WILEY

Information Technology and Optimal Firm Structure
Author(s): Amir Ziv
Source: *Journal of Accounting Research*, Vol. 38, No. 2 (Autumn, 2000), pp. 297-328
Published by: Wiley on behalf of Accounting Research Center, Booth School of Business, University of Chicago
Stable URL: https://www.jstor.org/stable/2672935
Accessed: 31-03-2020 19:31 UTC

Information Technology and Optimal Firm Structure

AMIR ZIV*

1. Introduction

In this paper I use a principal–agent framework to explore the relation between the hierarchical structure of firms and the accounting information technologies available to them. My analysis is related to that in Melumad, Mookherjee, and Reichelstein [1992] and Ziv [1993]. Melumad, Mookherjee, and Reichelstein model a principal who employs two privately informed agents and chooses either a flat structure where both agents contract and communicate with the principal, or a hierarchical structure in which the principal contracts with only one agent, who subsequently writes a subcontract with a second agent, creating a two-layer organizational form.[1] Melumad, Mookherjee, and Reichelstein use the revelation principal to prove the general superiority of the flat structure. They add exogenous restrictions on communication (with respect to dimensionality and complexity of the message space) to demonstrate a demand for hierarchy. Ziv [1993], in a moral hazard setting, solves for the optimal number of agents in a one-layer firm, under different exogenously given information structures. In this paper, I take an approach that allows the principal to choose the number of layers in the firm, the

---

[1] I define the number of organization layers to be the number of layers of agents; the principal is not counted as an organization layer.

number of agents in each layer, and the quantity and quality of information in the firm (subject to the available information technology).

This paper also complements Baiman, Larcker, and Rajan's [1995] analysis of the optimal allocation of tasks from a parent firm to its business units. The determinants of this organizational structure decision are the parent's task expertise relative to that of the business unit and the relative importance of the business unit to the performance of the parent firm. My paper considers multiple *identical* agents, who may be assigned different tasks, and allows for a multitier organizational structure.

Finally, my analysis extends Calvo and Wellisz [1978], who consider a world where the probability an employee is monitored decreases with the ratio of employees to supervisors. They show an optimal firm should consist of either one principal or an infinite number of layers of supervisors. I show that the number of layers of supervisors is finite, a more descriptive result. The different result arises because Calvo and Wellisz assume there is an upper bound on any one person's effort. This assumption precludes the principal from increasing his/her effort in order to gain more control, thereby forcing him/her to hire more and more layers of supervisors. My model, in contrast, does not impose *external* limits on the level of effort, thus allowing the principal to increase his/her monitoring effort without adding organizational layers. I therefore can examine the costs and benefits of hierarchical structure.

I find that demand for a layer of supervisors exists only for a limited set of parameters. Furthermore, only in a few extreme cases do the benefits of additional layers of supervisors outweigh the costs. Obviously, there are reasons other than supervision for firms to use hierarchical structure; examples include the motivational effects of providing a ladder for promotions, different talent levels across employees, or reducing communication burdens.[2] The results of this paper demonstrate that from an information-gathering perspective, in many cases, the required information rent associated with a hierarchical structure may outweigh its benefit, and in this respect "flatter" organizations are optimal. Structural changes in the economy that make monitoring more difficult might increase the information rent in a hierarchy (e.g., more complex production functions where supervision is not straightforward, or monitoring employees who are working at home). Hence, the analysis in this paper may help explain the recent trend toward "flatter" organizational structures.

In section 2, I introduce the basic model. In section 3, I discuss the optimal hierarchical structure of the firm. Section 4 provides a summary. Highlights of the proofs are provided in Appendix A.

---

[2] Mookherjee and Reichelstein [1997] show in a participating budgeting model that hierarchies of varying depth can be equally effective in terms of incentives and performance, and speculate that a communication burden argument can give an advantage to a multitier hierarchy.

2. The Model

Consider a risk-neutral principal who operates a firm and wishes to maximize its expected profits, Π. The firm owns a production function $g_n(\underline{a})$ where $\underline{a}$ is vector of the inputs of the firm's employees, and $n \geq 1$ is the number of employees. The mean of the firm's output, Y, which is randomly distributed, is provided by the production function. Formally, this is $E(Y) \equiv g_n(\underline{a})$. I assume the firm can replace the input (effort) of one agent with the input of another. Specifically, I make the following assumption about the production function:

ASSUMPTION 1 (Production Function). $g_n(\underline{a}) = (\Sigma_1^n a_i)^{\beta}$, where $\beta \in (0, 1)$.

Under this production function, the firm's output depends on the sum of the agents' efforts, regardless of their source. The parameter β represents the concavity of the production function.[3] The firm hires employees from a large pool of identical workers, each of whom is both risk- and (increasingly) work-averse. Each agent has a utility function $u(z, a)$, where z is the monetary compensation and a is the agent's effort. Each agent has a market alternative, $\hat{u}$, with a strictly positive certainty equivalent, denoted $\bar{u}$. For tractability I make the following assumption:

ASSUMPTION 2 (Utility Function). $u(z, a) = -\exp^{-r(z-a^{\omega})}$, where $\omega > 1$. Agents' effort is bounded from below by an arbitrary small constant, ε.

The coefficients r and ω measure the agents' risk and work aversion, respectively. I assume that any agent hired provides a strictly positive level of effort.[4] This assumption prevents the principal from hiring infinitely many agents, each of whom is doing essentially nothing.

To compare different organizational designs (for example, with respect to optimal contracts or tasks assigned), I assume that agents employed as supervisors are identical to the production agents, i.e., all the agents employed by the firm are chosen from the same pool. The principal can thus use the same employee for different tasks. Also, when the principal or any supervisor exerts monitoring effort, his/her personal costs are identical to a production agent's costs of effort, i.e., a^{ω}.

Agents are subject to moral hazard. Their effort is not observable and therefore cannot be contracted upon. Thus, the principal must use

---

[3] A slightly more general case involves a Cobb-Douglas production function, $g_n(\underline{a}) = (\Sigma \ a_i)^{\beta_1} n^{\beta_2}$, where there may exist mutual disturbance among agents ($\beta_2 < 0$), or agents' efforts are not perfect substitutes. Here, the more agents there are, the less productive they are, so the same level of total effort produces less. I elected not to use this production function because an important focus of this paper is the information-related trade-off between effort and organization size. The production function I use is neutral to this decision. It is easy to show the qualitative nature of the results is not affected by generalization to a Cobb-Douglas production function. In n. 10, I provide an example for a solution under a Cobb-Douglas production function.

[4] I use this technical assumption only in the proof of Claim 2 in Theorem 1.

compensation schemes based on performance measures, like signals produced by the accounting system, to motivate the agents. When accounting signals are generated by supervisors, it seems reasonable that more supervision (either in terms of the number of supervisors or in terms of their total effort) increases the precision of the signals generated. I assume the set of available information (signals) is normally distributed, specifically, $\underline{x} \sim N(\underline{a}, \Sigma)$. The covariance matrix, Σ, is a function of the information technology available to the firm. Specifically, it is a function of monitoring inefficiency, k, the impact of control reduction, θ, and returns on supervision effort, δ. I assume $x_i \sim N(a_i, n^\theta \frac{k}{e^\delta})$, where e is the total effort employed by all supervisors (monitors), $k \geq 0$ reflects the inefficiency of the monitoring system (for $k = 0$ monitoring is perfect, while for $k \to \infty$ monitoring provides no useful information), and $\delta > 0$ captures the returns on monitoring effort. To compare the owner's monitoring with alternative monitoring arrangements, I assume the same monitoring technology is available to all monitors. Later, I allow for different monitoring inefficiencies for production, k, and supervision, k_s, activities.

Finally, n^θ captures the possibility that the precision of each agent's signal decreases with the number of agents observed, even when total input is unchanged. The impact of the reduced control on signals' precision is captured by $\theta \geq 0$. When $\theta = 0$, there is no reduced control (and the precision of the signal is independent of number of agents employed); as θ increases, so does the reduction in control due to the presence of multiple agents. In some cases, it is possible to *derive* the value of θ from the properties of the accounting information system. For example, consider a principal whose capacity is q identically, independently distributed (*iid*) observations, each with variance s^2. Now, vary the number of agents in the firm, n. Increasing the number of agents reduces the average number of observations per agent, implying that the variance of the mean of each agent's observation is higher or the signal is less accurate. Formally, I show that holding the monitoring capacity constant at q observations, the variance of the signals in the firm increases *linearly* with the number of the agents; hence, $\theta = 1$.

OBSERVATION 1.   When the principal has limited monitoring capacity of q observations, the variance of each agent's signal, V, increases linearly with the number of agents, i.e., $\theta = 1$, or $V = n\sigma^2$, where $\sigma^2 = \frac{s^2}{q}$.

The discussion regarding the distribution of the signals is summarized by the following:

ASSUMPTION 3.   Signals are generated by the normal distribution; specifically, $\underline{x} \sim N(\underline{a}, \Sigma)$, where Σ is a diagonal matrix and each of its elements equals $n^\theta \frac{k}{e^\delta}$.

Finally, I consider only linear compensation rules. This assumption has become standard in the literature and can be justified descriptively (see, for example, Demski and Dye [1999]) and, in certain settings, theoretically (see Holmstrom and Milgrom [1987; 1991]).

ASSUMPTION 4 (Compensation Contracts).   The principal offers each agent a *linear contract* $s_i(\underline{x}) = \gamma + \underline{\alpha}^T \underline{x}$.

As long as there is no correlation on different agents' signals, the principal contracts individually with each agent and offers him/her a contract that depends *only* on his/her own signals. The optimal contract does not depend on other agents' signals.[5] Given his/her contract, each agent selects an optimal level of effort.

To summarize, the time line for a one-layer firm is as follows. The principal chooses the number of agents to employ, n, and offers each the linear contract, $s_i(\underline{x})$. Given their contracts, agents simultaneously choose their effort. Next, the accounting system generates a report (the set of signals $\underline{x}$) which is observed by the principal and all agents. The precision of the report depends on the monitoring system efficiency and on total supervision effort. The compensation paid to the employees depends on these signals.[6] Finally, output and the resulting payoffs are realized.

I conclude this section by describing the principal's problem. The principal chooses the expected production level, the number of agents employed, their effort, and the contracts that induce this level of effort in order to maximize his/her expected profits. In the current setting, however, the problem of optimal expected production level is separable from that of efficient production. In particular, I initially focus on the problem of cost minimization for a given level of expected production and then solve for optimal expected production level.

---

[5] When the covariance matrix is not diagonal, one can transpose it into a diagonal matrix. However, even though all covariances of the transposed matrix are zero, the principal needs *all* signals in contracting with each individual agent. I thank Anil Arya for suggesting this explanation.

[6] Another source of information may be the realization of the output, Y. Inclusion of Y in contracts may provide further information on the agents' performance and reduces the information cost for the principal. This complicates the discussion with no effect on qualitative results. Either of the following assumptions will eliminate these complications: (1) 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. (2) The principal cannot contract on Y with the agents. This may happen if Y is realized after the payment of compensation, or if the principal himself is subject to moral hazard on the reporting of Y, and Y is unobservable to individual workers. (See Williamson [1985, p. 139] for a discussion of a related case.)

PROGRAM 0.   Minimization of Production Costs.[7]

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

subject to: $\forall_i = 1, \ldots, n$

    Individual Rationality (IR): $\int u_i(s_i(\underline{x}), a_i)\, f(\underline{x},\ \underline{a})\, d\underline{x} \geq \hat{u}$

    Incentive Compatibility (IC): $\displaystyle \int \frac{\partial u_i(s_i(\underline{x}), a_i)}{\partial a_i} f_{a_i}(\underline{x},\ \underline{a})\, d\underline{x} = 0$

    Production: $g_n(\underline{a}) \geq \overline{Y}.$

Program 0 includes three constraints. The first is that of individual rationality (participation). To hire agents, the firm must provide expected utility at least as high as their outside opportunities. The second constraint is incentive compatibility. When agents privately choose their effort, the effort level designed by the principal must be part of each agent's best response set; otherwise, the agent will choose a different action. The third constraint is the level of expected production.

Solving Program 0, and throughout the paper, I treat the number of agents, n, as a continuous variable, thereby avoiding the technical difficulty of solving for an optimal integer. A possible interpretation could be that the principal may hire at most one part-time agent.[8]

As a benchmark, I present the solution to the first-best case, where the incentive compatibility constraint does not exist:[9]

$$n = \left(\frac{\omega-1}{\overline{u}}\right)^{1/\omega} \overline{Y}^{\frac{1}{\beta}}, \ a_i = \left(\frac{\overline{u}}{\omega-1}\right)^{1/\omega}, \ s_i(\underline{x}) = \frac{\omega\overline{u}}{\omega-1}, \text{ and}$$

$$TC = \left[\omega\left(\frac{\overline{u}}{\omega-1}\right)^{\frac{\omega-1}{\omega}}\right] Y^{\frac{1}{\beta}}.^{10}$$

---

[7] I use the local IC constraint. As pointed out by Mirrlees [1999], this being a weaker constraint may be incorrect. Jewitt [1988] provides a set of sufficient conditions for the first-order approach to be valid in a single-agent setting. These conditions are satisfied for a broad set of distributions, including the exponential family. For all cases discussed in this paper, it is also possible to show directly that the first-order approach is valid.

[8] If one restricts the number of agents, n, to be an integer, one should use a continuous extension of n in the program. Otherwise, the first-order condition with respect to n involves abuse of notation. When solving this continuous approximation of the problem, a noninteger may be the solution. In this case, under certain regulatory assumptions, the solution is a nearby integer.

[9] The first-best is a special case of Proposition 1, where $k = 0$. Hence, its derivation is omitted.

[10] When the production function is $g_n(\underline{a}) = (\Sigma a_i)^{\beta_1} n^{\beta_2}$, the optimal effort is $a_i = \left(\dfrac{\overline{u}\beta_1}{\beta_1(\omega-1)+\beta_2\omega}\right)^{1/\omega}$ and the optimal compensation is $s_i(\underline{x}) = \dfrac{(\beta_1+\beta_2)\,\omega\overline{u}}{(\beta_1+\beta_2)\,\omega-\beta_1}$. It is easy to see that qualitative results are not affected.

The compensation is independent of the signal's realization; i.e., signals are not used for the first-best contracting (as in Holmstrom [1979]). This is because the principal is interested in the agent's input and not in the realization of $\underline{x}$. Optimal risk sharing imposes all the risk on the (risk-neutral) principal with constant compensation for the agents. Also, the optimal effort level is independent of the expected output $\bar{Y}$.[11] Simple economic intuition underlies this result. Since inputs are perfect substitutes, the principal may change either the number of agents or the effort induced from each agent when s/he wants to change expected production level. Changing the number of agents, n, has the expected fixed cost $E(s_i(\underline{x}))$ per agent, while the cost of changing each agent's effort, a, is increasing ($\omega > 1$). Hence, once an agent reaches the optimal level of effort, it is less costly to increase the number of agents than to increase that agent's effort.

This discussion demonstrates that treating the number of agents as given, as is often done in agency models, may lead to erroneous conclusions. For example, if there is a change in market conditions, like an increase in the competitive output price, and the number of agents is exogenous, then the response of the principal is to increase the effort level required from each agent. This paper shows that the principal's optimal response may be to change the number of employees rather than their effort level.

In the next section I solve the principal's problem, taking into account supervision costs, when the hierarchical structure of the firm is altered.

3. The Firm's Hierarchical Structure

This section deals with the optimal hierarchical structure of the firm. I begin with the case where the principal conducts all necessary monitoring (a one-layer firm). Then, I analyze the case where the principal hires and monitors supervisors, who in turn monitor the production agents (a two-layer firm). I solve the principal's optimization program for each of these two organization designs, compare the two solutions, and find conditions under which the principal prefers, for a given level of expected output, to hire supervisors. In analyzing the two-layer firm I distinguish between nonstrategic and strategic supervisors (who work under conditions of moral hazard), and demonstrate the effect of these two types of supervision on the induced organization structure. Finally, I demonstrate the organization design choice as a function of a competitive output market price.

_____

[11] This independence holds only if the number of agents is a continuous variable. If the number of agents must be an integer, the principal may need to adjust slightly the optimal effort when the change in the expected output, $\bar{Y}$, does not exactly correspond to an integer change in the number of agents (e.g., s/he needs to hire "half" an additional agent). When the principal optimally hires many agents, this change in the optimal effort is negligible. A similar comment applies for other comparative statics, throughout the paper.

3.1 A ONE-LAYER FIRM

Consider a one-layer organization where the principal conducts all monitoring. Denote the principal's monitoring effort by a_p, and recall that the principal's cost of effort is identical to the agents' cost, i.e., a_p^ω.[12] The principal minimizes the total cost of production and monitoring:

PROGRAM 1.[13]

$$\text{Min}_{\underline{a},a_p,n} \sum_{i=1}^{n} \left(\bar{u} + a_i^\omega + \frac{rk\omega^2 a_i^{2(\omega-1)} n^\theta}{2a_p^\delta} \right) + a_p^\omega,$$

subject to: $(\sum_1^n a_i)^\beta \geq \bar{Y}$ [Production].

The solution to Program 1 is characterized below.

PROPOSITION 1.   If the principal conducts the monitoring, then: (i) the optimal number of production workers is: $n = \dfrac{1}{a} \bar{Y}^{\frac{1}{\beta}}$, where the optimal effort of the production workers, a, is determined by:

$$\bar{u} = (\omega - 1)a^\omega + (2\omega - \theta - 3) \left[\delta^{-\delta} (\frac{1}{2} \, r \, k)^\omega \omega^{2\omega+\delta} a^{\omega(2\omega-2-\theta)+\delta} \bar{Y}^{-\frac{\theta\omega-\delta}{\beta}} \right]^{\frac{1}{\omega+\delta}}, \ (1)$$

and the optimal effort of the principal, a_p, is determined by:

$$a_p = \left[\frac{1}{2} \, \delta \, r \, k \, \omega \, a^{2\omega-\theta-3} \bar{Y}^{\frac{\theta+1}{\beta}} \right]^{\frac{1}{\omega+\delta}}. \tag{2}$$

(ii) The total costs of production are:[14]

$$TC_1(\bar{Y}) = (\bar{u} + a^\omega)\frac{1}{a} \bar{Y}^{\frac{1}{\beta}} + \left(1 + \frac{\omega}{\delta}\right)\left(\frac{1}{2} \, \delta \, r \, k \, \omega \, a^{2\omega-\theta-3} \bar{Y}^{\frac{\theta+1}{\beta}}\right)^{\frac{\omega}{\omega+\delta}}. \tag{3}$$

While I do not obtain a closed-form solution for the optimal number of agents or for the optimal effort of the production agents (equation (1)), I do provide comparative statics for the impact of different parameters on the optimal solution. Given any expected production level, $\bar{Y}$, the principal must decide first how to efficiently produce $\bar{Y}$ (i.e., what com-

---

[12] I assume that the principal does not operate under moral hazard. This could be the case, for example, if the principal's monitoring effort corresponds to the installment of a monitoring system, which is observed by the agents *before* they take their action. Alternatively, given that the principal is risk-neutral, and assuming s/he has sufficient wealth, it is possible the principal commits to some level of noise (variance) in his/her signals and is penalized (using an optimal insurance mechanism, not modeled here) for deviations. Note that as I compare the cases below, where the principal's effort plays a similar role, the solution is not biased toward any of the cases.

[13] The derivation of Program 1 appears in Appendix A.

[14] The subscripts used for cost functions correspond to the number of layers in the organization. Here, there is one layer.

binations of number of agents and individual efforts to choose) and, second, how much information to generate, by exerting supervision effort. Both decisions are related to the risk premium the principal has to pay to motivate the agents. Formally:

COROLLARY 1.   When the principal monitors the agents, and for a given expected production level, $\bar{Y}$: (i) The comparative statics for the optimal number of agents, n, have the opposite sign from those for the optimal effort of the agents. (ii) When $\theta < 2\omega - 3$ ($\theta > 2\omega - 3$), the optimal effort of the production agents is decreasing (increasing) in the inefficiency of the monitoring system, k. When $(2\omega - \theta - 3)(\omega\theta - \delta) > 0$ (< 0), the optimal level of the production agents' effort is decreasing (increasing) in the expected level of production, $\bar{Y}$. When $\theta = 2\omega - 3$, the optimal effort of the agents is equal to the first-best effort and is not affected by k or by $\bar{Y}$. The optimal effort of the production agents is always increasing in the agents' certainty equivalent, $\bar{u}$, and in the impact of reduced control, θ. (iii) The optimal effort of the principal is always increasing in k and in $\bar{Y}$. When $\theta \leq 2\omega - 3$ ($\theta \geq 2\omega - 3$) the optimal effort of the principal is increasing (decreasing) in $\bar{u}$.

The intuition behind the result regarding changes in the inefficiency of the monitoring system, k, is as follows. Suppose the firm expects to produce $\bar{Y}$ units, and assume the monitoring system becomes less efficient (i.e., k increases), so signals are less precise. Usually, higher variance implies a lower level of optimal agents' effort, which increases the demand for employees. But increasing the number of agents generates a negative externality—an additional decrease in the overall precision of the signals. To counteract this effect, the principal would want to increase each agent's optimal effort, not the number of employees. The direction of the overall change in the optimal effort and in the number of agents is determined by the relative magnitudes of agents' work aversion, ω (which creates the costs of increasing the optimal effort), and the impact of reduced control, θ (which creates the costs of hiring more agents to do the same work). The effects of work aversion and reduced control exactly offset each other when $\theta = 2\omega - 3$; here, the optimal effort equals the first-best effort level and is not affected by any of the problem parameters.[15] When $\theta > 2\omega - 3$, the optimal effort exceeds the first-best effort and is *increasing* in the inefficiency of the monitoring system, k. This result is in contrast to most traditional agency results, where the second-best optimal effort is lower than the first-best effort and is decreasing in the noise in the system.

The impact of changes in the expected production level, $\bar{Y}$, on the production agents' effort is identical to that of changes in the inefficiency of the monitoring system, k, if $\omega\theta - \delta > 0$, and is reversed when $\omega\theta$

---

[15] To see why the term $2\omega - \theta - 3$ is key, replace the effort in the risk premium term by the solution to the production constraint. Then, instead of $a^{2(\omega-1)}n^{\theta+1}$ we have $n^{\theta+3-2\omega}\,\bar{Y}^{\frac{2(\omega-1)}{\beta}}$.

$-\delta < 0$. Changes in $\bar{Y}$ have a direct impact on the agents' effort that is identical to the impact of changes in k, discussed above. However, an offsetting effect is the impact of the change in the *principal's* effort, induced by the change in the expected production level. This principal's effort effect dominates the direct effect when the returns on monitoring effort are high relative to the principal's work aversion and to the impact of the reduced control, specifically when $\delta > \omega\theta$.

As expected, the principal's effort is increasing in both k and $\bar{Y}$. The principal's and the production agents' effort levels move either in the same direction (when $2\omega - \theta - 3 < 0$) or in the opposite direction (when $2\omega - \theta - 3 > 0$), as a response to changes in k. The reason is that the principal is the only monitor and cannot share the monitoring effort.

Finally, total production costs (equation (3)) are separable in the costs of hiring agents and compensating them for their certainty equivalent and their exerted effort, and the private information costs.

3.2 A TWO-LAYER FIRM

In a two-layer organization, the principal delegates monitoring of the production workers to a layer of supervisors, which s/he, in turn, monitors. The principal must now consider supervisors' incentives as well as those of production workers. Strategic supervisors must be monitored; hence, an optimal firm structure may have multiple layers of supervisors. The top level of supervisors must of course be monitored by the principal.[16]

The compensation of the production workers is based on the signals produced by the supervisors, $x_i \sim N\left(a_i, \dfrac{n^\theta k}{\left(\sum_{j=1}^m a_{sj}\right)^\delta}\right)$ where $m \geq 1$ is the number of supervisors hired and a_{sj} is the effort of the jth supervisor. The supervisors' compensation depends on the signals produced by the principal, $x_{sj} \sim N\left(a_{sj}, \dfrac{m^\theta k_s}{(a_p)^\delta}\right)$.[17] The accuracy of the signals is a function of the monitor's (supervisors' or principal's) effort. Since monitoring production and monitoring supervision could necessitate different activities, and to simplify the nonstrategic supervision analysis below, I

---

[16] The principal must be able to commit to his/her effort level. Assuming no collusion or renegotiation, the analysis is not affected by the sequence of action of the supervisors and the production agents. In particular, the results are identical if the supervisors exert their effort prior to, simultaneously with, or after the production agents take their action. The reason is that supervisors' compensation is a function of their effort and the principal's effort and not of the production agents' effort. Hence, their best response function is not sensitive to the production agents' activities.

[17] The possibility of additional information relevant for contracting with the supervisors is discussed in section 3.4.

allow for different inefficiencies for the two activities. In particular, k and k_s represent the monitoring inefficiency of production and monitoring activities, respectively.

In designing contracts with the production workers and the supervisors under conditions of moral hazard, the principal minimizes costs (equivalently, maximizes total surplus) for any level of expected output. The minimization problem is:[18]

PROGRAM 2.

$$\text{Min}_{\underline{a},\underline{a}_s,a_p,n,m} \sum_{i=1}^{n} \left(\bar{u} + a_i^{\omega} + \frac{rk\omega^2 a_i^{2(\omega-1)} n^{\theta}}{2 \left(\sum_{j=1}^{m} a_s \right)^{\delta}} \right) + \sum_{j=1}^{m} \left(\bar{u} + a_{sj}^{\omega} \right.$$

$$+ \frac{rk_s \omega^2 a_{sj}^{2(\omega-1)} m^{\theta}}{2 a_p^{\delta}} \right) + a_p^{\omega},$$

subject to: $(\sum_{1}^{n} a_i)^{\beta} \geq \bar{Y}$ [Production].

The solution to Program 2 is characterized below.

PROPOSITION 2.   Suppose the principal hires one layer of supervisors. Then: (i) when $2\omega - \theta - 3 \neq 0$, the optimal number of production agents is $n = \frac{1}{a} \bar{Y}^{\frac{1}{\beta}}$, and the optimal number of supervisors, m, is $m =$

$$\left[\frac{[\bar{u} - (\omega - 1) a_s^{\omega}]^{\omega + \delta} \delta^{\delta}}{(2\omega - \theta - 3)^{\omega + \delta} \left(\frac{1}{2} rk_s \right)^{\omega} \omega^{2\omega + \delta} a_s^{2\omega(\omega-1)}} \right]^{\frac{1}{\omega\theta - \delta}}, \text{where the implicit solu-}$$

tion for the optimal effort of the supervisors, a_s, and of the production agents, a, is given by:

$$a_s = \left[\frac{(2\omega - \theta - 3) rk\omega^2 a^{2(\omega-1)} \bar{Y}^{-\frac{\theta}{\beta}}}{2 [\bar{u} - (\omega - 1) a^{\omega}] m^{\delta}} \right]^{\frac{1}{\delta}}$$

$$a = \left[\frac{m^{\delta+1} a_s^{\delta} [2(\omega-1)\bar{u} + a_s^{\omega} (\omega - \theta\omega - 2)]}{(2\omega - \theta - 3) \frac{1}{2}\delta rk\omega^2 \bar{Y}^{\frac{\theta+1}{\beta}}} \right]^{\frac{1}{2\omega - \theta - 3}}.$$

[18] An implicit assumption here is that each supervisor monitors all agents. As shown in Baldenius, Melumad, and Ziv [2000], as long as supervisors' observations are uncorrelated and are not aggregated, any arbitrary deterministic assignment of the monitoring tasks that involves the same number of observations about each agent results in the same costs of inducing agents' effort.

(ii) When $2\omega - \theta - 3 = 0$, the optimal number of production agents is $n = \dfrac{1}{a} \bar{Y}^{\frac{1}{\beta}}$, the optimal number of supervisors is determined by:

$$\left(\frac{\omega}{\omega-1}\right)\bar{u} + (\theta+1)\left[\frac{(rk_s)^\omega\,\omega^{2\omega+\delta}\bar{u}^{2(\omega-1)}\,m^{\theta\omega-\delta}}{2^\omega\delta^\delta(\omega-1)^{2(\omega-1)}}\right]^{\frac{1}{\omega+\delta}}$$

$$= \frac{\delta rk\omega^2\bar{u}^{\frac{2\omega-\theta-\delta-3}{\omega}}\bar{Y}^{\frac{\theta+1}{\beta}}}{2(\omega-1)^{\frac{2\omega-\theta-\delta-3}{\omega}}m^{\delta+1}},$$

and the optimal effort of the supervisors and the production agents is identical, $a = a_s = \left[\dfrac{\bar{u}}{\omega-1}\right]^{\frac{1}{\omega}}$. ($iii$) The optimal effort of the principal is

$$a_p = \left[\frac{1}{2}\,rk_s\omega\,a_s^{2(\omega-1)}m^{\theta+1}\right]^{\frac{1}{\omega+\delta}}.$$

While Proposition 2 characterizes the optimal two-layer organization, it does not have an explicit form, and direct analysis is impossible. Later, I use a numerical example in order to gain more insights.

A special case of Proposition 2 occurs when supervisors' effort is either directly observed and could be contracted on, or when the monitoring system allows for a precise inference of supervisors' effort, i.e., $k_s = 0$. I refer to this case as the nonstrategic supervisors or the monitors case. It is clear the principal will hire at most one layer of monitors and will exert no effort supervising them. The solution to the principal's problem is:

PROPOSITION 3.   When the principal hires nonstrategic supervisors, or when $k_s = 0$, then: (i) The optimal number of production workers is: $n = \dfrac{1}{a}\bar{Y}^{\frac{1}{\beta}}$, where a is the optimal effort of the production workers determined by:

$$\bar{u} = (\omega-1)a^\omega + (2\omega-\theta-3)\left[\frac{rk\omega^{2+\delta}\bar{u}^{\frac{(\omega-1)\delta}{\omega}}a^{(2\omega+\delta-\theta-2)}\bar{Y}^{\frac{\theta-\delta}{\beta}}}{2(\omega-1)^{\frac{(\omega-1)\delta}{\omega}}\delta^\delta}\right]^{\frac{1}{1+\delta}}. \quad (4)$$

(ii) The optimal number of monitors is $m = \left[\dfrac{\delta rk\omega\,(\omega-1)^{\frac{\omega+\delta}{\omega}}\bar{Y}^{\frac{2(\omega-1)}{\beta}}}{2\bar{u}^{\frac{\omega+\delta}{\omega}}}\right]^{\frac{1}{1+\delta}}$,

and the optimal effort of the monitors is $a_s = \left(\dfrac{\bar{u}}{\omega-1}\right)^{\frac{1}{\omega}}$.

Examining Proposition 3, we can show that most of the results of Corollary 1 remain intact. Below, however, I consider cases where the results are different from those of Corollary 1, and some cases specific to the setting of Proposition 3. It is easy to see the production agents' effort may be higher (when $\theta > 2\omega - 3$) or lower (when $\theta < 2\omega - 3$) than that of the monitors. The changes in the production workers' effort that correspond to a change in expected output, $\bar{Y}$, are similar to those reported in Corollary 1 except that the term $\theta - \delta$ replaces the term $\omega\theta - \delta$ in Corollary 1. The reason is that, given monitors, a principal who needs more supervision effort can hire more supervisors rather than increasing his/her individual effort. As a result, the principal's effort aversion parameter, ω, disappears.

Finally, I consider monitoring intensity, as measured by the ratio of monitors to production agents, $\frac{m}{n}$.

COROLLARY 2.   Assume nonstrategic supervisors. Then, the intensity of monitoring, $\frac{m}{n}$, is increasing in monitoring inefficiency, k, and is increasing (decreasing) in the expected output, $\bar{Y}$, when $\delta < \theta$ ($\delta > \theta$). An increase in the agents' certainty equivalent may either increase or decrease monitoring intensity.

Intuitively, when the desired expected output increases, and the impact of reduced control on signals' precision is relatively high, $\delta < \theta$, hiring more agents is costly. The principal increases the number of monitors per production agent, effectively reducing the cost of a unit of production agent effort.

A special case that allows for a closed-form solution occurs when work aversion, the impact of reduced control, and the returns on monitoring are exactly offsetting. Here, the agents' optimal effort equals the first-best effort and does not change when the program parameters change. The principal alters the number of agents as a response to such parameter changes. For example, this happens when $\omega = 2$, $\theta = 1$, and $\delta = 2$. These parameters do not represent a knife-edge case; all terms reported in Propositions 1–3, and their derivatives, are continuous in the neighborhood of this example. I first use this example to compare the strategic and nonstrategic supervisors cases, and then to analyze potential differences in compensation between supervisors and agents.

COROLLARY 3.   Suppose the principal hires one layer of strategic supervisors, $\omega = 2$, $\theta = 1$, and $\delta = 2$. Then: (i) the optimal number of production workers is $n = (\bar{u})^{-\frac{1}{2}} \, \bar{Y}^{\frac{1}{\beta}}$, and the optimal number of supervisors is

$$m = \left[\frac{(2rk) \, \bar{Y}^{\frac{2}{\beta}}}{\left(\bar{u}^{\frac{1}{2}} + (2rk_s)^{\frac{1}{2}} \right)} \right]^{\frac{1}{3}} \bar{u}^{-\frac{1}{2}}. \; (ii) \text{ The optimal effort of each production}$$

worker is identical to that of each supervisor: $a = a_s = (\bar{u})^{\frac{1}{2}}$, and the opti-

mal effort of the principal is $a_p = \dfrac{(2rk_s)^{\frac{5}{12}}\, \bar{Y}^{\frac{1}{3\beta}}}{\left[\bar{u}^{\frac{1}{2}} + (2rk_s)^{\frac{1}{2}}\right]^{\frac{1}{6}}}$. (*iii*) The total costs of

production are:

$$TC_{2Example}(\bar{Y}) \;=\; \left[2\bar{u}^{\frac{1}{2}} + 3(2\,r\,k)^{\frac{1}{3}}\left[\bar{u}^{\frac{1}{2}} + (2\,r\,k_s)^{\frac{1}{2}}\right]^{\frac{2}{3}}\bar{Y}^{-\frac{1}{3\beta}}\right]\bar{Y}^{\frac{1}{\beta}}. \quad (5)$$

(*iv*) When k_s increases, i.e., supervisors become "more strategic," the principal does not change the number of agents, n; decreases the number of supervisors, m; decreases the intensity of production agents' supervision, $\frac{m}{n}$; does not change the level of effort required from the production agents, a, and the supervisors, a_s; and personally exerts more effort, a_p. Furthermore, both the marginal and the total costs of production are increasing in k_s.

The comparative statics for the strategic and the nonstrategic supervisors cases are as expected. When supervisors are strategic, the principal increases his/her monitoring effort and decreases the intensity of supervision of the production agents.

Under the conditions stipulated in Corollary 3, equal effort is induced from supervisors and production agents. However, although supervisors and production agents are ex ante identical and exert the same effort, their expected compensation need not be the same. The expected compensation also depends on the risk imposed on an agent in his/her contract and on the relative inefficiencies of the monitoring systems. Controlling for the monitoring inefficiency effect, by assuming $k = k_s$, the group whose contracts involve, optimally, more risk will receive more compensation. I show that for any level of monitoring inefficiency, $k = k_s$, a supervisor earns more than a production agent when the desired expected output, $\bar{Y}$, exceeds a certain threshold. Furthermore, when it is optimal to hire supervisors (i.e., when the conditions stipulated in Corollary 4 are satisfied), the desired expected output exceeds this threshold. Hence, when supervisors are hired, they earn more expected compensation than production workers, a result consistent with the casual observation that managers earn more than their subordinates.

Why would the principal want to impose more risk on supervisors? Intuitively, the risk imposed on the supervisors is a function of the principal's effort. Since the principal cannot share the monitoring of the supervisors, s/he trades off his/her personal effort with additional risk imposed on them. On the other hand, the risk imposed on the production agents can be reduced by hiring additional supervisors.

OBSERVATION 2.   Assume $\omega = 2$, $\theta = 1$, $\delta = 2$, $k = k_s$, and that it is optimal to hire supervisors. Then, the expected compensation paid to a supervisor exceeds the expected compensation paid to a production worker.

While Observation 2 relates only to one example, it should be noted that supervisor compensation *strictly* exceeds that of a production agent; hence, the result also holds in the neighborhood of the stipulated conditions (by invoking a continuity argument). Allowing for different monitoring inefficiencies, Observation 2 holds for any $k < k_s$. On the other hand, if $k_s < k$, production agents might earn more than the supervisors. In particular, nonstrategic supervisors, exerting the same effort as the production agents, will always earn less than production agents.

3.3 OPTIMAL FIRM STRUCTURE

So far, I have analyzed a one-layer organization (Proposition 1) and a two-layer organization (Proposition 2). When designing the firm, the principal chooses an organizational form that maximizes profits. Below, I discuss the conditions under which the principal opts for a two-layer organization, for a given level of expected output. For this set of parameters, a demand exists for supervisors.

The costs of hiring supervisors includes compensation for both their monitoring effort and the risk imposed by their contracts. However, supervisors improve on the monitoring efficiency of the production agents, by exerting more effort than the principal would on his/her own, hence increasing the precision of the agents' signals.[19] The benefits of hiring supervisors are also increasing in the desired expected output.

I use the inefficiencies of the monitoring system, k and k_s, the desired expected output, $\overline{Y}$, and the agents' certainty equivalent, $\overline{u}$, to characterize the demand for supervisors. I distinguish between two scenarios. Under the first, inefficiencies in monitoring production activities and supervision activities are positively correlated. In particular, I assume $k_s = \gamma k$, where $\gamma > 0$. Under the second scenario, I consider changes in monitoring efficiencies for the production agents, for a *given* inefficiency in monitoring supervision activities. The nonstrategic supervisors analysis represents a special case in which $k_s = 0$.

THEOREM 1.   (i) Consider a given level of expected output, and assume $k_s = \gamma k$, $\gamma > 0$. Then, the principal will not hire any supervisors when the level of monitoring inefficiency, k, is either sufficiently small or sufficiently large and the agents' certainty equivalent, $\overline{u}$, is sufficiently large. For any set of parameters and desired expected output, there exist some levels (midrange) of k, and sufficiently small $\overline{u}$, such that the principal hires at least one layer of supervisors.

(ii) *Consider a given level of expected output, and fix k_s. Then, the* principal will not hire any supervisors when the level of monitoring inefficiency, k, is sufficiently small and the agents' certainty equivalent, $\overline{u}$, is sufficiently large. For any set of parameters and desired expected
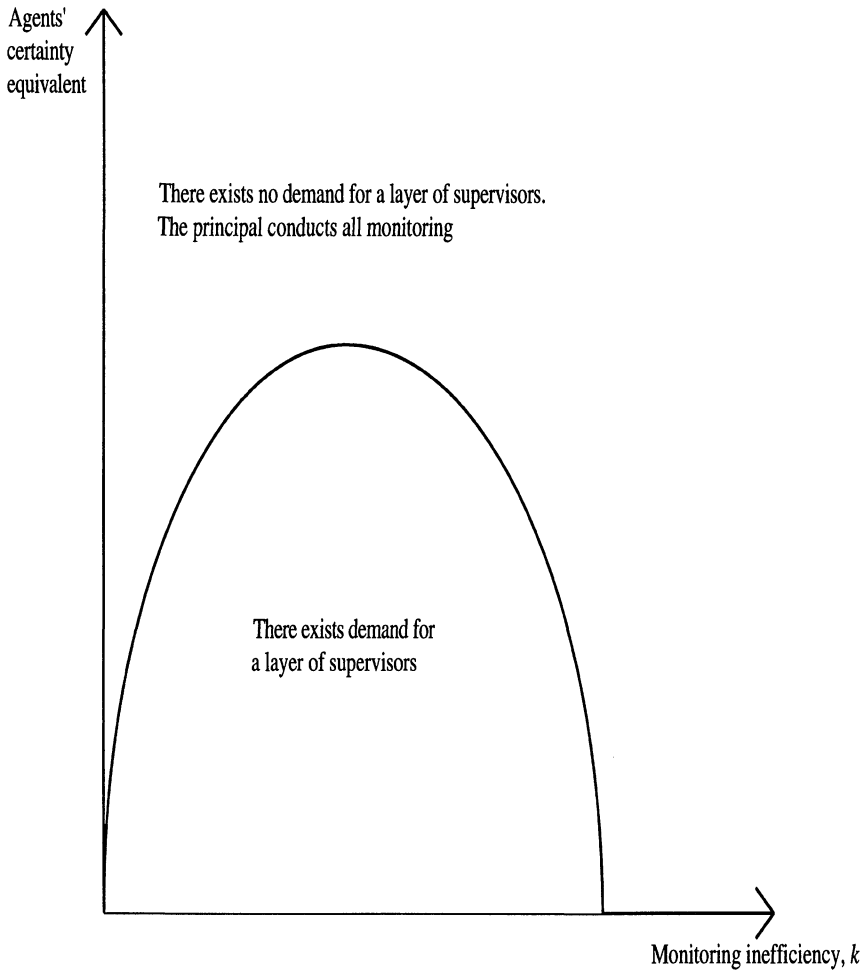
---

[19] See Claim 1 in the proof of Theorem 1.

FIG. 1.—Parameters where there exists a demand for a layer of supervisors, when $k = \gamma k_s$.

output, there exist sufficiently small levels of k, and sufficiently small $\bar{u}$, such that the principal hires at least one layer of supervisors.

Figures 1 and 2 illustrate Theorem 1. They demonstrate sets of monitoring inefficiencies and agents' certainty equivalent when the principal hires one layer of supervisors under the assumption $k = \gamma k_s$ (k_s is given). Intuitively, very efficient monitoring means not much is needed. The principal can enforce the desired production agents' effort without exerting much effort him/herself so s/he does not hire any supervisors. As monitoring inefficiencies increase, conducting all monitoring on his/her own becomes very costly. If supervisors are not too expensive ($\bar{u}$ is relatively small), the principal hires them, increasing production agents' monitoring.
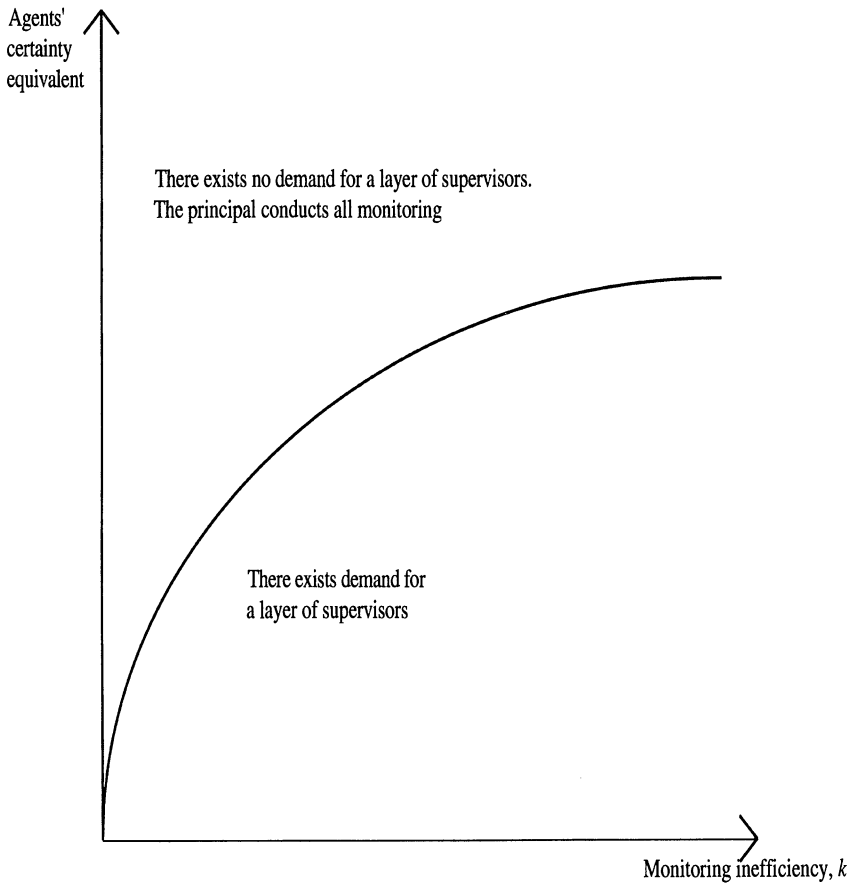
FIG. 2.—Parameters where there exists a demand for a layer of supervisors, when k_s is fixed.

The results of the two parts of the theorem are different for cases of high monitoring inefficiencies. Under the first scenario, all monitoring inefficiencies are high (they involve a high level of noise). Therefore, motivating supervisors is very costly because doing so requires that the principal exert great effort; hence, the principal reverts to conducting all monitoring. Under the second scenario, only monitoring of production activities is not efficient, and the demand for supervisors does not disappear (it actually increases). Finally, as the desired expected output, $\bar{Y}$, increases, so does the demand for a layer of supervisors.

The exact set of parameters when the principal prefers to hire a layer of supervisors can be derived for the example used earlier:

COROLLARY 4.   Assume $\omega = 2$, $\theta = 1$, and $\delta = 2$. Then, for a given level of desired expected output, the principal hires a layer of supervisors when $TC_2(\bar{Y}) \leq TC_1(\bar{Y})$, or:

$$\frac{3\left[\bar{u}^{\frac{1}{2}} + (2rk_s)^{\frac{1}{2}}\right]^{\frac{2}{3}}}{2(2rk)^{\frac{1}{6}}} \le \bar{Y}^{\frac{1}{3\beta}}. \tag{6}$$

Under certain circumstances, the principal may find neither the one-layer nor the two-layer organization optimal. S/he may either work on his/her own, without hiring any agents, or s/he may want to hire more than one layer of supervisors.

While working on his/her own saves the principal all monitoring costs, a sufficiently large production task, coupled with his/her own increasing work aversion, makes this alternative undesirable.

PROPOSITION 4.   When the principal works alone, his/her effort is a_p = $\bar{Y}^{\frac{1}{\beta}}$, and the total costs of production are $TC_0 = \bar{Y}^{\frac{\omega}{\beta}}$. When, in addition, $\omega = 2$, $\theta = 1$, and $\delta = 2$, the principal prefers working alone to hiring production agents if $TC_0 \le TC_1$, or $\bar{Y}^{\frac{1}{\beta}} \le 2\bar{u}^{\frac{1}{2}} + 2(2rk)^{\frac{1}{2}}$.[20]

When the principal could work alone, the demand for agents increases with the level of expected production and decreases with monitoring inefficiency. Even when the production agents are working under first-best conditions, for small enough expected production levels, the principal may be better off working alone.[21]

A second possibility is multiple layers of supervisors. While it is easy to construct a program similar to Program 2, assuming a firm with any number of layers, l, where the bottom layer of agents produces and all other layers supervise, even for the parameters of this example one cannot derive a closed-form solution. It is possible, however, to construct the marginal cost function for any number of layers, as a function of the optimal effort of the production workers, a, when $2\omega - \theta - 3 \ne 0$:

$$MC_{l+1}(\bar{Y}) = \left[\frac{2(\omega - 1)\bar{u} + a^{\omega}(\omega - \omega\theta - 2)}{a(2\omega - \theta - 3)}\right]\frac{1}{\beta}\bar{Y}^{\frac{1-\beta}{\beta}}.$$

Further numerical analysis provides no new insights: it seems the set of parameters supporting l layers of supervisors is a subset of the set of parameters supporting $l - 1$ layers of supervisors.

---

[20] It can be shown that this condition is sufficient for working alone as the optimal organization design; i.e., if this condition is satisfied, it cannot be that $TC_2 \le TC_0$.

[21] This observation requires the assumption that if the principal hires agents, s/he cannot work elsewhere and earn his/her market alternative. However, the principal might terminate the firm and work elsewhere. For this decision, the principal compares his/her profits from operating the firm and his/her market alternative.

Theorem 1 was derived for a given desired expected output. However, in general, the principal also chooses the optimal expected output, perhaps as a function of organizational structure. The following observation demonstrates this dependency under the assumption of a competitive market, where the output market price is P.

OBSERVATION 3.   When $\omega = 2$, $\theta = 1$, and $\delta = 2$, there exists a P^*, such that the principal hires supervisors only if the market price exceeds P^*. Furthermore, at P^* the optimal level of expected output is discontinuous, i.e., $\lim_{P \to P^{*-}}(\bar{Y}^*) < \lim_{P \to P^{*+}}(\bar{Y}^*)$. Also, when $TC_2(\bar{Y}) = TC_1(\bar{Y})$, it is always the case that $MC_2(\bar{Y}) < MC_1(\bar{Y})$.

As the market price increases, so does the optimal expected output and, as a result, the demand for supervision increases.

3.4 DISCUSSION AND EXTENSION

So far, I have assumed the information used in each agent's contract is provided by the signals generated by the agent's superior(s). While the assumption of uncorrelated signals precludes the use of relative performance evaluation, other available information might be used in the contract. In particular, the variance of the signals generated by a supervisor could be used to monitor the supervisor him/herself.[22] Recall that greater supervising effort reduces the variance of the signals generated (the mean is a function of the agent's effort and is not affected by the supervisor). I introduce this possibility, adding the variance of his/her reported observations, to each supervisor's contract.

Suppose the principal hires supervisors who provide public (hence nonmanipulative) signals regarding agents' efforts. The variance of these signals is informative about a supervisor's effort, so the principal would like to incorporate it in the supervisor's contract. As before, I assume linear contracts, now in the form: $s_j(\underline{x}) = \gamma_j + \alpha_{1j}x_{sj} - \alpha_{2j}\text{Var}(x_{aj})$, where x_{sj} denotes the signals generated by the principal on the effort of the jth supervisor, and $\text{Var}(x_{aj})$ denotes the variance of the signals generated by the jth supervisor on the production agents' effort. The additional term changes the risk imposed on each (risk-averse) supervisor. In particular, the variance of the new contract is: $\text{Var}(s_j(\underline{x})) = \alpha_{1j}^2\text{Var}(x_{sj}) + \alpha_{2j}^2\text{Var}(\text{Var}(x_{aj}))$.

Assume each supervisor has q observations, each normally distributed with a variance σ^2; then a standard statistical derivation (see Appendix A or Greene [1997]) shows the expected value of the variance of the supervisors' observations, denoted s^2, is σ^2 and its variance is $\dfrac{2\sigma^4}{q}$. Given the specification of my model, it can be shown that:

---

[22] Because the effort of the production agents has no impact on the variances, using variances in their compensation function has no value.

$$\text{Var}(x_{sj}) \;=\; \frac{k_s}{a_p^{\delta}}, \text{ and } \text{Var}(\text{Var}(x_{aj})) \;=\; \frac{4k^2}{a_s^{2\delta}\,q}.$$

Assuming mean-variance preferences, the supervisor's utility is:[23]

$$U(s_j(\underline{x}) - a_{sj}^{\omega}) \;=\; \gamma_j + \alpha_{1j}a_{sj} - \alpha_{2j}\text{Var}(x_{aj}) - \frac{1}{2}\,r\!\left(\alpha_{1j}^2\frac{k_s}{a_p^{\delta}} + \alpha_{2j}^2\frac{4k^2}{a_{sj}^{2\delta}\,q} \right) - a_{sj}^{\omega}.$$

Note that, unlike any other case discussed above, the supervisor's effort affects the risk s/he bears, so s/he has some control over his/her risk exposure. The first-order condition for each supervisor when choosing his/her effort is:

$$\alpha_{1j} + \frac{\alpha_{2j}k\delta}{a_s^{\delta+1}} - \frac{4r\alpha_{2j}^2k^2\delta}{qa_s^{2\delta+1}} \;=\; \omega\,a_s^{\omega-1}.$$

Several combinations of α_{1j} and α_{2j} induce a given level of effort. The principal's program becomes:

PROGRAM 2*a*.

$$\text{Min}_{\underline{a},\underline{a}_s,a_p,n,m,\underline{\alpha}} \sum_{i=1}^{n}\left(\bar{u} + a_i^{\omega} + \frac{rk\omega^2\,a_i^{2(\omega-1)}\,n^{\theta}}{2j\left(\sum_{j=1}^{m}a_{sj}\right)^{\delta}} \right)$$

$$+ \sum_{j=1}^{m}\left(\bar{u} + a_{sj}^{\omega} + \frac{1}{2}\,r(\alpha_{1j}^2m^{\theta}\frac{k_s}{a_p^{\delta}} + \alpha_{2j}^2\frac{4k^2}{a_{sj}^{2\delta}\,q})\right) + a_p^{\omega},$$

subject to: $(\Sigma_1^n\,\underline{a}_i)^{\beta} \geq \bar{Y}$ [Production]

$$\alpha_{1j} + \frac{\alpha_{2j}k\delta}{a_{sj}^{\delta+1}} - \frac{4r\alpha_{2j}^2k^2\delta}{qa_{sj}^{2\delta+1}} \;=\; \omega a_{sj}^{\omega-1}\qquad \text{[Supervisors' }IC\text{]}.$$

Solving the above, where $\omega = 2$, $\delta = 2$, and $\theta = 1$, yields the following coefficients:

$$\alpha_{1j} = 2a_s - \frac{2qk_s^{\frac{1}{2}}\left[a_{sj} - 2\,(2rk_s)^{\frac{1}{2}}\right]}{a_{sj}(2r)^{\frac{1}{2}}\left[a_{sj} - 4\,(2rk_s)^{\frac{1}{2}}\right]^{\frac{1}{2}}} \text{ and } \alpha_{2j} = \frac{qk_s^{\frac{1}{2}}\,a_{sj}^2}{(2r)^{\frac{1}{2}}k\left[a_{sj} - 4\,(2rk_s)^{\frac{1}{2}}\right]}$$

---

[23] Since the payoffs are not normally distributed, one cannot use the negative exponential utility assumption and derive mean-variance preferences.

where a_{sj} is each supervisor's optimal effort. When the number of observations per supervisor is very small, $q \rightarrow 0$, the principal cannot use the supervisor's variance, and Corollary 3 applies. I conjecture that α_{1j} is decreasing in q, while α_{2j} is increasing in q.[24] When q is very large, the principal generates no signals, chooses $\alpha_{1j} = 0$, and relies only on the information included in the variance term.

In the best possible case, the variance of the supervisors' observations is precise enough to infer their effort perfectly (equivalent to a perfect monitoring system where $k_s = 0$). More generally, using the variance to monitor the supervisors could be modeled as shifting monitoring inefficiency, k_s. Regardless of what might shift k_s, however, the results reported earlier in the paper, when $k_s \neq k$, are unchanged.

While my model considers only moral hazard, other information asymmetry problems, in particular adverse selection, may also be relevant. If the signals created by the supervisors are privately observed, their report may not be accurate. Proposition 2 involves supervisors' compensation based only on signals generated by the principal, not by the supervisors. Consequently, invoking the revelation principle (see Myerson [1982]), one can show that a truth-telling constraint is not violated, and the solution is unaffected. On the other hand, when the principal uses the variance of the supervisor's signals, the question of truth telling emerges because a supervisor's report affects his/her compensation. Since a supervisor possesses perfect private information about the variance of his/her observations, one can invoke the following from Melumad and Reichelstein [1987]:

OBSERVATION 4.   When supervisors report their observations to the principal, they possess perfect private information; hence, there is no value to communication, and the principal will not use the variance of the supervisors' observations in designing their optimal contract.

Finally, implementability arises in any multiagent setting. In particular, the principal (as well as the agents) should consider the possibility of other equilibria (see Demski and Sappington [1984] and Ma, Moore, and Turnbull [1988]) and the possibility of collusion (see Baiman, Evans, and Nagarajan [1991] and Villadsen [1995]). In my model, the contract with each agent is based only on his/her signals, with no implications for other contracts. On the other hand, because supervisors are reporting on agents' activities, and the principal "reports" his/her observations, implementation may become an important consideration. This issue is beyond the scope of the current paper.

4. Summary

In a principal–agent framework and given information (monitoring) technology, I analyzed the demand for supervisors (monitors) to observe

---

[24] The dependency of α_1 and α_2 on a_s does not allow for a formal derivation.

production agents who work under moral hazard. I showed a demand for monitoring services and studied its relation to monitoring inefficiency, the impact of supervision effort on signals' precision, and the impact of reduced control on signals' precision. The demand for supervision is shown to be a function of monitoring inefficiencies, agents' market alternatives, and total expected production level. In particular, a demand for supervisors exists only for moderate monitoring inefficiency coupled with a relatively low certainty equivalent. Otherwise, the principal is better off monitoring the production agents him/herself. As the total expected output increases, the demand for supervision increases as well. Finally, I showed that when the principal can use other sources of information to contract with the supervisors (like the variance of the signals generated by these supervisors), the demand for supervision increases.

APPENDIX A
Outline of the Proofs

OBSERVATION 1.   The total number of observations is q, so each of the n agents is observed, on average, $\frac{q}{n}$ times.[25] The mean of all signals is a sufficient statistic for contracting, and has a variance of $\dfrac{s^2}{\frac{q}{n}} = n\dfrac{s^2}{q} = n\sigma^2$.   □

DERIVATION OF PROGRAM 1.   There are three steps in the derivation of Program 1 from Program 0. First, it is straightforward to show that when using linear contracts, maximizing the negative exponential utility for any random cash flow is equivalent to maximizing the expected value of this cash flow less its variance multiplied by half the risk aversion coefficient.

Second, I use the following lemma to arrive at the optimal slope of linear contract.

LEMMA 1. An agent who faces the linear contract $s_i(\underline{x}) = \gamma_i + \alpha_i x_i$ selects an effort level of $a_i = \left(\dfrac{\alpha_i}{\omega}\right)^{\frac{1}{\omega-1}}$. Alternatively, to induce effort level a_i, the principal chooses $\alpha_i = \omega a_i^{\omega-1}$.

Proof.   The certainty equivalent for each agent $CE(s_i(\underline{x}) - a_i^\omega) = \gamma_i + \alpha_i E(x_i) - \frac{1}{2} r \alpha_i^2 \text{Var}(x_i) - a_i^\omega$. Using Assumption 3, the first-order condition is:

---

[25] The principal assigns each agent the same number of observations because the variance function is decreasing and convex in the number of observations. Hence, the marginal value of an additional observation is decreasing.

$$\frac{\partial CE(s_i(\underline{x}) - a_i^{\omega})}{\partial a_i} = \alpha_i - \omega a_i^{\omega-1} = 0 \text{ or } a_i = \left(\frac{\alpha_i}{\omega}\right)^{\frac{1}{\omega-1}}. \qquad \square$$

To arrive at Program 1, I replace the incentive compatibility constraint by using the optimal slope identified in Lemma 1 in the objective function.

Third, it is easy to see that the intercept of the linear contract serves only to allocate the total certainty equivalent between the principal and the agent; hence, one can solve an identical program with no intercept, where the individual rationality constraint does not appear. Taken together, these three steps simplify the efficient production problem drastically. One needs only to maximize the total certainty equivalent subject to the production constraint, i.e., Program 1. The Lagrangian for Program 1 is:[26]

$$\mathcal{L} = n\left(\bar{u} + a^{\omega} + \frac{rk\omega^2 a^{2(\omega-1)}n^{\theta}}{2a_p^{\delta}}\right) + a_p^{\omega} - \mu((na)^{\beta} - \bar{Y}).$$

PROPOSITION 1. The first-order conditions for Program 1 are:

$$\frac{\partial \mathcal{L}}{\partial a} = n\omega a^{\omega-1} + rk(\omega - 1)\omega^2 \frac{a^{2\omega-3}n^{\theta+1}}{a_p^{\delta}} - \mu\beta(na)^{\beta-1}n = 0,$$

$$\frac{\partial \mathcal{L}}{\partial n} = \bar{u} + a^{\omega} + \frac{1}{2}rk(\theta + 1)\omega^2 \frac{a^{2(\omega-1)}n^{\theta}}{a_p^{\delta}} - \mu\beta(na)^{\beta-1}a = 0,$$

$$\frac{\partial \mathcal{L}}{\partial a_p} = -\frac{\delta rk\omega^2 a^{2(\omega-1)}n^{\theta+1}}{2a_p^{\delta+1}} + \omega a_p^{\omega-1} = 0.$$

Algebraic manipulation of the first-order conditions and the production constraint provides the terms for the optimal number of production agents, n; the optimal effort of the production agents, a (equation (1)); and the optimal effort of the principal, a_p (equation (2)). Total production costs (equation (3)) are the value of Program 1 at the optimum. $\square$

COROLLARY 1. The derivatives of equation (1) (in its implicit form) provide for the comparative statics for the production agents effort, and as a result for the number of agents. The derivatives of equation (2) provide

---

[26] To save on notation, I remove the subscript i from the agents' optimal effort and solve for the effort required from each identical agent. It is easy to show the results are unaffected.

for the comparative statics for the principal's effort (taking into account the impact on the production agents' effort).   □

PROPOSITION 2.   The first-order conditions for Program 2 are:

$$\frac{\partial \mathcal{L}}{\partial a} = n\omega a^{\omega-1} + rk(\omega - 1)\omega^2 \frac{a^{2\omega - 3}n^{\theta + 1}}{(ma_s)^\delta} - \mu\beta(na)^{\beta-1}n = 0,$$

$$\frac{\partial \mathcal{L}}{\partial n} = \bar{u} + a^\omega + \frac{1}{2}rk(\theta + 1)\omega^2 \frac{a^{2(\omega - 1)}n^\theta}{(ma_s)^\delta} - \mu\beta(na)^{\beta-1}a = 0,$$

$$\frac{\partial \mathcal{L}}{\partial a_s} = -\frac{\delta rk\omega^2 a^{2(\omega - 1)}n^{\theta + 1}}{2a_s^{\delta + 1}m^\delta} + m\omega a_s^{\omega-1} + rk_s(\omega - 1)\omega^2 \frac{a_s^{2\omega - 3}m^{\theta + 1}}{a_p^\delta} = 0,$$

$$\frac{\partial \mathcal{L}}{\partial m} = -\frac{\delta rk\omega^2 a^{2(\omega - 1)}n^{\theta + 1}}{2a_s^\delta m^{\delta + 1}} + \bar{u} + a_s^\omega + \frac{1}{2}rk_s(\theta + 1)\omega^2 \frac{a_s^{2(\omega - 1)}m^\theta}{a_p^\delta} = 0,$$

$$\frac{\partial \mathcal{L}}{\partial a_p} = -\frac{\delta rk_s\omega^2 a_s^{2(\omega - 1)}m^{\theta + 1}}{2a_p^{\delta + 1}} + \omega a_p^{\omega-1} = 0.$$

Algebraic manipulation of the first-order conditions and the production constraint provides the terms in the proposition.   □

PROPOSITION 3.   Use $k_s = 0$ in the first-order conditions of Proposition 2.   □

COROLLARY 2.   First calculate the ratio $\frac{m}{n} =$

$$\left[\frac{\delta rk\omega\,(\omega - 1)^{\frac{\omega + \delta}{\omega}}\bar{Y}^{\frac{(2\omega - \delta - 3)}{\beta}}a^{1 + \delta}}{2\bar{u}^{\frac{\omega + \delta}{\omega}}}\right]^{\frac{1}{1 + \delta}}.$$ An alternative way of presenting this

ratio, which holds when $2\omega - \theta - 3 \neq 0$ and immediately provides the re-

sults, is:[27] $\frac{m}{n} = \frac{\delta\,(\omega - 1)}{\omega\,(2\omega - \theta - 3)}\left[1 - \frac{(\omega - 1)\,a^\omega}{\bar{u}}\right]$. The second presentation

for the ratio $\frac{m}{n}$ is affected through changes in the effort of the production

agents, a. When $\theta < 2\omega - 3$ $(\theta > 2\omega - 3)$, this ratio is decreasing (increasing) in a, while a is decreasing (increasing) in k. As for changes in the expected production level, $\bar{Y}$, the optimal effort is increasing in $\bar{Y}$ when θ

---

[27] Note this ratio is always positive, because when $2\omega - \theta - 3 > 0$ (< 0), the term in the square brackets of the equation in the text is positive (negative), due to the relation between $(\omega - 1)\,a^\omega$ and $\bar{u}$ (see equation (4)).

$< 2\omega - 3$ and $\delta < \theta$ or when $\theta > 2\omega - 3$ and $\delta > \theta$, and is decreasing in $\bar{Y}$ when $\theta < 2\omega - 3$ and $\delta > \theta$ or when $\theta > 2\omega - 3$ and $\delta < \theta$. As before, changes in the ratio are a result of the changes in the optimal effort. When $2\omega - \theta - 3 = 0$, the optimal effort, a, equals its first-best level, $\left(\dfrac{\bar{u}}{\omega - 1}\right)^{1/\omega}$, and is not affected by any of the parameters under discussion. It is clear the ratio is increasing in the monitoring inefficiency, k. For the result with respect to the expected production level, $\bar{Y}$, use $2\omega - 3 = \theta$ in the exponent of $\bar{Y}$. $\square$

COROLLARY 3.   Use $\omega = 2$, $\theta = 1$, and $\delta = 2$ in the first-order conditions of Proposition 2. The comparative statics with respect to k_s are immediate. $\square$

OBSERVATION 2.   Consider the solution to Program 2 as reported in Proposition 2, and use $k = k_s$. The expected compensation paid to a production agent is $E(s_i(\underline{x})) = \bar{u} + a^2 + 2rk\,\dfrac{a^2 n}{(m a_s)^2}$, while a supervisor earns $E(s_j(\underline{x})) = \bar{u} + a_s^2 + 2rk_s\,\dfrac{a_s^2 m}{a_p^2}$. Use the optimal effort $a = a_s = \bar{u}^{\frac{1}{2}}$ and the first-order condition with respect to a_p: $a_p = (2rk)^{\frac{1}{4}}(m a_s)^{\frac{1}{2}}$, to get: $E(s_j(\underline{x})) > E(s_i(\underline{x}))$, if and only if $(2rk)^{\frac{1}{2}} < \dfrac{m^2}{n}\,\bar{u}^{\frac{1}{2}}$. Next, consider the first-order condition with respect to a_s: $a_s + (2rk)^{\frac{1}{2}} = \dfrac{(2rk)(an)^2}{(m a_s)^3}$. Again, use the terms for a and a_s and solve for m: $m = \left[\dfrac{(2rk)\,n^2}{\left[\bar{u}^{\frac{1}{2}} + (2rk)^{\frac{1}{2}}\right]\bar{u}^{\frac{1}{2}}}\right]^{\frac{1}{3}}$. Recall, $n = \bar{u}^{-\frac{1}{2}}\bar{Y}^{\frac{1}{\beta}}$, to get $E(s_j(\underline{x})) > E(s_i(\underline{x}))$, if and only if $\dfrac{\left[\bar{u}^{\frac{1}{2}} + (2rk)^{\frac{1}{2}}\right]^{\frac{2}{3}}}{(2rk)^{\frac{1}{6}}} < \bar{Y}^{\frac{1}{3\beta}}$.

Now, as I show in Corollary 4, the principal hires supervisors if $\dfrac{3}{2}\dfrac{\left[\bar{u}^{\frac{1}{2}} + (2rk)^{\frac{1}{2}}\right]^{\frac{2}{3}}}{(2rk)^{\frac{1}{6}}} < \bar{Y}^{\frac{1}{3\beta}}$ (equation (6)). Obviously, if the principal hires supervisors, the condition $E(s_j(\bar{x})) > E(s_i(\underline{x}))$ is also satisfied. $\square$

THEOREM 1.   The theorem is proved using the following claims:

CLAIM 1.   If the principal optimally employs a layer of supervisors, then it is always the case that $ma_s > a_p$.

Proof.   Suppose not, i.e., $ma_s \leq a_p$, and consider the optimal contracts in the organization. Now, eliminate the layer of supervisors, and (weakly) reduce the principal's effort to be ma_s. Note the production agents' incentives are intact; hence, the total expected output and production costs remain the same. It is clear total monitoring costs are lower with no supervisors, a contradiction to the optimality of the two-layer organization.   □

CLAIM 2.   Assume $k_s = \gamma k$, $\gamma > 0$. Then, for sufficiently large monitoring inefficiency, k, the principal does not hire supervisors.

Proof.   Assume one layer of supervisors is optimal, and consider the first-order condition with respect to the supervisors' effort, using $k_s = \gamma k$:[28]

$$\omega a_s^{\omega - 1} = rk\omega^2 \left[\frac{\delta a^{2(\omega - 1)} n^{\theta + 1}}{2(ma_s)^{\delta + 1}} - \frac{(\omega - 1)\gamma a_s^{2\omega - 3} m^{\theta}}{a_p^{\delta}} \right].$$

The left-hand side is positive; hence it must be: $\left(\dfrac{a_p}{ma_s} \right)^{\delta} >$

$\dfrac{2(\omega - 1)\gamma a_s^{2(\omega - 1)} m^{\theta + 1}}{\delta a^{2(\omega - 1)} n^{\theta + 1}}$. Use $n = \dfrac{1}{a} \bar{Y}^{\frac{1}{\beta}}$ and Claim 1 to conclude:

$a_s^{2(\omega - 1)} m^{\theta + 1} < \dfrac{\delta}{2(\omega - 1)\gamma} \bar{Y}^{\frac{2(\omega - 1)}{\beta}}$. Now, since $m \geq 1$ and $a_s \geq \varepsilon$, we

know $a_s \leq \left[\dfrac{\delta}{2(\omega - 1)\gamma} \bar{Y}^{\frac{2(\omega - 1)}{\beta}} \right]^{\frac{1}{2(\omega - 1)}}$ and $m \leq \left[\dfrac{\delta}{2(\omega - 1)\gamma} \right.$

$\left. \bar{Y}^{\frac{2(\omega - 1)}{\beta}} \right]^{\frac{1}{2(\omega - 1)}} \left(\dfrac{1}{\varepsilon} \right)^{\frac{2(\omega - 1)}{\theta + 1}}$, so both terms are bounded from above as

well.

Following algebraic manipulation of the first-order conditions with respect to the principal's effort when one layer of supervisors is hired:

---

[28] In the proof of Theorem 1, I assume the constraints that the effort and the number of agents are bounded from below are not binding. If any of these constraints is binding, the proof is much simpler, and it is omitted.

$\left(\dfrac{a_p}{ma_s}\right)^{\omega+\delta} = \dfrac{1}{2}\delta r\gamma k\omega\, a_s^{\omega-\delta-2} m^{\theta+1-\omega-\delta}$. As both m and a_s are bounded from

above and below, for sufficiently large monitoring inefficiency, k, the right-hand side must be greater than one, which implies the left-hand side is also greater than one, a contradiction to the optimality of a layer of supervisors. $\square$

CLAIM 3. For a sufficiently small monitoring inefficiency, k, the principal does not hire supervisors.

Proof. Assume the principal hires, optimally, m supervisors. His/her monitoring costs are at least $m\bar{u}$. Now, assume the principal enforces the first-best effort, number of production agents, and total expected production, without hiring any supervisors. (Note this may not be the optimal arrangement when the principal conducts all monitoring.) The difference in costs between the first-best case and the contract suggested here

is the risk premium for the agents, $\dfrac{1}{2}\delta^{-\frac{\delta}{\omega+\delta}}(rk\omega^2)^{\frac{\omega}{\omega+\delta}} a^{\frac{2\omega(\omega-1)}{\omega+\delta}} n^{\frac{\omega\theta-\delta}{\omega+\delta}}$, and

the cost of the principal's monitoring needed to provide the production agents with incentives to exert the first-best effort level, $a_p^{\omega} = [\delta rk\omega^2 a^{2(\omega-1)}$

$n^{\theta+1}]^{\frac{\omega}{\omega+\delta}}$. When monitoring inefficiency, k, is sufficiently small, the sum of these two terms is less than $\bar{u}$, the lower bound for the costs of the supervisors. Hence, monitoring by the principal generates higher payoffs for the principal. $\square$

CLAIM 4. For a sufficiently large certainty equivalent, $\bar{u}$, the principal does not hire supervisors.

Proof. Consider the optimal contracts when the principal hires one layer of supervisors. There are m supervisors, and each exerts an effort of a_s. The costs are $m\bar{u} + ma_s^{\omega}$ + some risk premium. Now, assume the principal replaces the supervisors and exerts as much effort: $a_p = ma_s$. Note the production workers' incentives are intact. Ignore the risk premium paid to the supervisors, and compare the cost of monitoring. It is clear when $\bar{u}$ is large enough, $m\bar{u} + ma_s^{\omega} > (ma_s)^{\omega}$. (Recall a_s and m are bounded from above.) $\square$

CLAIM 5. For any set of parameters and desired expected output, there exist some level (midrange) of monitoring inefficiency, k, and sufficiently small agents' certainty equivalent, such that the principal hires at least one layer of supervisors.

Proof. Consider the optimal one-layer organization. Total production costs are given by equation (3), where the optimal effort of the production agents and of the principal are determined by equations (1) and (2), respectively.

Assume the principal hires one layer of supervisors, and designs their contracts such that their total effort is identical to his/her effort when s/he conducts all monitoring. (This may not be an optimal arrangement for a two-layer organization.) In order to enforce the desired effort from the supervisors, the principal should exert effort, in particular (see Proposition 2): $a_p = \left[\frac{1}{2} \delta r k_s \omega a_s^{2(\omega-1)} m^{\theta+1} \right]^{\frac{1}{\omega+\delta}}$. Note the production agents' contracts, assignments, and expected compensation are, by design, intact. It can be shown that the total costs are:

$$TC_2(\bar{Y}) = (\bar{u} + a^\omega) \frac{1}{a} \bar{Y}^{\frac{1}{\beta}} + \frac{\omega}{\delta} \left(\frac{1}{2} \delta r k \omega a^{2\omega-\theta-3} \bar{Y}^{\frac{\theta+1}{\beta}} \right)^{\frac{\omega}{\omega+\delta}}$$

$$+ m(\bar{u} + a_s^\omega) + (1 + \frac{\omega}{\delta})\left(\frac{1}{2} \delta r k_s \omega a_s^{2(\omega-1)} m^{\theta+1} \right)^{\frac{\omega}{\omega+\delta}}.$$

Now, consider $m = 2$, so $a_s = \frac{1}{2}\left[\frac{1}{2} \delta r k \omega a^{2\omega-\theta-3} \bar{Y}^{\frac{\theta+1}{\beta}} \right]^{\frac{1}{\omega+\delta}}$. Use equation (1) for the agents' certainty equivalent, and compare total production costs between the optimal one-layer and the constructed two-layer organizations. It appears, $TC_{2\,-constructed}(\bar{Y}) < TC_1(\bar{Y})$ if:[29]

$$2(\omega - 1)a^\omega + A_3(2\omega - \theta - 3)k^{\frac{\omega}{\omega+\delta}} a^{\frac{2\omega(\omega-1)+\delta-\theta\omega}{\omega+\delta}}$$

$$+ A_1 k^{\frac{2(\omega-1)\omega}{\omega+\delta}} k_s^{\frac{\omega}{\omega+\delta}} a^{\frac{2\omega(\omega-1)(2\omega-\theta-3)}{(\omega+\delta)^2}} < A_2 k^{\frac{\omega}{\omega+\delta}} a^{\frac{(2\omega-\theta-3)\omega}{\omega+\delta}}, \qquad (7)$$

where $A_1 = 2^{\frac{\omega(\theta+3-2\omega)}{\omega+\delta}} \left(1 + \frac{\omega}{\delta} \right) \left(\frac{1}{2}\delta r\omega \right)^{\frac{3\omega+\delta-2}{(\omega+\delta)^2}} \bar{Y}^{\frac{2\omega(\omega-1)(\theta+1)}{\beta(\omega+\delta)^2}} > 0$, $A_2 =$

$\left[\frac{1}{2}\delta r\omega \bar{Y}^{\frac{\theta+1}{\beta}} \right]^{\frac{\omega}{\omega+\delta}} \left(1 - \frac{1}{2^{\omega-1}} \right) > 0$, and $A_3 = 2\omega\delta^{\frac{-\delta}{\omega+\delta}} \left(\frac{1}{2} r\omega \right)^{\frac{\omega}{\omega+\delta}} \bar{Y}^{\frac{\theta\omega-\delta}{\beta(\omega+\delta)}}$

> 0. It is sufficient to show that each of the terms on the left-hand side of equation (7) is less than one-third of the right-hand side.[30]

For the first term on the left-hand side, we need:

___

[29] I use here the production agents' effort and not the certainty equivalent directly. This is done for convenience, because one cannot explicitly solve equation (1) for the production agents' effort. For any given a and k identified in the proof below, there is a unique corresponding certainty equivalent $\bar{u}$.

[30] I present here the proof only for the case where $\theta < 2\omega - 3$. The proof for the other two cases, $\theta > 2\omega - 3$ and $\theta = 2\omega - 3$, is similar and less complicated. When $\theta > 2\omega - 3$, the second term on the left-hand side is negative and could be ignored. Also, in this case, one can easily verify that, given the selection of a and k, $\bar{u}$ is strictly positive.

$$2(\omega - 1)a^\omega < \frac{1}{3} A_2 k^{\frac{\omega}{\omega + \delta}} a^{\frac{(2\omega - \theta - 3)\omega}{\omega + \delta}}$$

or:
$$\left[\frac{6(\omega - 1)}{A_2}\right]^{\frac{\omega + \delta}{\omega}} a^{\delta + \theta + 3 - \omega} < k. \tag{8}$$

For the second term on the left-hand side, we need:

$$A_3(2\omega - \theta - 3)k^{\frac{\omega}{\omega + \delta}} a^{\frac{2\omega(\omega - 1) + \delta - \theta\omega}{\omega + \delta}} < \frac{1}{3} A_2 k^{\frac{\omega}{\omega + \delta}} a^{\frac{(2\omega - \theta - 3)\omega}{\omega + \delta}}$$

or:
$$a < \frac{A_2}{3A_3(2\omega - \theta - 3)} . \tag{9}$$

For the third term on the left-hand side, we need:

$$A_1 k_s^{\frac{\omega}{\omega + \delta}} k^{\frac{2\omega(\omega - 1)}{(\omega + \delta)^2}} a^{\frac{2\omega(\omega - 1)(2\omega - \theta - 3)}{(\omega + \delta)^2}} < \frac{1}{3} A_2 k^{\frac{\omega}{\omega + \delta}} a^{\frac{(2\omega - \theta - 3)\omega}{\omega + \delta}}$$

or:
$$k_s^{\omega + \delta} k^{\omega - \delta - 2} < \left[\frac{A_2}{3A_1}\right]^{\frac{(\omega + \delta)^2}{\omega}} a^{(2\omega - \theta - 3)(\delta + 2 - \omega)}. \tag{10}$$

There are three cases to consider here: (i) Assume $k_s = \gamma k$, $\gamma > 0$. Equation (10) becomes $k < \left[\frac{A_2}{3A_1}\right]^{\frac{(\omega + \delta)^2}{2\omega(\omega - 1)}} a^{\frac{(2\omega - \theta - 3)(\delta + 2 - \omega)}{2(\omega - 1)}} \gamma^{\frac{-(\omega + \delta)}{2(\omega - 1)}}$. Notice that k has a lower (equation (8)) and an upper (equation (10)) bound. Hence, one needs to select an effort level, a, such that both bounds are satisfied. It is easy to see that the lower bound is smaller than the upper bound when:

$$a^{\omega + \omega\theta + \delta + \theta\delta} < \left[\frac{A_2}{3A_1}\right]^{\frac{(\omega + \delta)^2}{\omega}} \left[\frac{A_2}{6(\omega - 1)}\right]^{\frac{2(\omega + \delta)(\omega - 1)}{\omega}} . \tag{11}$$

Obviously, when a is sufficiently small, equations (9) and (11) are satisfied. Hence, there always exist k and $\bar{u}$ (corresponding to a above) where the principal prefers hiring at least one layer of supervisors.

(ii) Assume k and k_s are not proportional to each other, and that $\omega - \delta - 2 > 0$. Here, equation (10) becomes: $k < \left[\frac{A_2}{3A_1}\right]^{\frac{(\omega + \delta)^2}{2\omega(\omega - 1)}} k_s^{\frac{-(\omega + \delta)}{2\omega(\omega - \delta - 2)}} a^{-(2\omega - \theta - 3)}$, and the lower bound for k is smaller than the upper bound when $a < \left[\frac{A_2}{3A_1}\right]^{\frac{(\omega + \delta)}{\omega(\omega - \delta - 2)}} \left[\frac{A_2}{6(\omega - 1)}\right]^{\frac{\omega}{(\omega + \delta)^2}} k_s^{\frac{-1}{\omega(\omega - \delta - 2)}}$. Again, when a

is sufficiently small, both equations (9) and (11) are satisfied, and the principal prefers hiring at least one layer of supervisors.

(*iii*) Assume k and k_s are not proportional to each other, and that $\omega - \delta - 2 < 0$. Here, equation (10) becomes $k > \left[\dfrac{3A_1}{A_2}\right]^{\frac{(\omega+\delta)^2}{\omega(\delta+2-\omega)}} k_s^{\frac{(\omega+\delta)}{\omega(\delta+2-\omega)}} a^{2\omega-\theta-3}$. Hence, recalling that a is decreasing in k, when $k > $ Max

$$\left[\left[\dfrac{6(\omega-1)}{A_2}\right]^{\frac{(\omega+\delta)}{\omega}} a^{\delta+\theta+3-\omega}, \left[\dfrac{3A_1}{A_2}\right]^{\frac{(\omega+\delta)^2}{\omega(\delta+2-\omega)}} k_s^{\frac{(\omega+\delta)}{\omega(\delta+2-\omega)}} a^{2\omega-\theta-3}\right], \text{ and } a$$

$< \dfrac{A_2}{3A_3(2\omega-\theta-3)}$, the principal prefers hiring at least one layer of supervisors.   □

Claims 1–5 are used to establish Theorem 1, part (*i*), while Claims 1 and 3–5 are used to establish Theorem 1, part (*ii*).   □

COROLLARY 4.   Use $\omega = 2$, $\theta = 1$, and $\delta = 2$ in calculating the total production costs for the one-layer organization (equation (3)) and for the two-layer organization (equation (5)): $TC_1(\bar{Y}) = \left[2\bar{u}^{\frac{1}{2}} + 2(2rk)^{\frac{1}{2}}\right]\bar{Y}^{\frac{1}{\beta}}$ and

$TC_2(\bar{Y}) = \left[2\bar{u}^{\frac{1}{2}} + 3(2rk)^{\frac{1}{3}}\left[\bar{u}^{\frac{1}{2}} + (2rk_s)^{\frac{1}{2}}\right]^{\frac{2}{3}} \bar{Y}^{-\frac{1}{3\beta}}\right]\bar{Y}^{\frac{1}{\beta}}$. Comparing the above total costs terms yields the result.   □

PROPOSITION 4.   The principal's program when working alone is:

$$\text{Min}_{a_p} a_p^{\omega}, \text{ subject to } (a_p)^{\beta} \geq \bar{Y} \quad \text{[Production]}.$$

The production constraint determines a_p and yields the first result. Use $\omega = 2$, $\theta = 1$, and $\delta = 2$, and compare to TC_1 (equation (3)), to arrive at the second result.   □

OBSERVATION 3.   The principal decides about the optimal expected production level by comparing the market price to the marginal costs. Given the market price, there are two possible optimal expected output levels: $P = MC_1(\bar{Y})$ and $P = MC_2(\bar{Y})$. Comparing the two marginal cost terms, it is easy to see $MC_2(\bar{Y}) < MC_1(\bar{Y})$ when $\dfrac{\left[\bar{u}^{\frac{1}{2}} + (2rk)^{\frac{1}{2}}\right]^{\frac{2}{3}}}{(2rk)^{\frac{1}{6}}} \leq \bar{Y}^{\frac{1}{3\beta}}$.

Note the left-hand side is smaller than the left-hand side of equation (6). Hence, when $TC_2(\bar{Y}) = TC_1(\bar{Y})$, it must be that $MC_2(\bar{Y}) < MC_1(\bar{Y})$.

If both possible optimal expected output levels are in the range where $TC_2(\bar{Y}) < TC_1(\bar{Y})$ ($TC_2(\bar{Y}) > TC_1(\bar{Y})$), the principal prefers two- (one-) layer organization. On the other hand, if the optimal expected output $\bar{Y}_1^*(\bar{Y}_2^*)$ when $MC_1(\bar{Y}) = P$ ($MC_2(\bar{Y}) = P$) is such that $TC_2(\bar{Y}_1^*) > TC_1(\bar{Y}_1^*)$ ($TC_2(\bar{Y}_2^*) < TC_1(\bar{Y}_2^*)$), then both solutions are within their relevant range, and the principal should compare his/her profits.

Finally, note that when $P = MC_1(\bar{Y}_1) = MC_2(\bar{Y}_2)$ and $\bar{Y}_2 > \bar{Y}_1$, the slope of the function $MC_2(\bar{Y})$ is smaller than the slope of the function $MC_1(\bar{Y})$. Hence, if for a price P the principal prefers a two-layer organization, s/he also prefers a two-layer organization for any higher price.   □

USING THE VARIANCE IN THE SUPERVISOR'S CONTRACT.   We know $x_i \sim$

$N(\mu, \sigma^2)$ $\forall i = 1 \ldots q$. Hence, $\bar{x} \sim N\left(\mu, \dfrac{\sigma^2}{q}\right)$ and $(x_i - \bar{x}) \sim N\left(0, \dfrac{q-1}{q}\sigma^2\right)$.

It follows that $\dfrac{q}{q-1}\ \Sigma\left(\dfrac{x_i - \bar{x}}{\sigma}\right)^2 \sim \chi^2(q)$, so $\dfrac{q}{\sigma^2}\left[\dfrac{1}{q-1}\ \Sigma(x_i - \bar{x})^2\right] \sim \chi^2(q)$.

Define $s^2 = \left[\dfrac{1}{q-1}\ \Sigma(x_i - \bar{x})^2\right]$, so $\dfrac{q}{\sigma^2}s^2 \sim \chi^2(q)$. The mean and the vari-

ance of a χ^2 distribution with q degrees of freedom is given by $E\left(\dfrac{qs^2}{\sigma^2}\right) =$

q and $\mathrm{Var}\left(\dfrac{qs^2}{\sigma^2}\right) = 2q$.

REFERENCES

BAIMAN, S.; J. EVANS; AND N. NAGARAJAN. "Collusion in Auditing." *Journal of Accounting Research* (Spring 1991): 1–18.

BAIMAN, S.; D. LARCKER; AND M. RAJAN. "Organizational Design for Business Units." *Journal of Accounting Research* (Autumn 1995): 205–30.

BALDENIUS, T.; N. MELUMAD; AND A. ZIV. "Monitoring in Multi-Agent Organizations." Working paper, Columbia University, July 2000.

CALVO, G., AND S. WELLISZ. "Supervision, Loss of Control, and the Optimal Size of the Firm." *Journal of Political Economy* (October 1978): 943–52.

DEMSKI, J., AND R. DYE. "Risk, Return, and Moral Hazard." *Journal of Accounting Research* (Spring 1999): 27–55.

DEMSKI, J., AND D. SAPPINGTON. "Optimal Incentive Contracts with Multiple Agents." *Journal of Economic Theory* 33 (1984): 152–71.

GREENE, W. *Econometric Analysis.* Englewood Cliffs, N.J.: Prentice-Hall, 1997.

HOLMSTROM, B. "Moral Hazard and Observability." *Bell Journal of Economics* (Spring 1979): 74–91.

HOLMSTROM, B., AND P. MILGROM. "Aggregation and Linearity in the Provision of Inter-temporal Incentives." *Econometrica* (March 1987): 303–28.

———. "Multi-Task Principal–Agent Analysis: Incentive Contracts, Asset Ownership and Job Design." *Journal of Law, Economics, and Organization* (Supplement 1991): 24–52.

JEWITT, I. "Justifying the First Order Approach to Principal–Agent Problems." *Econometrica* (September 1988): 1177–90.

MA, C.; J. MOORE; AND S. TURNBULL. "Stopping Agents from Cheating." *Journal of Economic Theory* 46 (1988): 355–72.

MELUMAD, N., AND S. REICHELSTEIN. "Centralization versus Delegation and the Value of Communication." *Journal of Accounting Research* (Supplement 1987): 1–18.

MELUMAD, N.; D. MOOKHERJEE; AND S. REICHELSTEIN. "A Theory of Responsibility Centers." *Journal of Accounting and Economics* (December 1992): 445–84.

MIRRLEES, J. "The Theory of Moral Hazard and Unobservable Behavior, Part I." *Review of Economic Studies* (January 1999): 3–21.

MOOKHERJEE, D., AND S. REICHELSTEIN. "Budgeting and Hierarchical Control." *Journal of Accounting Research* (Autumn 1997): 129–56.

MYERSON, R. "Optimal Coordination Mechanisms in Generalized Principal–Agent Relationships." *Journal of Mathematical Economics* (1982): 67–81.

VILLADSEN, B. "Communication and Delegation in Collusive Agencies." *Journal of Accounting and Economics* (March 1995): 315–44.

WILLIAMSON, O. E. *The Economic Institutions of Capitalism.* New York: Free Press, 1985.

ZIV, A. "Performance Measure and Optimal Organization." *Journal of Law, Economics, and Organization* 9 (1993): 30–50.