effort required from each agent. This may look counterintuitive, as in many cases one observes that in small teams agents are working very

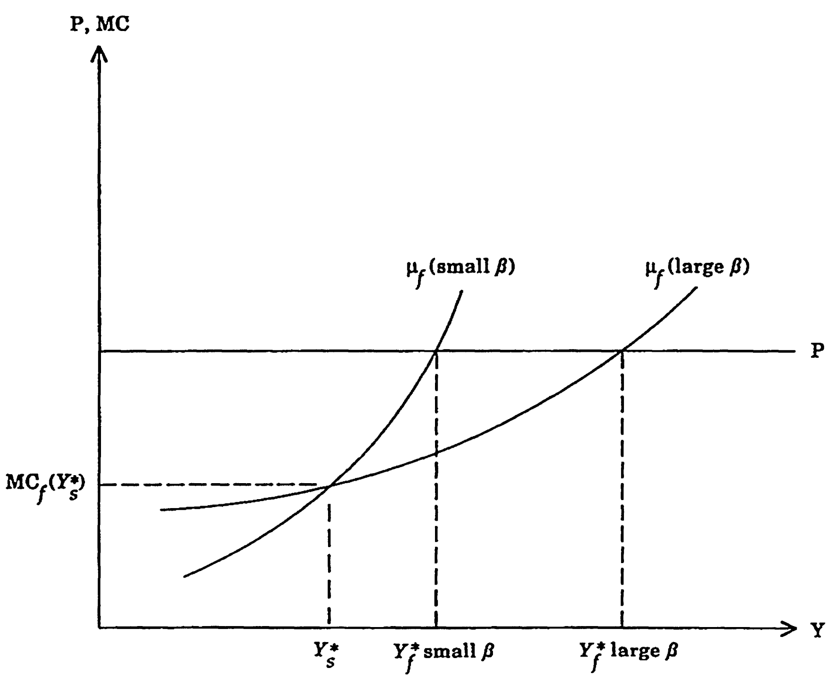

Figure 1. Relations between first- and second-best firm size.

hard while in larger teams work intensity is lower. The prediction of positive correlation is reversed when I introduce externalities among agents' performance measures.

First I do so by altering the signals' distribution. The next proposition demonstrates that when one changes the distribution to the exponential distribution, ordering of firm size may be reversed.

Proposition 3. Consider the case where the signals are exponentially distributed. Then, in some cases, the second-best firm is larger than the first-best firm.

Why should one find a different result here? For the exponential distribution the variance of the signal is increasing with effort. If $x_i \sim \exp(1/a_i)$, then $E(x_i) = a_i$ and $\mathrm{var}(x_i) = a_i^2$. This introduces an externality, in that for a given level of output, increasing firm size (and decreasing the a_i) increases the precision of the signal. This externality results in a decrease in individuals' efforts and an increase in firm size.[18]

3.2 Individual Signals, Increasing Variance (II)

A major assumption in the IC case is that the variance of each signal is independent of the number of agents employed. When signals are produced by a monitor with limited processing capacity (i.e., with a limited number of observations), this assumption is clearly unacceptable. Assume the monitoring device has a capacity of z observations, and the variance of each observation is s^2. Define $\sigma^2 = s^2/z$ to be the variance if all the observations were made on the same employee. In this case:

Lemma 1. When signals are produced by a monitor with a limited capacity, the variance for each agent's signal is increasing *linearly* with the number of agents. Formally, $V = n\sigma^2$.

This case represents *loss of control,* such that the more agents the principal supervises the less accurate is the individual signal.[19] Formally, I assume $x_i \sim N(a_i, n\sigma_i^2)$. Intuitively, one would expect in this case that the firm will prefer to increase the effort required from each agent rather than to hire more agents, since hiring will increase noise in all the signals, and not only in the signals related to the new agents.

The result is even stronger. It appears that at this level of control loss, the firm enforces the *first-best effort* level, regardless of the signal accuracy (σ_i^2). As shown in Proposition 1, higher variance implies a lower level of optimal effort; but in this case the diseconomies due to having a large number of agents affects the optimal effort level in the opposite direction. In fact, the two

18. Having a similar property for the normal distribution invalidates the proof of the first-order approach. Further, the first-order conditions do not have a closed-form solution.

19. In spirit, this case resembles Calvo and Wellisz (1978), where the larger the number of agents under a supervisor, the less accurate is his control. Calvo and Wellisz deal with the case of $\beta = 1$.

forces exactly cancel each other and one observes no change in the effort, even when σ_i^2 is changed.

The above finding may explain why even though firms have different accuracy levels in their information structures one does not observe high variability in the number of working hours (as a proxy for effort) across firms. The optimal effort level under loss of control is not sensitive to the accuracy of the performance signals.

While the optimal effort in the II case is identical to the first best, the marginal cost, of course, is not. Because the increased noise is costly to the firm, the marginal cost is higher than in the IC case or in the first-best case. I can rewrite the marginal cost function as

$$\mu_{\text{II}} = \mu_f + (2\sigma_i^2/\beta)\cdot Y^{(2-\beta)/\beta}.$$

Observe that there are two components to the production cost. The first is the actual production cost (cost of effort) and the second is the cost of information. *Information appears as an additional input in the production function.*

The marginal cost equation suggests a very steep cost structure, and the reason is clear: A large n implies very large individual variances, which increase the information cost for the firm. The above discussion is summarized in the next proposition.

Proposition 4. Consider the case of individual signals and increasing variance (II). Then, (a) the optimal effort level is identical to the first-best effort level ($a_{\text{II}}^* = a_f^*$); and (b) for any production plan the marginal cost is larger under the II case than under the IC case ($\mu_{\text{II}} > \mu_{\text{IC}} > \mu_f$) and optimal production level is such that $Y_{\text{II}}^* < Y_{\text{IC}}^* < Y_f^*$, for $n > 1$.

In comparing the firm sizes of IC and II, I find that the two forces operate in the same direction: An II firm produces less and enforces a higher effort level for each agent; hence, it needs fewer agents to meet its production assignment.

Corollary 1. The optimal size of the firm in the II case is smaller than the size in either the IC case or the first-best case ($n_{\text{II}} < n_{\text{IC}} < n_f$).

If one introduces a stronger version of loss of control—that is, $x_i \sim N(a_i, n^k\sigma_i^2)$ with $k > 1$—it can be shown that the effort level is *increasing* when the signal becomes less accurate. This result is *in contrast* with Proposition 1, where effort is inversely related to σ_i^2, and could be obtained only when the principal is free to change the number of agents.

3.3 Aggregated Signals

Information is not always as detailed as in the preceding cases. This may result from the inability to distinguish among individuals' efforts (see Holmstrom, 1982) or because of limits on the amount of information the system can process (as in Melumad, Mookherjee, and Reichelstein, 1991). Many aggregations are performed before any (accounting) report is issued. In the extreme case, the principal may be able to observe only one signal with respect to all agents, in which case the compensation scheme can only be a function of this one signal observed.

One possibility is that in the case of a single signal, its variance is independent of the number of individual signals aggregated. This may be the case if the noise comes from measurement error in evaluating the signal (e.g., if we weigh the output and the scale inaccuracy is independent of the weight level). Formally, we refer to this case as the *aggregated signal, constant variance* (AC) case, where $\Sigma x_i \sim N(\Sigma a_i, \sigma^2)$.

Alternatively, it may be that each signal has its own variance, and when signals are aggregated, one adds up variances. This is represented by the *aggregated signal, increasing variance* (AI) case where $\Sigma x_i \sim (\Sigma a_i, n\sigma^2)$.

It is clear that in general aggregated cases are inferior to the individual signal cases. In this setting, the aggregate measure is a sufficient statistic for all other signals and, therefore,

Observation 1. The aggregated signal cases AC and AI are identical to the IC and II cases, respectively.

Under the optimal contract, every agent is evaluated based on the *total* deviation from the expected value of the signal, as if that one agent were solely responsible for the deviation. This result is in line with a similar result, in a substantially different setting, of McAfee and McMillan (1986). Their result is derived by the risk neutrality of the agents. Here I deal with risk-averse agents; however, the composition of the specific risk aversion that I assumed and the distribution function imply that the induced payoff is linear in the information variable; hence, similar economic intuitions apply.

Since an agent's evaluation is based on other agents' signals (and actions), collusion among agents is possible. It can be shown (see Ziv, 1990) that in this case the equilibrium involves strictly dominant strategies; hence, the principal need not worry about implicit collusion or multiplicity of equilibria (as in Demski and Sappington, 1983; Ma, Moore, and Turnbull, 1988; and Holmstrom and Milgrom, 1990). If agents are allowed to sign binding contracts and commit to effort levels (explicit collusion, see Holmstrom and Milgrom, 1990), the equilibrium should be modified. Given the contracts introduced above, agents have incentives to overinvest in effort, since they get positive externality from other agents' effort. In this case, the principal should treat all the agents as if they were a single syndicate-agent (see Wilson, 1968).

A third possible case of aggregation involves adding up signals where each one of them has an increasing variance; hence, the total variance is $n^2\sigma_{\tilde{\imath}}^2$. In this case,

Proposition 5. Consider the case of increasing variance signals and suppose one observes only an aggregate signal (such that the total variance is $n^2\sigma_{\tilde{\imath}}^2$). Then, (a) the optimal effort is increasing in the variance $\sigma_{\tilde{\imath}}^2$, and (b) the optimal effort is larger than first-best effort ($a_{\tilde{\mathrm{AI}}} > a_f$).

The firm now produces lower quantity and requires a higher level of effort from each agent. Hence, we have the following corollary.

Corollary 2. The size of the firm in the case considered in Proposition 5 is smaller than the size of the firm in any of the cases discussed above. Further, effort and firm size are negatively correlated.

Next we use the exponential distribution to demonstrate the possibility that the optimal effort is decreasing in the firm's output.

Assume that $\Sigma x_i \sim \exp(1/\Sigma a_i)$.[20] From the first-order conditions I get $E(s(\Sigma x_i)) = (\bar{u} + a_i^2)^2/4 + (\Sigma a_i)^2 a_i^2$. This, when minimized under the production constraint, yields

$$(3 + 4n^2)a_i^4 + 2\bar{u}a_i^2 - \bar{u}^2 = 0. \tag{4}$$

With this last equation, I show that the optimal effort is always smaller than the first-best effort, and for large enough firms it is smaller than the effort in the IC case. I also show that the optimal effort is decreasing in the firm's output.

Proposition 6. Consider the case where the aggregated signal is drawn from an exponential distribution. Then, (a) optimal effort is always smaller than the first-best effort ($a^*_{\text{AC}} < a^*_f$); (b) when the number of agents employed is at least 2, optimal effort is smaller than the effort in the IC case ($a^*_{\text{AC}} < a^*_{\text{IC}}$); and (c) the optimal effort a^*_{AC} is a decreasing function of $\bar{Y}$.

The empirical prediction from Proposition 6 is that effort and firm size are negatively correlated. This prediction is consistent with the observation that in a start-up company there are few employees who are all working very hard, while more established organizations employ a larger number of workers at a lower intensity level.

The last part of the proposition comes from the fact that for any given $\bar{Y}$ the variance is given by $V = (\Sigma a_i)^2 = \bar{Y}^{2/\beta}$; in other words, the variance of the aggregated signal is given by the output and does not change with the production structure. Hence the firm will reduce the optimal individual effort to save information costs. This result is consistent with Proposition 1 (effort is a decreasing function of the signal's variance).

It is important to note that results like Proposition 6(c) may appear only in a setting where the number of agents is a decision variable. Holding the number of agents constant yields the *opposite* result (a is an increasing function of $\bar{Y}$).

20. Aggregation of signals that are coming from the exponential distribution gives a gamma distribution, which is different from the assumed distribution. Solving the model with the gamma distribution is possible only under the *assumption* that all agents exert the same level of effort. In such a case, Equation (4) below is slightly different [the coefficient of a_i^4 is $(3 + 8n)$]. The results of Proposition 6 are identical.

4. Conclusions

In this article, I illustrated the importance of analyzing the nature of the performance measure system in determining optimal firm size and compensation contracts. The analysis showed that optimal firm size depends on the nature of the information available. The factors determining firm size (in particular the trade-off between individual effort and number of agents) cannot be identified by traditional principal–agent(s) models that hold the number of agents constant. Further, single-agent models cannot deal with the variations in the information structure as discussed here. I also investigated the consequences of changes in the accuracy of a given performance measure and the effect output price has on these decisions.

To gain tractability, I made a number of simplifying assumptions. Even with these restrictive assumptions, I established that the net effect of the conflicting forces can go both ways. I showed that an increase in the signal's precision may imply an increase or a decrease in the level of optimal effort. Similarly, I showed that less detailed information may increase or decrease the size of the firm. I view the contribution of this article as the identification of factors determining firm size and in particular the role that alternative information technologies play in determining optimal firm structure.

Appendix: Proofs

I start the Appendix with a technical result. The problem as presented in Section 2 is not tractable. The reason is that in order to use standard "calculus of variations" one needs $\mathbf{x} \in R^k$, which is satisfied only when n is an integer. Once there is an integer in the program, one cannot apply Euler's theorem. To overcome this problem, I employ the following Step Optimization.[21,22]

(i) First I find the optimal structure of $s_i(\mathbf{x})$ for any given a,n. This structure should hold at the optimal solution. For example, in the second-best cases, $s_i(\mathbf{x}) = (\lambda_i + \eta_i f_{a_i}(\mathbf{x},\mathbf{a})/f(\mathbf{x},\mathbf{a}))^2$.

(ii) I then introduce a new optimization problem with an *assumed* similar structure of $s_i(\mathbf{x})$ but with different parameters—say, $s_i(\mathbf{x}) = (\alpha_i + \gamma_i f_{a_i}(\mathbf{x},\mathbf{a})/f(\mathbf{x},\mathbf{a}))^2$—and optimize with respect to those new parameters as well as a and n.[23]

Lemma 2. The Step Optimization solution is identical to the original problem solution.

Proof. Suppose not, then there exist $\hat{a}$ and $\hat{n}$ that solve the general program and have lower cost than a^* and n^* that solve the step optimization. It is clear that $\hat{a}$ and $\hat{n}$ satisfy the structure of the optimal compensation scheme as

21. The general idea is that $\max_{x,y} f(x,y) = \max_x[f^*(x)]$, where $[f^*(x)] = \max_y f(x,y)$.

22. I still need to assume that n is an integer to make the program meaningful. However, when I solve the continuous approximation of the problem, as I do below, I may have a noninteger as a solution. In this case the solution is a nearby integer.

23. Even though it turns out that $\alpha_i = \lambda_i$ and $\gamma_i = \eta_i$ (as one would expect), observe that including λ_i,η_i in the second optimization will be wrong. It is easy to see that the solution to the second problem is consistent with the general structure of $s_i(\mathbf{x})$.

identified by the first step, and hence are feasible in the second step. This contradicts the optimality of a^* and n^* in the second step. □

Proof of Proposition 1. (a) Using Equation (2), the Lagrangian becomes

$$\mathcal{L} = \int \sum_{i=1}^{n} \left\{ \left[\left(\alpha_i + \gamma_i \frac{x_i - a_i}{\sigma_i^2} \right)^2 \right.\right.$$

$$\left. - \lambda_i \left(2\alpha_i + 2_i \frac{x_i - a_i}{\sigma_i^2} - a_i^2 \right) \right] f(\mathbf{x},\mathbf{a})$$

$$\left. - \eta_i \left(2\alpha_i + 2\gamma_i \frac{x_i - a_i}{\sigma_i^2} \right) f_{a_i}(\mathbf{x},\mathbf{a}) + \lambda_i \bar{u} + 2a_i \eta_i \right\} d\mathbf{x}$$

$$- \mu \left(\left(\sum a_i \right)^{\beta} - \bar{Y} \right).$$

The derivatives with respect to a_i and n give

$$\frac{\partial \mathcal{L}}{\partial a_i} = 0 \Rightarrow 2\lambda_i a_i + 2\eta_i = \mu\beta \left(\sum a_i \right)^{\beta - 1},$$

$$\frac{\partial \mathcal{L}}{\partial n} = 0 \Rightarrow \alpha_i^2 + \frac{2\gamma_i^2}{\sigma_i^2} = \mu\beta \left(\sum a_i \right)^{\beta - 1} a_i.$$ [24]

These equations simplify to

$$3a_{\mathrm{IC}}^4 + 2a_{\mathrm{IC}}^2(\bar{u} + 2\sigma_i^2) - \bar{u}^2 = 0. \tag{A1}$$

Solving the above gives the solution as in Equation (3).

(i) I have to show $\partial a_{\mathrm{IC}}^* / \partial \sigma_i^2 < 0$:

$$\frac{\partial a_{\mathrm{IC}}^*}{\partial \sigma_i^2} = \frac{1}{6a_{\mathrm{IC}}^*} \left[-2 + \frac{2(2\sigma_i^2 + \bar{u})}{2\sqrt{\bar{u}^2 + \sigma_i^4 + \bar{u}\sigma_i^2}} \right]$$

$$= \frac{1}{2a_{\mathrm{IC}}^* \sqrt{\bar{u}^2 + \sigma_i^4 + \bar{u}\sigma_i^2}} \left[\frac{\bar{u} + 2\sigma_i^2 - 2\sqrt{\bar{u}^2 + \sigma_i^4 + \bar{u}\sigma_i^2}}{3} \right]$$

$$= \frac{-a_{\mathrm{IC}}^*}{2\sqrt{\bar{u}^2 + \sigma_i^4 + \bar{u}\sigma_i^2}} < 0.$$

(ii) By definition

$$a_{\mathrm{IC}}^* = a_f^* + \int_0^{\sigma_i^2} \frac{\partial a_{\mathrm{IC}}^*(\bar{\sigma}_i^2)}{\partial \bar{\sigma}_i^2} \, d\bar{\sigma}_i^2.$$

24. Clearly, the term $\partial\mathcal{L}/\partial n$ involves an abuse of notation. One can look at $m(n)$, which is the continuous extension of $\mathcal{L}$ with respect to n. See also note 22.

From part (i), I have $\partial a^*_{\mathrm{IC}}/\partial\sigma_i^2 < 0$ for all values of σ_i^2, so the integral is always negative.

(b) Using the first-order conditions, it is easy to show that $\mu = (E(s(\mathbf{x}))/a) \times \bar{Y}^{(1-\beta)/\beta}/\beta$. For a given $\bar{Y}$, I need to show that

$$\frac{E(s_{\mathrm{IC}}(\mathbf{x}))}{a_{\mathrm{IC}}} > \frac{E(s_f(\mathbf{x}))}{a_f}.$$

Now, as the second best is a constrained version of the first best, and the additional constraint is binding, it must be that

$$\sum_1^{n_f} E(s_f(\mathbf{x})) < \sum_1^{n_{\mathrm{IC}}} E(s_{\mathrm{IC}}(\mathbf{x})),$$

but since $(na)^\beta = \bar{Y}$, I can use $n = \bar{Y}^{1/\beta}/a_i$. Plug the last term into the above inequality to get

$$E\left(\frac{s_f(\mathbf{x})}{a_f}\, Y^{1/\beta}\right) < E\left(\frac{s_{\mathrm{IC}}(\mathbf{x})}{a_s}\, Y^{1/\beta}\right).$$

(c) Follows directly from (b). □

Proof of Proposition 2. I show that $\partial n^*_{\mathrm{IC}}/\partial\sigma_i^2 < 0$. I have

$$\frac{\partial n^*_{\mathrm{IC}}}{\partial\sigma_i^2} = P^{1/(1-\beta)}\frac{1}{1-\beta}\left[\frac{\beta}{a_{\mathrm{IC}}^{*2-\beta}(a_{\mathrm{IC}}^{*2} + \bar{u} + 2\sigma_i^2)}\right]^{[1/(1-\beta)]-1}$$

$$\times\, \beta\, \frac{-\partial(a_{\mathrm{IC}}^{*\,2-\beta}(a_{\mathrm{IC}}^{*\,2} + \bar{u} + 2\sigma_i^2))/\partial\sigma_i^2}{[a_{\mathrm{IC}}^{*\,2-\beta}(a_{\mathrm{IC}}^{*\,2} + \bar{u} + 2\sigma_i^2)]^2}$$

$$\overset{s}{=} -\frac{\partial}{\partial\sigma_i^2}(a_{\mathrm{IC}}^{*\,2-\beta}(a_{\mathrm{IC}}^{*\,2} + \bar{u} + 2\sigma_i^2))$$

$$= -a_{\mathrm{IC}}^{*\,1-\beta}\left[\frac{\partial a^*_{\mathrm{IC}}}{\partial\sigma_i^2}(2a_{\mathrm{IC}}^{*\,2} + (2-\beta)(a_{\mathrm{IC}}^{*\,2} + \bar{u} + 2\sigma_i^2)) + 2a^*_{\mathrm{IC}}\right]$$

$$\overset{s}{=} \frac{a^*_{\mathrm{IC}}}{2\sqrt{\bar{u}^2 + \sigma_i^4 + \bar{u}\sigma_i^2}}(2a_{\mathrm{IC}}^{*\,2} + (2-\beta)(a_{\mathrm{IC}}^{*\,2} + \bar{u} + 2\sigma_i^2)) - 2a^*_{\mathrm{IC}}$$

$$\le a^*_{\mathrm{IC}}\left[\frac{2a_{\mathrm{IC}}^{*\,2} + \bar{u} + 2\sigma_i^2}{\sqrt{\bar{u}^2 + \sigma_i^4 + \bar{u}\sigma_i^2}} - 2\right].$$

So it is enough to show that the term inside the brackets is negative, which when squaring (all parts positive) is negative if and only if

$$4a_{\mathrm{IC}}^{*4} + 2a_{\mathrm{IC}}^{*2}(2\bar{u} + 4\sigma_i^2) - 3\bar{u}^2 < 0.$$

Recalling Equation (A1),

$$-2a_{\mathrm{IC}}^{*4} - \bar{u}^2 < 0,$$

which is always true.

(b) $n_{\mathrm{IC}}^*(\sigma_i^2 = 0) = n_f^*$ and, with part (a),

$$n_{\mathrm{IC}}^*(\sigma_i^2) = n_{\mathrm{IC}}^*(\sigma_i^2 = 0) + \int_0^{\sigma_i^2} \frac{\partial n_{\mathrm{IC}}^*(\tilde{\sigma}_i^2)}{\partial \tilde{\sigma}_i^2}\, d\tilde{\sigma}_i^2 < n_{\mathrm{IC}}^*(\sigma_i^2 = 0) = n_f^*. \qquad \square$$

Proof of Proposition 3. I first solve the principal's problem. The Lagrangian becomes

$$\mathcal{L} = \int \sum_{i=1}^{n} \left\{ \left[\left(\alpha_i + \gamma_i \frac{x_i - a_i}{a_i^2} \right)^2 - \lambda_i \left(2\alpha_i + 2\gamma_i \frac{x_i - a_i}{a_i^2} - a_i^2 \right) \right] f(\mathbf{x},\mathbf{a}) \right.$$

$$\left. - \eta_i \left(2\alpha_i + 2\gamma_i \frac{x_i - a_i}{a_i^2} \right) f_{a_i}(\mathbf{x},\mathbf{a}) + \lambda_i \bar{u} + 2a_i\eta_i \right\} d\mathbf{x}$$

$$- \mu \left(\left(\sum a_i \right)^{\beta} - \bar{Y} \right).$$

Solve the first-order conditions to get

$$a_{\mathrm{IC}}^* = \left(\frac{\bar{u}}{5}\right)^{1/2}, \qquad \mu_{\mathrm{IC}} = \frac{2}{\sqrt{5}} \frac{\bar{u}^{3/2}}{\beta}\, \bar{Y}^{(1-\beta)/\beta},$$

$$Y_{\mathrm{IC}}^* = \left(\frac{\sqrt{5}\,\beta}{2\,\bar{u}^{3/2}}\right)^{\beta/(1-\beta)} p^{\beta/(1-\beta)}, \qquad n_{\mathrm{IC}}^* = \left(\frac{5}{\bar{u}}\right)^{1/2} \left(\frac{\sqrt{5}\,\beta}{2\,\bar{u}^{3/2}}\right)^{1/(1-\beta)} p^{1/(1-\beta)}.$$

Now,

$$\frac{n_f^*}{n_{\mathrm{IC}}^*} = \frac{3^{(4-\beta)/(2-2\beta)}}{5^{(2-\beta)/(2-2\beta)}\, 2^{1/(1-\beta)}}.$$

This term is less than 1 for small β (up to about $\beta = 0.45$) and greater than 1 for higher levels of β, providing circumstances where the second-best firm is larger than the first-best firm. $\square$

Proof of Lemma 1. Since each agent is observed the same number of times, there are z/n observations per agent. The signal used for contracting with the agent is the mean of all signals (a sufficient statistic), which has a variance of $s^2/(z/n) = n\sigma^2$. □

Proof of Proposition 4. (a) The Lagrangian becomes

$$\mathcal{L} = \int \sum_{1}^{n} \left\{ \left[\left(\alpha_i + \gamma_i \frac{x_i - a_i}{n\sigma_i^2} \right)^2 - \lambda_i \left(2\alpha_i + 2\gamma_i \frac{x_i - a_i}{n\sigma_i^2} - a_i^2 \right) \right] f(\mathbf{x},\mathbf{a}) \right.$$

$$\left. - \eta_i \left(2\alpha_i + 2\gamma_i \frac{x_i - a_i}{n\sigma_i^2} \right) f_{a_i}(\mathbf{x},\mathbf{a}) + \lambda_i \bar{u} + 2a_i\eta_i \right\} d\mathbf{x}$$

$$- \mu \left(\left(\sum a_i \right)^{\beta} - \bar{Y} \right).$$

Simple optimization gives

$$\frac{\partial \mathcal{L}}{\partial a_i} = 0 \Rightarrow 2\lambda_i a_i + 2\eta_i = \mu\beta \left(\sum a_i \right)^{\beta - 1},$$

$$\frac{\partial \mathcal{L}}{\partial n} = 0 \Rightarrow \alpha_i^2 + \frac{2\eta_i\gamma_i}{n\sigma_i^2} = \mu\beta \left(\sum a_i \right)^{\beta - 1} a_i.$$

The above and the first-order conditions with respect to α_i, γ_i, λ_i, and η_i give $a_{\text{II}}^* = (\bar{u}/3)^{1/2}$.

(b) Use the fact that

$$\mu_{\text{II}} = \frac{a_{\text{II}}^*(a_{\text{II}}^{*2} + \bar{u} + 2n\sigma_i^2)}{\beta} Y^{(1-\beta)/\beta}.$$

The first result follows from $n > 1$ and $a_{\text{II}}^* > a_{\text{IC}}^*$, while the second result follows from the first-order conditions for profit maximizaton. □

Proof of Corollary 1. From Proposition 4 I know that $Y_{\text{II}}^* < Y_{\text{IC}}^*$, so

$$\left(\sum_{1}^{n_{\text{II}}} a_{\text{II}} \right)^{\beta} < \left(\sum_{1}^{n_{\text{IC}}} a_{\text{IC}} \right)^{\beta}$$

and $a_{\text{II}} > a_{\text{IC}}$, implying $n_{\text{II}} < n_{\text{IC}}$. □

Observation 1. Careful statement of the principal's problem will give the same set of first-order conditions and hence the same solution. The details were omitted because of space considerations. □

Proof of Proposition 5. (a) In this case, the first-order conditions imply $3a_i^4 + 2a_i^2(\bar{u} - 4n\sigma_i^2) - \bar{u}^2 = 0$. Solving for a_i and taking the derivative with respect to σ_i^2 implies that a_i is increasing in σ_i^2. (b) The argument is identical to Proposition 1 (a (ii)) with positive integral. □

Proof of Proposition 6. (a) and (b) follow from the fact that $4n^2 > 0$ and $3 + 4n^2 > 15$ for all $n \geq 2$. For (c), rewrite Equation (4) with the production constraint

$$3a_i^4 + 2a_i^2(\bar{u} + 2\bar{Y}^{2/\beta}) - \bar{u}^2 = 0$$

and observe that the coefficient of a_i^2 is increasing in Y. □

References

Baiman, S. 1990. "Agency Research in Managerial Accounting: A Second Look," 15 *Accounting Organization and Society* 341–71.

Calvo, G., and S. Wellisz. 1978. "Supervision, Loss of Control, and the Optimal Size of the Firm," 86 *Journal of Political Economy* 943–52.

Coase, R. 1937. "The Nature of the Firm," 4 *Economica* 386–405.

Demski, J., and D. Sappington, 1983. "Optimal Incentive Contracts with Multiple Agents," 17 *Journal of Economic Theory* 152–71.

Holmstrom, B. 1979. "Moral Hazard and Observability," 13 *Bell Journal of Economics* 74–91.

———. 1982. "Moral Hazard in Teams," 13 *Bell Journal of Economics* 324–40.

———, and P. Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives," 55 *Econometrica* 303–28.

———, and ———. 1990. "Regulating Trade Among Agents," 146 *Journal of Institutional and Theoretical Economics* 85–105.

———, and ———. 1991. "Multitask Principal–Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design," 7(Sp) *Journal of Law, Economics, & Organization* S24–52.

Jewitt, I. 1988. "Justifying the First Order Approach to Principal–Agent Problems," 56 *Econometrica* 1177–90.

Lipsey, R. 1966. *An Introduction to Positive Economics*. 2nd ed. London: Weidenfeld and Nicolson.

Ma, C., J. Moore, and S. Turnbull. 1988. "Stopping Agents from Cheating," 46 *Journal of Economic Theory* 355–72.

McAfee, R. P., and J. McMillan. 1986. "Optimal Contracts for Teams," mimeo, University of Western Ontario.

Melumad, N., D. Mookherjee, and S. Reichelstein. 1991. "Hierarchical Decentralization of Incentive Contracts," mimeo, Stanford University.

———, and S. Reichelstein. "Value of Communication in Agencies," 47 *Journal of Economic* Theory 334–68.

Mirrlees, J. 1974. "Notes on Welfare Economics, Information and Uncertainty," pp. 243–258 in M. Balch, D. McFadden, and S. Wu, eds., *Essays in Economic Behavior Under Uncertainty*. Amsterdam: North Holland.

———. 1976. "The Optimal Structure of Incentives and Authority within an Organization," 7 *Bell Journal of Economics* 105–31.

Myerson, R. 1982. "Optimal Coordination Mechanisms in Generalized Principal–Agent Relationships," 10 *Journal of Mathematical Economics* 67–81.

Rogerson, W. 1985. "The First Order Approach to Principal–Agent Problems," 53 *Econometrica* 1357–68.

Rothschild, M., and J. Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets, An Essay on The Economics of Imperfect Information," 90 *Quarterly Journal of Economics* 629–49.

Williamson, O. 1985. *The Economic Institutions of Capitalism.* New York: Free Press.
Wilson, R. 1968. "The Theory of Syndicates," 36 *Econometrica* 119–32.
Ziv, A. 1990. "Informational Issues in Organizations." Ph.D. dissertation, Graduate School of Business, Stanford University.
———. 1992. "The Normal Distribution and Principal–Agent Models," mimeo, Yale School of Organization and Management.