

# Hierarchical Contract, Firm Size, and Pay Sensitivity

Hyeng Keun Koo

Gyoocheol Shim

Jaeyoung Sung\*

March 13, 2007

## Abstract

We present a moral-hazard-based hierarchical contracting model, where investors contract the top manager and the top manager contracts all middle managers. We compare effects of hierarchical contracting on managerial contract sensitivities with those of a direct contracting benchmark where investors directly contract all managers. We argue that under hierarchical contracting, the top manager shifts his compensation risk to middle managers by providing middle managers with higher-powered incentive contracts than would be desired by investors under direct contracting. It is striking that this top managerial risk-shifting behavior motivates investors to design the top managerial contract in such a way that the top-managerial hierarchical contract sensitivity approaches either the first best or zero, as the firm size grows. However, under some reasonable conditions such as correlated managerial effort outcomes, the top managerial sensitivity quickly approaches zero as the firm size increases, and consequently, the sensitivity for large firms can be far lower than predicted by the standard agency theory. This result can serve as an explanation of widely observed firm-size effects on CEO compensations, namely, lower pay sensitivities for large firms than those for small firms. We also argue that even when investors are risk-neutral and then individual performance outcomes of nonexecutive employees may be very weakly correlated to the total outcome of the firm, company-wide bonus plans for nonexecutive employees can still be justified under hierarchical contracting.

*Keywords:* hierarchical contract, firm size, pay sensitivity, CEO compensation, middle managerial contract, principal-agent problem, many agents, moral hazard, team, performance measure, continuous-time model

*JEL Classifications:* D86, G30, L14, L22, L25

---

\*Koo: School of Business Administration, Ajou University, Suwon, 443-749, Korea; hkoo@ajou.ac.kr. Shim: The Bank of Korea, Seoul, 100794, Korea; gshim@bok.or.kr. Sung: Department of Finance (M/C 168), University of Illinois at Chicago, 601 South Morgan Street, Chicago, Illinois 60607-7124; (312) 996-0720 (Phone); (312) 413-7948 (Fax); jsung@uic.edu. This paper was presented at the 2006 International Workshop on Mathematical Finance and Insurance, Lijang, China, HKUST and CUHK.

# 1 Introduction

We present a hierarchical contracting model under moral hazard with many agents, where investors (the principal) contract the top manager who in turn subcontracts many middle managers. The model enables us to examine effects of not only hierarchical contracts but the firm size on incentives of individual employees including the top manager. Throughout the paper, the firm size is synonymous with the number of middle managers.

Publicly held corporations can be viewed as organizations established with various sets of contracts among investors and employees. In financial economics, standard principal agent models, structured with one principal and one agent, are frequently utilized to analyze contractual relationships between investors and the top manager of the firm, completely ignoring other employees and the organizational form of the firm. Such frequent practices are based on the implicit assumption that the top manager can represent all employees including himself, and his incentives are not affected by the organizational form. Being made mostly for tractability of models, the assumption is at best a very crude abstraction of complicated issues that may arise because of the organizational form, and interaction with other employees.

In real life, typical contracts for employees of the firm are hierarchically organized with “a cascade of principal-agent relationships.” In the literature, there are many reasons offered to justify such hierarchical forms of organizations: for information processing and communication (Radner [1992, 1993], Bolton and Dewatripont [1994], Marschak and Reichelstein [1998], and Stein [2002]); for authority and control in decision making (see Rosen [1982], Aghion and Tirole [1997], and Hart and Moore [2005]); and for monitoring subordinates’ outputs/effort levels, (see Williamson (1967), Mirrlees (1976), Calvo and Wellisz (1978,1979), and Qian [1994]).<sup>1</sup> However, none of the above studies examine how the hierarchy of the firm can affect optimal managerial contracts.

Radner (1992) states:

In fact, most organizations combine aspects of both the partnership and principal-agent models. A hierarchy of authority can be thought of as a cascade of principal-agent relationships, each supervisor acting as a principal to his subordinates, and as an agent in relationship to his own supervisor. On the other hand, in most cases the valued outcomes of organizational activity depend on the joint action of several agents, as in the partnership model, so that the assignment of individual responsibility for specific outcomes as required by the principal-agent model may not be justified. Unfortunately, I am not aware of significant progress on more

---

<sup>1</sup>All of these authors study various motivations for hierarchical organizations, but provide no formal analysis on managerial contracting problems.

comprehensive theoretical models of the firm that combine these two submodels in a systematic way.

We believe, even up to this date, Radner's comment still holds. In this paper, we try to start such a comprehensive model using moral hazard problems as an underlying basis.<sup>2</sup> Our model may also be viewed as a modest step beyond the existing "teams" literature such as Holmstrom [1982] and Mookherjee [1984] by introducing a cascade of principal-agent problems in a hierarchically organized firm with its total outcome resulting from joint/individual actions of the top and middle managers.

In this paper, investors are risk neutral and all managers are risk averse. There are multiple tasks in the firm and each task is carried out by one manager. The top manager exerts effort which affects not only his own but all middle managerial effort outcomes. He also acts as both a principal and a monitor to each middle manager. In particular, he monitors all middle-managerial effort outcomes, and compensates each middle manager based on all outcomes. On the other hand, investors can only observe the aggregate outcome net of the aggregate middle managerial compensation.<sup>3</sup>

We compare the optimal structure of managerial hierarchical contracts with that of managerial contracts resulting from direct contracting with full information under which investors directly contract each manager and observe all managerial effort outcomes. In the literature, it has been widely recognized that if the principal is risk neutral and all managerial outcomes are independent of each other, optimal performance measures for each manager can be constructed by using sufficient statistics of the managerial effort (see Holmstrom [1982] and Mookherjee [1984]). Although we reconfirm the sufficient statistics result for our direct contracting case, the sufficient statistics do not provide a necessary set of performance measures for our optimal hierarchical contracts, because the top manager acting as a principal to middle managers is risk averse.

Under our hierarchical contracting, each optimal middle managerial contract between the top and middle managers is based on three performance measures: the individual middle-managerial effort outcome, top managerial effort outcome, and the total outcome of the firm. Note that the first measure is a sufficient statistic of the individual middle managerial effort. The other two measures are also added because the risk-averse top manager tries to share his own compensation risk with middle managers, even when the two measures have almost nothing to do with middle managerial effort levels. This observation may provide a rationale

---

<sup>2</sup>Melumad, Mookherjee, and Reichelstein [1995] examine effects of adverse selection under a hierarchical contracting environment. In their model, there is only one middle manager, and all managers are risk-neutral.

<sup>3</sup>The monitoring structure of the firm in this paper is somewhat similar to that of Mirrlees [1976, section 5], although he does not discuss optimal contracts.

to justify common managerial compensation practices such as company-wide bonus programs and stock option plans for nonexecutive employees (see Core and Guay [2001] for stock option plans for nonexecutive employees).<sup>4</sup>

We also compare optimal top and middle-managerial contract sensitivities of our hierarchical contracts with those of the direct contracting benchmarks. As compared with the direct contracting with full information, inefficiency can arise under our hierarchical contracting, as the risk-averse top manager tries to shift his own compensation risk to middle managers by providing too high-powered middle managerial effort incentives.<sup>5</sup> As a consequence, middle managers work harder than they do under direct contracting.<sup>6</sup> Furthermore, we show that as the firm size increases, sensitivities of middle managerial hierarchical contracts decrease toward those of direct contracts. The reason is that as the firm size increases, the top manager keeps shifting his own compensation risk to middle managers, but this shifted risk is gradually diversified away because the risk is shared by a large number of middle managers.

On the other hand, we show that the top managerial (contract) sensitivity depends on several economic factors such as common uncertainty across all managerial effort outcomes, returns to scale on the firm's production, the top managerial effort productivity relative to that of middle managers, and the aggregate middle managerial risk-sharing premium. In particular, for large firms, the sensitivity of the top managerial hierarchical contract can be either much higher or lower than that of the direct contracting case.

We discuss the case of independent outcomes, first, and then the case of correlated outcomes with common uncertainty. For the case of independent outcomes where all managerial outcomes are independent of each other, we show that if the firm's production function exhibits nonincreasing returns to scale in labor, and the relative effort productivity of the top manager is lower (higher) than a certain threshold, the optimal top managerial sensitivity approaches zero (the first best), as the firm size increases. The threshold depends on top managerial effort productivity and the aggregate middle managerial risk-sharing premium.

The intuition for the two extreme top-managerial sensitivities (i.e., either zero or the first best) for large firms is as follows: Given a top managerial contract in place, the risk averse

---

<sup>4</sup>The dependence of middle managerial contracts on top managerial effort outcome and the aggregate outcome may also provide a new insight into relative performance evaluation. Traditional literature on relative schemes (e.g., Lazear and Rosen 1981, Nalebuff and Stiglitz 1983, Gibbons and Murphy 1990, Dye 1992, Core and Guay 2003) emphasize their usefulness for filtering common noises and identifying true effort levels of individual workers. We show, however, that dependence of a middle manager's compensation of outputs of others in the same firm is not necessarily based on filtering, but may be caused by the top manager's incentive to share risk.

<sup>5</sup>Of course, investors can be better off with our hierarchical contracting than direct contracting if they can observe only the aggregate of all managerial effort outcomes under direct contracting. In this paper, we do not discuss this straightforward direct contracting case with the limited information structure.

<sup>6</sup>This result may be indirectly related to a recent empirical finding by Aggarwal and Samwick [2003] who document positive and significant (pay-divisional) sensitivities of middle managerial contracts. Our results indicate that more future empirical research is called for.

top manager assigns an excessively high-powered incentive contract to each middle manager in order to reduce his own compensation risk, and each middle manager demands a risk-sharing premium for the excessive compensation risk imposed by the top manager. The aggregate risk-sharing premium is eventually shifted to investors, decreasing investors' expected wealth. However, the risk-shifting behavior engenders not only a cost but a benefit to investors, because with reduced compensation risk, the top managerial incentive to work is improved. Therefore, investors strike a balance between the cost and benefit, in order to determine whether to encourage or discourage such risk-shifting/risk-sharing behavior.

If the marginal gain from an improvement of the top managerial effort incentives dominates (is dominated by) the marginal aggregate risk-sharing premium to be paid to middle managers, investors would like to encourage (discourage) the top managerial risk-shifting behavior by increasing (decreasing) the top managerial sensitivity. Thus, when the relative effort productivity of the top manager is sufficiently higher (lower) than the aforementioned threshold, the top managerial sensitivity is higher (lower) than that of the direct contracting benchmark, and in fact, the sensitivity approaches the first best (zero) as the firm size grows.

Although the above independent-outcome case helps us understand various tradeoffs, it may be more realistic for managerial effort outcomes to be correlated with each other through some sort of common uncertainty within the firm. We find that even in the presence of common uncertainty, the top managerial contract sensitivity still depends on the basic tradeoff between the marginal top managerial incentives and the marginal aggregate middle managerial risk-sharing premium. However, the risk-sharing premium in the case of common uncertainty grows much faster than it does in the case of independent outcomes.

The reason is that unlike risks of independent outcomes, the common uncertainty, being present across all managerial effort outcomes, cannot be diversified away through risk-sharing among managers. Consequently, the aggregate middle managerial risk-sharing premium can explode as the firm size increases. This implies that unless the top managerial effort productivity explodes even at a faster speed than the aggregate common risk-sharing premium can, the premium in the case of common uncertainty can become much more important to investors' expected wealth than it can in the case of independent outcomes.

In particular, we show that in the presence of common uncertainty, if the aggregate top managerial contribution to middle managerial outcomes is bounded, and the production technology of the firm does not exhibit too high a degree of increasing returns to scale in labor, then the sensitivity of the top managerial hierarchical contract is much lower than that of the top managerial direct contract, approaching zero as the firm size grows. The reason is that under the above stated conditions, the aggregate risk-sharing premium can grow with the

firm size so fast that investors are much more willing to sacrifice top managerial incentives in order to reduce the aggregate middle managerial risk-sharing premium than they are in the case of independent outcomes. Note that the condition on the firm’s production technology for this result is general enough to encompass Baker and Hall’s [2004] empirical finding that firms typically have decreasing returns to scale technologies.

We believe the above result can serve as an explanation of the puzzle presented by Jensen and Murphy [1990]: the average wealth changes of CEOs for large and small firms are, respectively, \$3.25 and about \$8.00 for every \$1,000 change in shareholder wealth.<sup>7</sup> Moreover, our result is also consistent with an intuition that as the firm size becomes larger, top managerial effort decisions can be less important than his decisions on middle managerial contracts, and thus investors would like to decrease the top managerial effort incentives in order to induce the top manager to make decisions more like investors in contracting middle managers.

The paper is organized as follows: In the next section we start with the general description of our hierarchical contracting environment. In Section 3, we consider a direct contracting benchmark where investors can directly contract all managers with full information about all individual managerial effort outcomes. For this benchmark, we extend Holmstrom and Milgrom [1987] with many agents. Section 4 is the main section of the paper. This section consists of two subsections: the first is a case where all managerial outcomes are independent of each other, and the other is a case where all outcomes are correlated with each other through common uncertainty of the firm. In Section 5, we compare both top and middle managerial sensitivities under the two different direct and hierarchical contracting environments. Finally, Section 6 provides a brief summary of results of the paper. All proofs are presented in the Appendix.

## 2 The Model

There are two dates, 0 and 1, and a firm with  $N + 1$  managers  $0, 1, \dots, N$  where manager 0 is the top manager and the rest are middle managers. Throughout the paper, we interpret  $N$

---

<sup>7</sup>Hall and Liebman (1999), Aggarwal and Samwick [2003], and Baker and Hall [2004] document higher sensitivities than Jensen and Murphy. However, the magnitude of these sensitivities are generally small and the firm-size effect still exists: the larger the firm, the lower the sensitivity. Another comment is in order. In the literature, using an implication from the standard one-principal and one-agent model, it is commonly argued that the sensitivities of large firms should be lower than those of small firms, because the volatilities of aggregate outcomes of large firms are larger than those of small firms. Unfortunately, this argument may not be correct because it ignores the firm size effect on the productivity of the CEO (the agent) of the firm. Since the CEO marginal productivity is positively related to the sensitivity, the standard model can also produce the opposite of the firm size effect if the volatility of the profit of the firm grows sufficiently slower than the productivity does as the firm size increases. However, given that it does not tell us why the volatility of the profit of the firm should grow sufficiently slower or faster than the productivity does as the firm size increases, the standard principal-agent model does not provide a satisfactory explanation of the firm-size effect.

synonymously with the firm size. Within the firm, there can be  $(N + 1)$  individual productions (outcomes)  $Y^i$ ,  $i = 0, 1, \dots, N$ , resulting in a total production of  $\sum_{i=0}^N Y^i$ . Middle manager  $i$  exerts effort  $\mu^i$  to increase  $Y^i$  at a personal (monetary) cost of  $c^i(\mu^i)$ . The top managerial effort  $\mu^0$ , however, can affect, at a cost of  $c^0(\mu^0)$ , not only his own outcome  $Y^0$  but all middle managerial outcomes. We assume that for all  $i = 0, 1, \dots, N$ ,  $c_\mu^i \geq 0$ ,  $c_{\mu\mu}^i > 0$ . Throughout the paper, we say cost functions are *quadratic* if  $c^i(\mu) = (K^i/2)\mu^2$ ,  $i = 0, 1, \dots, N$ . The individual outcomes have the following structures.

$$\begin{aligned} Y^0 &= \mu^0 N^{-\delta^0} + \sigma^c \xi^c + \sigma^0 \xi^0, \\ Y^i &= (\mu^i + \kappa^i(N)\mu^0) N^{-\delta^i} + \sigma^c \xi^c + \sigma^i \xi^i, \quad i = 1, \dots, N, \end{aligned}$$

where  $\kappa^i(N) \geq 0$  is a function of  $N$ ;  $\xi^c$  and  $\xi^i$  are independent identically distributed standard normal random variables; and  $\sigma^c \geq 0$ , and  $\sigma^i > 0$ , for  $i = 0, 1, \dots, N$ . The random variable  $\xi^c$  captures the source of risk common to all managers in the firm, and  $\xi^i$ 's are idiosyncratic risk sources unique to individual managers.

The parameter  $\delta^i$  represents a scale effect on the production function of the firm: if  $\delta^i > (=, <) 0$  for all  $i$ , each production function in the firm exhibits decreasing (constant, increasing, respectively) returns to scale in labor.<sup>8</sup> For simplicity, we assume that  $\delta^i$ 's,  $i = 0, 1, \dots, N$ , have the same signs across all managers, e.g., if  $\delta^0$  is positive, then all  $\delta^i$ 's,  $i = 1, \dots, N$  are positive. When  $\delta^i > 0$ , manager  $i$  would find it increasingly more difficult to increase the expected outcome for a given level of effort, as the firm size grows. In other words, if  $\delta^i$ 's are positive and the total number of employees is doubled with all effort levels held constant, then the total output will be less than doubled. The case of decreasing returns to scale may be the most reasonable description of the firm in the long run. Nonetheless, we shall examine all three cases, because some small firms can sometimes exhibit increasing returns to scale in the short run, or during some contracting periods.

The term  $\kappa^i(N)\mu^0 N^{-\delta^i}$ ,  $i = 0, 1, \dots, N$ , describes contribution by the top manager to middle manager  $i$ 's effort outcome  $Y^i$ , and thus the sum  $\sum_{i=1}^N \kappa^i(N) N^{-\delta^i} \mu^0$  is the top managerial contribution to the aggregate middle managerial output. This term plus  $\mu^0 N^{-\delta^0}$  captures in spirit the influence of the top managerial effort on the total output of the firm as in Rosen [1982]. We assume that the top managerial contribution to the aggregate middle managerial

---

<sup>8</sup>Scale effects in production functions in this paper can be equivalently interpreted as those in managerial cost functions, because the model specification can be equivalently transformed as follows.

$$\begin{aligned} Y^0 &= m^0 + \sigma^c \xi^c + \sigma^0 \xi^0, & Y^i &= m^i + \sigma^c \xi^c + \sigma^i \xi^i, \quad i = 1, \dots, N, \\ c^0 &= c^0(m^0 N^{\delta^0}), & c^i &= c^i(m^i N^{\delta^i} - \kappa^i(N)m^0), \quad i = 1, \dots, N. \end{aligned}$$

Under this equivalent specification, the top managerial effort helps reduce middle managerial costs of effort.

output increases and approaches a finite nonnegative number  $q$  as the firm size increases, i.e.,  $\sum_{i=1}^N \kappa^i(N)N^{-\delta^i} \uparrow q < \infty$ , as  $N \rightarrow \infty$ . This assumption implies that the marginal productivity of the top managerial effort to increase the aggregate middle managerial outcome cannot grow to infinity as the firm size grows.

Investors are risk neutral, and managers exhibit constant absolute risk aversion (CARA) with CARA coefficient  $R^i$ ,  $i = 0, 1, \dots, N$ . We assume the reservation utility of manager  $i$ , for  $i = 0, 1, \dots, N$ , is  $e^{-R^i W_0^i}$ . It is well accepted that investors exhibit risk neutrality as far as contracting is concerned, because risks relevant to contracting are firm-specific, whereas managers exhibit risk aversion, because significant portions of their wealth are exposed to firm-specific risks of the firm, and these exposures cannot be diversified away in capital markets. As will be seen later, top managerial risk aversion plays a particularly important role in hierarchical contracting.

We consider two different types of contracting problems. The first is the case of direct contracting where investors observe all individual outcomes, and directly contract each manager. The other is the hierarchical contracting case where investors can only observe the total outcome net of all middle managerial compensations, and contract the top manager who has technology to observe all individual outcomes and is given a mandate to contract all middle managers. We assume all managerial contracts are linear in observable outcomes of managerial efforts. Further details of admissible contracts shall be given in Sections 3 and 4. The linearity assumption is without loss of generality, as long as we interpret our results in the context of a continuous-time model described in the Appendix.

For brevity, we express managerial effort outcomes with vector  $Y$  as follows:

$$Y = e + D\xi,$$

where

$$\begin{aligned} Y &= \left[ Y^0 \ Y^1 \ \dots \ Y^N \right]^\top, \\ e &= \left[ \mu^0 N^{-\delta^0} \ (\mu^1 + \kappa^1(N)\mu^0)N^{-\delta^1} \ \dots \ (\mu^N + \kappa^N(N)\mu^0)N^{-\delta^N} \right]^\top, \\ \xi &= \left[ \xi^c \ \xi^0 \ \xi^1 \ \dots \ \xi^N \right]^\top, \end{aligned}$$

and  $D$  is an  $(N+1) \times (N+2)$  matrix such that

$$D = \begin{bmatrix} \sigma^c & \sigma^0 & 0 & \dots & \dots & 0 \\ \sigma^c & 0 & \sigma^1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma^c & \dots & \dots & \dots & 0 & \sigma^N \end{bmatrix}.$$



We often use notation  $Y_{-i}$  to denote an  $N$ -vector constructed from the  $(N+1)$ -vector  $Y$  with its  $i$ -th element removed. That is,

$$Y_{-i} = \left[ Y^0 \quad Y^1 \quad \dots \quad Y^{i-1} \quad Y^{i+1} \quad \dots \quad Y^N \right]^\top.$$

Throughout the paper, any vector with subscript  $-i$  means the vector with its  $i$ -th element removed.

On the other hand, the variance-covariance matrix of  $Y$ , denoted by  $Q$ , is an  $(N+1) \times (N+1)$  matrix such that

$$Q = DD^\top = \begin{bmatrix} (\sigma^c)^2 + (\sigma^0)^2 & (\sigma^c)^2 & \dots & \dots & (\sigma^c)^2 \\ (\sigma^c)^2 & (\sigma^c)^2 + (\sigma^1)^2 & (\sigma^c)^2 & \dots & (\sigma^c)^2 \\ \dots & \dots & \dots & \dots & \dots \\ (\sigma^c)^2 & \dots & \dots & (\sigma^c)^2 & (\sigma^c)^2 + (\sigma^N)^2 \end{bmatrix},$$

which is clearly positive definite. Define  $Q_{-i}$  to be an  $N \times N$  submatrix of  $Q$ , resulting from deleting the  $(i+1)$ -th row and column of  $Q$  for  $i = 0, 1, \dots, N$ . Then  $Q_{-i}$  is the variance-covariance matrix of  $Y_{-i}$  for  $i = 0, 1, \dots, N$ . Also define, for  $i = 0, 1, \dots, N$ , an  $(N+1)$ -vector  $p$  as follows:

$$p = \left[ Cov(Y^i, Y^0) \quad \dots \quad Cov(Y^i, Y^N) \right]^\top = \left[ (\sigma^c)^2 \quad \dots \quad (\sigma^c)^2 \right]^\top.$$

Later, it will be seen that quantity  $p_{-i}$  can be used in filtering common noise out of individual effort outcomes.

### 3 Direct Contracting with Complete Information

Before starting our discussion on our hierarchical contracting problem, we first examine a direct-contracting benchmark problem, the solution to which we will later compare with that of the hierarchical contracting problem.

In the benchmark problem, investors can observe all individual outcomes and directly sign contracts with all managers. Economic implications of such a case are in fact relatively thoroughly examined in the literature on moral hazard in teams consisting of one risk-neutral principal and many risk-neutral/risk-averse agents. We simply recover existing results with additional features which allow the top manager to affect outcomes of other managers, and all outcomes to depend on the firm size.<sup>9</sup> These features are to distinguish the top manager from middle managers, and enable us to better examine the firm size effect.

---

<sup>9</sup>Our companion paper, Koo, Shim, and Sung (2006) studies a case with a risk averse principal and production functions which allow general interaction among agents.

Since individual outcomes  $Y^i$ ,  $i = 0, 1, \dots, N$ , are observable, all salary functions are allowed to depend on all individual outcomes as follows:

$$S^i(Y) = \alpha^i + (\beta^i)^\top Y, \quad i = 0, 1, \dots, N,$$

where

$$\beta^i = \left[ \beta^{i0} \quad \beta^{i1} \quad \dots \quad \beta^{iN} \right]^\top, \quad i = 0, 1, \dots, N.$$

Let

$$\mu = \left[ \mu^0 \quad \mu^1 \quad \dots \quad \mu^N \right]^\top.$$

Then, the investors' problem is to choose  $S^i$ ,  $i = 0, 1, \dots, N$ , to maximize

$$E \left[ \mathbf{1}_{N+1}^\top Y - \sum_{i=0}^N S^i \right]$$

subject to, for  $i = 0, 1, \dots, N$ ,

$$\mu^i \in \arg \max_{\hat{\mu}^i} E[-\exp\{-R^i(S^i - c^i(\hat{\mu}^i))\} | \mu_{-i}],$$

and

$$E[-\exp\{-R^i(S^i - c^i(\mu^i))\} | \mu_{-i}] \geq -\exp\{-R^i \mathcal{W}_0^i\}.$$

Note that we look at Nash equilibria where each manager chooses his/her own effort level, as if effort levels of all other managers were exogenously given. Throughout the paper, we only focus on Nash equilibria.

Let

$$\begin{aligned} F &:= Q_{-0} - 2N^{\delta^0} (p_{-0})\theta^\top + [(\sigma^c)^2 + (\sigma^0)^2]N^{2\delta^0}\theta\theta^\top, \\ G &:= (p_{-0}) - [(\sigma^c)^2 + (\sigma^0)^2]N^{\delta^0}\theta, \\ \theta &:= \left[ \kappa^1(N)N^{-\delta^1} \quad \dots \quad \kappa^N(N)N^{-\delta^N} \right]^\top. \end{aligned}$$

Note that  $\theta$  is an  $N$ -vector of top managerial marginal effort-products through all middle managerial outcomes.

The following proposition provides the solution to our benchmark problem, and extends Holmstrom and Milgrom's [1987] model by allowing the principal to contract many agents.

**Proposition 1.** *Let  $(\mu^0, \mu^i)$  be the solution to the following equations: for  $i = 1, \dots, N$ ,*

$$\begin{aligned} & \left[ N^{-\delta^0} + \sum_{i=1}^N \kappa^i(N)N^{-\delta^i} \right] - c_\mu^0(\mu^0) \\ & - R^0 N^{2\delta^0} c_\mu^0(\mu^0) c_{\mu\mu}^0(\mu^0) \left[ (\sigma^c)^2 + (\sigma^0)^2 - G^\top F^{-1} G \right] = 0, \end{aligned} \quad (1)$$

$$N^{-\delta^i} - c_\mu^i(\mu^i) - R^i N^{2\delta^i} c_\mu^i(\mu^i) c_{\mu\mu}^i(\mu^i) \left[ (\sigma^c)^2 + (\sigma^i)^2 - (p_{-i})^\top (Q_{-i})^{-1} p_{-i} \right] = 0. \quad (2)$$

Then, the optimal contracts for the top and middle managers are as follows:

$$\begin{aligned}
S^0 &= \mathcal{W}_0^0 + N^{\delta^0} c_\mu^0(\mu^0) (F^{-1}G)^\top e_{-0} - \left(1 + N^{\delta^0} (F^{-1}G)^\top \theta\right) c_\mu^0(\mu^0) \mu^0 + c^0(\mu^0) \\
&\quad + \frac{R^0}{2} (N^{\delta^0} c_\mu^0(\mu^0))^2 \left[ (\sigma^c)^2 + (\sigma^0)^2 - G^\top F^{-1}G \right] \\
&\quad + N^{\delta^0} c_\mu^0(\mu^0) \left\{ Y^0 - (F^{-1}G)^\top (Y_{-0} - \theta N^{\delta^0} Y^0) \right\}, \tag{3}
\end{aligned}$$

$$\begin{aligned}
S^i &= \mathcal{W}_0^i + N^{\delta^i} c_\mu^i(\mu^i) (p_{-i})^\top (Q_{-i})^{-1} e_{-i} - c_\mu^i(\mu^i) (\mu^i + \kappa^i(N) \mu^0) + c^i(\mu^i) \\
&\quad + \frac{R^i}{2} (N^{\delta^i} c_\mu^i(\mu^i))^2 \left[ (\sigma^c)^2 + (\sigma^i)^2 - (p_{-i})^\top (Q_{-i})^{-1} p_{-i} \right] \\
&\quad + N^{\delta^i} c_\mu^i(\mu^i) \left\{ Y^i - (p_{-i})^\top (Q_{-i})^{-1} Y_{-i} \right\}. \tag{4}
\end{aligned}$$

Proposition 1 implies that in the presence of common noise, the optimal managerial performance measure for middle manager  $i$ ,  $i = 1, \dots, N$ , is  $Y^i - (p_{-i})^\top (Q_{-i})^{-1} Y_{-i}$ , where  $(p_{-i})^\top (Q_{-i})^{-1}$  is a vector of slope coefficients of the multiple regression of  $Y^i$  on  $Y_{-i}$ . In particular, the vector of slopes is constructed in such a way that the resulting volatility of the performance measure is minimized.<sup>10</sup> On the other hand, the top managerial performance measure is  $Y^0 - (F^{-1}G)^\top (Y_{-0} - \theta N^{\delta^0} Y^0)$ , which also minimizes the volatility of the performance measure.<sup>11</sup> Alternatively,  $(F^{-1}G)^\top$  can be viewed as a vector of regression slope coefficients when  $Y^0$  is regressed on  $Y_{-0} - \theta N^{\delta^0} Y^0$ . Note that the vector  $Y_{-0} - \theta N^{\delta^0} Y^0$  is independent of  $\mu^0$  and it is used to filter a part of the common noise out of  $Y^0$ .

The above implies that when correlated signals for managerial effort are available, the signals can be utilized to reduce the volatility of the managerial performance measure. If the signals have no information content about managerial effort, that is, if all outcomes are independent such that  $\sigma^c = 0$ , and  $\kappa^i(N) = 0$ , then each manager's salary depends on his/her own outcome alone.

<sup>10</sup>That is,  $[(\sigma^c)^2 + (\sigma^i)^2 - (p_{-i})^\top (Q_{-i})^{-1} p_{-i}]$  is the minimum value of  $x^\top DD^\top x$ , where  $x$  is an arbitrary  $(N+1)$ -vector with its  $(i+1)$ -th component being one. Since  $DD^\top$  and  $(Q_{-i})^{-1}$  are positive definite, and since  $p_{-i}$  is a nonzero vector, we have  $(p_{-i})^\top (Q_{-i})^{-1} p_{-i} > 0$ ,  $i = 1, \dots, N$ , which implies that

$$(\sigma^c)^2 + (\sigma^i)^2 > (\sigma^c)^2 + (\sigma^i)^2 - (p_{-i})^\top (Q_{-i})^{-1} p_{-i} > 0, \quad i = 1, \dots, N.$$

<sup>11</sup> For an arbitrary  $(N+1)$ -vector  $x = [x_0 \ x_1 \ \dots \ x_N]^\top$ , and  $x_{-0} = [x_1 \ \dots \ x_N]^\top$ ,  $(\sigma^c)^2 + (\sigma^0)^2 - G^\top F^{-1}G$  is the minimum value of  $x^\top DD^\top x$  subject to  $x_0 = 1 - N^{\delta^0} (x_{-0})^\top \theta$ . Since  $DD^\top$  and  $F^{-1}$  are positive definite, and since  $G$  is a nonzero vector, we have

$$(\sigma^c)^2 + (\sigma^0)^2 > (\sigma^c)^2 + (\sigma^i)^2 - G^\top F^{-1}G > 0.$$

Furthermore, it can be shown that all of the above performance measures both for the top and middle managers are sufficient statistics for their effort levels. See Koo, Shim and Sung [2006] for a general proof. Consequently, the performance measures in this section are consistent with Holmstrom's [1982] sufficient statistics result.

In the next section, we shall examine the case of hierarchical contracting, assuming identical middle managers, and then in Section 5, we will compare sensitivities of optimal hierarchical contracts with those of direct contracts. In order to facilitate the comparison, we simplify Proposition 1 with identical middle managers as in the following corollary.

**Corollary 1.** *Suppose all middle managers are identical in the following sense:*

$$R^i \equiv R^M, \quad c^i(\cdot) \equiv c^M(\cdot), \quad \sigma^i \equiv \sigma^M, \quad \mathcal{W}_0^i \equiv \mathcal{W}_0^M, \quad \delta^i \equiv \delta^M, \quad \kappa^i(N) \equiv \kappa(N), \quad i = 1, \dots, N.$$

*Then, the optimal contracts for the top and middle managers are given as in (3) and (4), respectively, with*

$$(F^{-1}G)^\top = \frac{(\sigma^c)^2 - [(\sigma^c)^2 + (\sigma^0)^2]\kappa(N)N^{\delta^0 - \delta^M}}{(\sigma^M)^2 + N[(\sigma^c)^2(1 - \kappa(N)N^{\delta^0 - \delta^M})^2 + (\sigma^0)^2(\kappa(N)N^{\delta^0 - \delta^M})^2]} \mathbf{1}_N^\top$$

$$(p_{-i})^\top (Q_{-i})^{-1} = \frac{(\sigma^c)^2}{(\sigma^M)^2[(\sigma^c)^2 + (\sigma^0)^2] + (N-1)(\sigma^c)^2(\sigma^0)^2} \begin{bmatrix} (\sigma^M)^2 & (\sigma^0)^2 & \dots & (\sigma^0)^2 \end{bmatrix},$$

and  $\mu^0$  and  $\mu^M$  satisfy

$$\begin{aligned} & [N^{-\delta^0} + \kappa(N)N^{1-\delta^M}] - c_\mu^0(\mu^0) - R^0 N^{2\delta^0} c_\mu^0(\mu^0) c_{\mu\mu}^0(\mu^0) \\ & \times \frac{[(\sigma^c)^2 + (\sigma^0)^2](\sigma^M)^2 + N(\sigma^0)^2(\sigma^c)^2}{(\sigma^M)^2 + N[(\sigma^c)^2(1 - \kappa(N)N^{\delta^0 - \delta^M})^2 + (\sigma^0)^2(\kappa(N)N^{\delta^0 - \delta^M})^2]} = 0 \end{aligned} \quad (5)$$

$$\begin{aligned} & N^{-\delta^M} - c_\mu^M(\mu^M) - R^M N^{2\delta^M} c_\mu^M(\mu^M) c_{\mu\mu}^M(\mu^M) \\ & \times \frac{(\sigma^M)^2 \left( (\sigma^M)^2 [(\sigma^c)^2 + (\sigma^0)^2] + N(\sigma^c)^2(\sigma^0)^2 \right)}{(\sigma^M)^2[(\sigma^c)^2 + (\sigma^0)^2] + (N-1)(\sigma^c)^2(\sigma^0)^2} = 0. \end{aligned} \quad (6)$$

From the above corollary, one can double check that  $F^{-1}G$  and  $((p_{-i})^\top Q_{-i})^{-1}$  are vectors of multiple regression coefficients, respectively, from regressing  $Y^0$  on  $Y_{-0} - \mathbf{1}_N \kappa(N) N^{\delta^0 - \delta^M} Y^0$ , and from regressing  $Y^i$  on  $Y_{-i}$ .

## 4 Hierarchical Contracting

In this section, for simplicity, we assume that all middle managers are identical with  $R^i \equiv R^M$ ,  $c^i(\cdot) \equiv c^M(\cdot)$ ,  $\sigma^i \equiv \sigma^M$ ,  $\mathcal{W}_0^i \equiv \mathcal{W}_0^M$ ,  $\delta^i \equiv \delta^M$ ,  $\kappa^i(N) \equiv \kappa(N)$ ,  $i = 1, \dots, N$ . At

time zero, investors and the top manager sign a compensation contract  $S^0$ , and then the top manager recruits  $N$  middle managers with contract  $S^i$ ,  $i = 1, \dots, N$ . After all contracts are signed, both the top and middle managers together carry out  $N + 1$  tasks.

Both the top and middle managers can observe all  $Y^i$ 's, outcomes of individual tasks. When  $Y$  is realized, middle manager  $i$  is paid  $S^i$  based on  $Y$ , i.e.,  $S^i = \alpha^i + (\beta^i)^\top Y$ , for  $i = 1, \dots, N$ . The top manager reports to investors the net profit,  $X$ , which is the total production net of middle managerial salaries, i.e.,  $X = \mathbf{1}_{N+1}^\top Y - \sum_{i=1}^N S^i$ . Then the top manager receives compensation  $S^0$  based on the net profit,  $X$ , i.e.,  $S^0(X) = A + BX$ . Investors claim the rest  $\mathbf{1}_{N+1}^\top Y - S^0 - \sum_{i=1}^N S^i$ . We assume investors cannot monitor the structure of  $S^i$ ,  $i = 1, \dots, N$  without incurring a prohibitably high cost.<sup>12</sup> Thus the top manager is in effect given complete discretion to design  $S^i$ ,  $i = 1, \dots, N$ .<sup>13</sup>

We first consider the case where agents' outcomes are independent in Subsection 4.1, and then generalize it with correlated outcomes in Subsection 4.2. Although the correlated outcome case can be more realistic, it is algebraically more complex than the independent case. The independent outcome case enables us to see main economics of our hierarchical contracts much clearly and help entangle economics hidden behind complex algebraic equations for the correlated outcome case.

#### 4.1 Independent Outcomes

In this subsection, we assume that all outcomes are independent, that is,  $\sigma^c = 0$ . In our hierarchical world, the top and middle managerial contracts cannot be determined independently of each other, because the top managerial contract necessarily affects middle managerial contracts. However, for ease of exposition, we first present necessary forms of middle managerial contracts in Proposition 2 and then we do the same for the top managerial contract in Proposition 3. Proofs of both propositions are however omitted because they are special cases of Propositions 4 and 5 with  $\sigma^c = 0$ .

Let us first define  $\beta$ ,  $\gamma^0$  and  $\gamma^M$  as follows:

$$\beta = N^{\delta^M} c_\mu^M(\mu^M), \quad \gamma^0 = \frac{R^0 B}{NR^0 B + R^M}, \quad \text{and} \quad \gamma^M = \frac{R^0 B(1 - \beta)}{(N - 1)R^0 B + R^M},$$

---

<sup>12</sup>In this paper we do not explicitly model the monitoring costs that the top manager should incur in order to observe individual middle managerial performances. It appears that monitoring costs can be related to the optimal hierarchical structure of the firm, and we leave this potentially interesting issue of optimal hierarchy for future research.

<sup>13</sup>Even without invoking high monitoring costs, the hierarchical contracting may be rationalized if investors are uncertain about middle managers' ability levels, but after the top managerial contract is signed, the top manager can identify their ability levels with higher precisions than investors can. However, we do not model adverse selection problems here.

which then become sensitivities for each middle managerial contract, respectively, to his/her own outcome, the top managerial outcome and all other middle managerial outcomes, as can be seen in the following proposition.

**Proposition 2.** *Suppose that the top managerial contract  $S^0(X) = A + BX$  is given. Under optimal middle managerial contracts, all middle managers exert the same levels of effort such that  $\mu^i = \mu^j = \mu^M$ , for  $i, j = 1, \dots, N$ , and each optimal middle managerial contract is given in the following form: For  $i = 1, \dots, N$ ,*

$$S^i(Y) = \mathcal{W}_0^M - (\beta^i)^\top e + c^M(\mu^M) + \frac{R^M}{2} \left\{ (\beta^{i0})^2(\sigma^0)^2 + \sum_{j=1}^N (\beta^{ij})^2(\sigma^M)^2 \right\} + (\beta^i)^\top Y, \quad (7)$$

where  $\beta^i$  is an  $(N + 1)$ -vector consisting of  $\beta^{i0} = \gamma^0$ ,  $\beta^{ii} = \beta$ , and  $\beta^{ij} = \gamma^M$ , for  $j = 1, \dots, N$  and  $j \neq i$ .

Proposition 2 implies that although individual outcomes are independent of each other, each middle managerial salary function  $S^i$  can depend on three metrics:  $Y^i$ ,  $Y^0$ , and  $\sum_{j=1, j \neq i}^N Y^j$  which are individual  $i$ 's own, the top managerial and the aggregate of all other middle managerial outcomes. Note that  $Y^i$  is a sufficient statistic for middle manager  $i$ 's effort level, and that the two other metrics are non-sufficient statistics that are orthogonal to the sufficient statistic. The sufficient statistic is obviously to give the middle manager incentives to work, whereas the other two non-sufficient statistics are for risk-sharing.

If the top manager were risk neutral, i.e., if  $R^0 = 0$ , then  $\gamma^0 = \gamma^M = 0$ , and the middle managerial contract would be independent of the other two metrics,  $Y^0$ , and  $\sum_{j=1, j \neq i}^N Y^j$ . However, when the top manager is risk-averse, he is concerned not only with middle managerial incentives but also with shifting a part of his compensation risk to middle managers. The latter risk-sharing need motivates the top manager to require that each middle managerial compensation also depend on both the top managerial and aggregate outcomes, although the two metrics have nothing to do with the middle manager effort incentives. *Therefore, under hierarchical contracting, even when the principal is risk neutral and all managerial outcomes are independent, the optimal performance measure for each middle managerial contract consists of not only a sufficient statistic but a non-sufficient statistic that is orthogonal to the sufficient statistic.*

Alternatively, one may rewrite  $S^i$  as follows:

$$S^i = \alpha^M + (\beta - \gamma^M)Y^i + (\gamma^0 - \gamma^M)Y^0 + \gamma^M \sum_{j=0}^N Y^j, \quad (8)$$

where  $\alpha^i = \alpha^j = \alpha^M$  for  $i, j = 1, \dots, N$ . Under this alternative expression,  $S^i$  depends on a new set of three metrics: individual  $i$ 's own, the top managerial, and the aggregate of all outcomes, i.e.,  $Y^i$ ,  $Y^0$ , and  $\sum_{j=0}^N Y^j$ , respectively. Even when it may be very weakly correlated with agent  $i$ 's performances, the aggregate economic profit of the firm before all salaries,  $\sum_{j=0}^N Y^j$ , is included in the optimal performance measure for agent  $i$ , in order to improve risk-sharing. Consequently, the expression in (8) may provide a rationale to justify common managerial compensation practices such as company-wide bonus programs including stock option plans for nonexecutive employees. See Core and Guay [2001] for empirical evidence of stock option plans for nonexecutive employees.

Sharing risk with middle managers brings about both a benefit and a cost to the top manager: a marginal decrease in the top managerial compensation risk and a marginal increase in the aggregate middle managerial compensation-risk premium. (We will shortly discuss this top managerial cost-benefit tradeoff from investors' perspective.) From (7), note that each middle managerial compensation risk premium consists of two terms:

$$\frac{R^M}{2}\beta^2(\sigma^M)^2 + \frac{R^M}{2} [(\gamma^0)^2(\sigma^0)^2 + (N-1)(\gamma^M)^2(\sigma^M)^2]. \quad (9)$$

The first term of the above quantity is a usual risk premium associated with middle managerial effort performance risk, and the second is an individual middle managerial compensation-risk premium attributed to risk-sharing. We call the first term *the middle managerial (effort) performance risk premium* and the second *the middle managerial risk-sharing premium*.

It is noteworthy that roles of risk-sharing are mostly ignored in the standard agency literature, partly because shareholders of a firm are typically assumed to be well-diversified and thus exhibit risk-neutrality over idiosyncratic risk such as contracting risk, and partly because risk-sharing issues are considered as well-understood in the neoclassical world. However, in a hierarchical contracting environment, the risk-sharing issue reemerges in a nontrivial way, because although shareholders may be risk neutral, the top manager exhibits risk aversion over idiosyncratic risk of the firm and acts as a principal to middle managers. Thus, in contracting middle managers, it is important for the top manager to take into account not only middle managerial incentives but also sharing risk with them. In other words, the top manager is willing to pay middle managers risk-sharing premia for their sharing his own compensation risk. Of course, the cost of the top managerial risk-sharing incentive or the middle managerial risk-sharing premia will eventually be shifted to shareholders. Thus, although they are risk neutral, shareholders, in turn, have to consider middle managerial risk-sharing premia in designing the top managerial compensation contract.

In this paper, it is this top managerial risk-sharing motivation that complicates both top and middle managerial incentive problems. Now, let us look at the top managerial contract.

**Proposition 3.** *The optimal top managerial contract  $S^0(X)$  under the hierarchical contracting is given in the following form.*

$$S^0(X) = \mathcal{W}_0^0 + c^0(\mu^0) - BE[X] + \frac{R^0}{2}B^2Var(X) + BX,$$

where

$$B = \frac{c_\mu^0(\mu^0)}{N^{-\delta^0} + \kappa(N)N^{1-\delta^M}} \quad (10)$$

$$E[X] = (N^{-\delta^0} + \kappa(N)N^{1-\delta^M})\mu^0 + N \left[ N^{-\delta^M} \mu^M - \mathcal{W}_0^M - c^M(\mu^M) - \frac{R^M}{2} \left\{ (\gamma^0)^2 (\sigma^0)^2 + (\beta^2 + (N-1)(\gamma^M)^2) (\sigma^M)^2 \right\} \right], \quad (11)$$

$$Var(X) = \left( \frac{R^M}{NR^0B + R^M} \right)^2 (\sigma^0)^2 + N \left( \frac{(1-\beta)R^M}{(N-1)R^0B + R^M} \right)^2 (\sigma^M)^2, \quad (12)$$

and  $\mu^0$  and  $\mu^M$  satisfy

$$N^{-\delta^M} - c_\mu^M(\mu^M) - R^M N^{2\delta^M} c_\mu^M(\mu^M) c_{\mu\mu}^M(\mu^M) (\sigma^M)^2 + R^0 B \left( \frac{R^M (1-\beta) N^{\delta^M} c_{\mu\mu}^M(\mu^M) (\sigma^M)^2}{(N-1)R^0B + R^M} \right) = 0. \quad (13)$$

Condition (13) immediately implies  $0 < \beta (= \beta^{ii} = N^{\delta^M} c_\mu^M) < 1$ . This condition is from one of the first order conditions (FOCs) of the top managerial expected utility maximization given a top managerial contract with sensitivity  $B$ , subject to middle managerial incentive and participation constraints. Thus, conditions (10) and (13) completely describe the top managerial decision rule on his own effort choice and middle managerial effort incentives.

The first three terms of the LHS of condition (13) are familiar terms that appear when investors directly contract each middle manager.<sup>14</sup> The three terms are related to a marginal change in  $E[X]$  with respect to a marginal increase in  $\beta$ . The fourth term, representing a hierarchical contracting effect, is associated with a marginal reduction in the top managerial compensation risk burden (premium) on  $X$ , as the top manager increases each middle managerial effort incentive  $\beta$ . (Equation (12) implies that the top manager can shift his compensation

<sup>14</sup>For the three terms under direct contracting, see Eq.(2) with  $\sigma^c = 0$ . Assuming identical middle managers, Eq.(2) becomes  $N^{-\delta^M} - c_\mu^M(\mu^M) - R^M N^{2\delta^M} c_\mu^M(\mu^M) c_{\mu\mu}^M(\mu^M) (\sigma^M)^2 = 0$ .



risk by increasing middle managerial contract sensitivities.) *It is this marginal reduction that motivates the top manager to shift his own compensation risk to middle managers by increasing middle managerial contract sensitivities more than desired by investors, and as a result, middle managers work harder than desired by investors under direct contracting.*

The more risk averse the top manager, the above effect is more pronounced. Indeed Equation (13) implies that the higher the top managerial risk aversion  $R^0$ , the higher the middle managerial effort incentive  $\beta$ . With the same level of top managerial effort incentive  $B$ , the variance of total output after payment of middle managerial salaries is smaller for a more risk averse top manager. (See Equation (11).)

Taking top-managerial effort and risk-sharing incentives into account, investors design a top managerial contract by choosing  $\mu^0$  and  $\mu^M$  to maximize the following expected net profit subject to (13):

$$\begin{aligned}
& E[X - S^0(X)] \\
&= N \left[ N^{-\delta^M} \mu^M - \mathcal{W}_0^M - c^M(\mu^M) - \frac{R^M}{2} \left\{ \left( \frac{R^0 B}{NR^0 B + R^M} \right)^2 (\sigma^0)^2 \right. \right. \\
&\quad \left. \left. + \left( \beta^2 + (N-1) \left( \frac{R^0 B(1-\beta)}{(N-1)R^0 B + R^M} \right)^2 \right) (\sigma^M)^2 \right\} \right] \\
&\quad + (N^{-\delta^0} + \kappa(N)N^{1-\delta^M})\mu^0 - \mathcal{W}_0^0 - c^0(\mu^0) \\
&\quad - \frac{R^0}{2} B^2 \left[ \left( \frac{R^M}{NR^0 B + R^M} \right)^2 (\sigma^0)^2 + N \left( \frac{(1-\beta)R^M}{(N-1)R^0 B + R^M} \right)^2 (\sigma^M)^2 \right], \quad (14)
\end{aligned}$$

where

$$B = \frac{c_\mu^0(\mu^0)}{N^{-\delta^0} + \kappa(N)N^{1-\delta^M}} \quad \text{and} \quad \beta = N^{\delta^M} c_\mu^M(\mu^M).$$

The quantity inside the first square bracket in the RHS of (14) is the net profit from each middle managerial contract, and the other terms all together are the net profit from the top managerial contract. However, the above maximand is highly nonconcave.

Propositions 2 and 3 shed light on the investors' problem stated in (14), in terms of the top managerial responses to investors' decision on  $B$ . A few implications of the two propositions are summarized in the following corollary.

**Corollary 2.** *Assume  $c_{\mu\mu}^M = K^M$ , a constant. Holding other things constant, if the sensitivity  $B$  of  $S^0$  is increased, Propositions 2 and 3 imply that the top manager responds by changing sensitivities of middle managerial contracts as follows:*

$$\frac{\partial \beta}{\partial B} > 0, \quad \frac{\partial \gamma^0}{\partial B} > 0, \quad \frac{\partial \gamma^M}{\partial B} > 0, \quad \text{and} \quad \frac{\partial \text{Var}(X)}{\partial B} < 0.$$

Corollary 2 implies that as the top-managerial sensitivity  $B$  is increased, the top manager shifts more of his own compensation risk to middle managers by increasing middle managerial contract sensitivities  $\gamma^0$ ,  $\gamma^M$ , and  $\beta$ , and as a result the volatility of the top managerial performance measure decreases. However, increasing the top managerial sensitivity introduces two conflicting effects on the investors' expected wealth: (1) a marginal effort efficiency gain from top managerial contracting, because the top managerial compensation risk declines as the top manager shifts his own compensation risk to middle managers; and (2) a marginal aggregate middle managerial contract efficiency loss, because middle managerial contract sensitivities are excessively increased by the top manager. The marginal efficiency loss consists of both a marginal increase in the aggregate middle managerial risk-sharing premium and an excessive marginal increase in the aggregate middle managerial effort level.<sup>15</sup> Thus, it is reasonable to conjecture that if the top managerial efficiency gain dominates (is dominated by) the aggregate middle managerial efficiency loss, then the top managerial sensitivity would be higher (lower) than that of the direct contracting case discussed in Section 3. Moreover, the tradeoff between the two effects can be affected by the firm size.

Thus, the main question of this paper is: What would be the effect of hierarchical contracting on the top managerial sensitivity, as the firm size increases? Would the positive effect eventually dominate the other, or the other way around?

In order to obtain an insight into the above questions, we simplify the investors' problem using quadratic cost functions. Let us suppose that

$$c^0(\mu^0) = \frac{K^0}{2}(\mu^0)^2, \quad \text{and} \quad c^M(\mu^M) = \frac{K^M}{2}(\mu^M)^2. \quad (15)$$

Then, Eq.(13), the main constraint to the investors' problem, can be rewritten as follows.

$$1 - \beta - a\beta + \left( \frac{aB(1 - \beta)}{(N - 1)B + r} \right) = 0, \quad (16)$$

where

$$a := R^M N^{2\delta^M} K^M (\sigma^M)^2, \quad b := a + 1, \quad \text{and} \quad r := \frac{R^M}{R^0}.$$

Hence, we have

$$1 - \beta = \frac{a((N - 1)B + r)}{(bN - 1)B + rb}, \quad \text{or} \quad \beta = \frac{(N - 1 + a)B + r}{(bN - 1)B + rb}. \quad (17)$$

---

<sup>15</sup>Constraint (13) implies the excessive marginal increase occurs because the top manager always tries to induce the middle managerial effort levels higher than desired by investors in the direct contracting case.

Using (16) and (17), the investors' problem in (14) can be simplified as follows: choose  $B$  to maximize

$$\begin{aligned}
\Phi(B; N) = & \frac{N}{2bN^{2\delta^M} K^M} - \frac{a^2 N}{2bN^{2\delta^M} K^M} \left( \frac{aB}{(bN-1)B+rb} \right)^2 \\
& + \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} \left( B - \frac{B^2}{2} \right) \\
& - \frac{R^M}{2} \left[ (N+r) \left( \frac{B}{NB+r} \right)^2 (\sigma^0)^2 \right. \\
& \left. + N(N-1+r) \left( \frac{aB}{(bN-1)B+rb} \right)^2 (\sigma^M)^2 \right]. \tag{18}
\end{aligned}$$

Even after the simplification with quadratic cost functions as above, the investors' problem still is highly complex, and may have multiple local maxima. The following lemma tells us a possible range where the global maximum lies.

**Lemma 1.** *Let  $B^* \in \arg \max_B \Phi(B, N)$ . Then  $B^* \in (0, 1)$ .*

Lemma 1 implies that the optimal top managerial sensitivity is between zero and one, with one being the first best sensitivity. That is,  $\Phi(B^*, N)$  should be greater than the larger of the two,  $\Phi(0, N)$  or  $\Phi(1, N)$ . We investigate the objective function  $\Phi$  for a large  $N$ , or for a large firm. Later in Theorem 1, we shall show that as the firm size grows, the top managerial sensitivity under our hierarchical contracting approaches either of two extremes: either as low as zero or as high as the first best.

In fact, a close look at Eq.(18) reveals that investors can derive profits from two sources: one from the top manager and the other from middle managers. However, the two sources are intertwined in such a way that an increase in  $B$  can improve the first source of profits, but it can hurt the second source. The reason is that an increase in  $B$  motivates the top manager not only to work harder but to shift his own compensation risk to middle managers. In order to shift his risk, he could provide middle managers with excessively higher-powered incentives than desired by investors, to the extent that a marginal decrease in the net aggregate profit to investors from middle managerial effort exceeds a marginal increase in the net gain from the top managerial effort. Thus, in choosing  $B$ , investors try to strike a delicate balance between the two profit sources.

Let  $d^* > 0$  be the unique positive solution to

$$K^0(K^M)^2 d^3 - (1+q)^2[(K^M)^2 d^2 + 2K^M d + 1] = 0. \tag{19}$$

As can be seen Theorem 1,  $d^*$  sometimes becomes an important quantity for investors to determine priorities between the two profit sources. In particular,  $d^*$  is positively related to  $q$ , the top managerial maximal marginal influence on the aggregate middle managerial effort outcome.<sup>16</sup> Here is one of the two main theorems of the paper.<sup>17</sup>

**Theorem 1.** *Assume quadratic cost functions as in (15), and let  $B_N^* \in \arg \max_B \Phi(B; N)$ .*

*I. Limits of  $B_N^*$  are as follows.*

- i. If  $\delta^i > 0$  for  $i = 0, M$ , and  $K^0 R^M (\sigma^M)^2 - q^2 > 0 (< 0, \text{ resp.})$ , then the optimal top managerial sensitivity approaches zero (the first best, resp.), i.e.,  $B_N^* \rightarrow 0 (1, \text{ resp.})$ , as  $N \rightarrow \infty$ .*
- ii. If  $\delta^i = 0$  for  $i = 0, M$ , and  $R^M (\sigma^M)^2 - d^* > 0 (< 0, \text{ resp.})$ , then the optimal top managerial sensitivity approaches zero (the first best, resp.).*
- iii. If  $\delta^i < 0$  for  $i = 0, M$ , then the optimal top managerial sensitivity approaches the first best.*

*II. As the firm size grows, the optimal middle managerial sensitivity is always higher than, but approaches that of the directing contracting case.*

Part II of Theorem 1 results from comparing constraint (13) for optimal hierarchical contracts with (6) for optimal direct contracts. The simple intuition is that the top manager would like to minimize the aggregate middle managerial risk-sharing premium by designing middle managerial contract sensitivities such that the portion of the top managerial compensation risk shifted to each middle manager is diversified away in the limit. Because of this diversification effect, individual middle managers under hierarchical contracts demand zero risk-sharing premia and their effort levels in the limit converge to that of direct contracts.

As one may have expected, Part I-iii is consistent with the intuition that when production functions exhibit increasing returns to scale in labor, the optimal sensitivity of the top managerial contract approaches the first best as the firm size grows. With increasing returns to scale, the profit from aggregate middle managerial production can be more than doubled whenever the number of middle managers doubles. However, one may imagine that, in the real world, technologies of most firms may be realistically characterized by decreasing returns to scale, although those of some fast growing small firms may temporarily exhibit constant or

<sup>16</sup>More precisely, one can show  $d^*$  is increasing in  $q$  and decreasing in  $K^0$  and  $K^M$ .

<sup>17</sup>Both Theorems 1 and 2 are about managerial contract sensitivities in the limit as the firm size grows, proofs of the two are somewhat different from each other, because in Theorem 2, common uncertainty creates important terms in risk-premium parts that can affect investors' contract decisions significantly even in the limit.

even increasing returns to scale. In their model setting that is somewhat different from ours, Baker and Hall [2004, JLE] empirically show that the elasticity of managerial effort productivity with respect to firm size is 0.4, which can be approximately interpreted as our  $\delta^i$  being 0.6, i.e., decreasing returns to scale.

It is striking that Parts I-i and I-ii suggest when a very large firm has production technologies with nonincreasing returns to scale, the top managerial sensitivity can be either very low or very high, depending upon magnitudes of parameters such as  $R^M(\sigma^M)^2$ ,  $K^0$  and  $K^M$ . The sensitivity in the limit can be as low as zero if either of the following two conditions holds: (1) decreasing returns to scale and  $K^0 R^M(\sigma^M)^2 > q^2$ ; or (2) constant returns to scale and  $R^M(\sigma^M)^2 > d^*$ . Let us call these two conditions *the top managerial zero-sensitivity conditions*. On the other hand, the sensitivity in the limit can be as high as the first best if either of the following two conditions holds: (1) decreasing returns to scale and  $K^0 R^M(\sigma^M)^2 < q^2$ ; or (2) constant returns to scale and  $R^M(\sigma^M)^2 < d^*$ . We call these two *the top managerial first-best-sensitivity conditions*.

It turns out that the inequalities in the above conditions are mainly from comparing sizes of two quantities: rates of increase over  $N$  in aggregate middle managerial risk-sharing premium (recall from (9) the definition of the middle managerial risk-sharing premium), and in total top managerial effort contribution to all managerial effort outcomes including his own. Theorem 1 suggests that, in the limit, the former quantity is greater (less) than the latter if the top managerial zero-sensitivity (first-best-sensitivity) conditions hold. Thus, roughly speaking, Parts I-i and I-ii tell us that *if the managerial production functions exhibit nonincreasing returns to scale and if the top managerial effort contribution to all managerial effort outcomes including his own is less (more) important than the aggregate middle managerial risk-sharing premium, the top managerial sensitivity approaches zero (the first best) as the firm size increases*.<sup>18</sup>

Let us look at more closely how the tradeoff between aggregate middle managerial risk-sharing premium and top managerial effort contribution can arise in explaining the above two extreme limits of the top managerial sensitivity. As compared with the case of direct contracting, our hierarchical contracting problem calls for several additional factors in determining the top managerial contract, such as marginal changes in effort incentives and compensation-risk premia for both the top and middle managers, as the top manager tries to shift his compensation risk to middle managers.

Recall from (9) that each middle managerial compensation risk premium consists of two parts: effort-performance risk, and risk-sharing premium. The effort performance-risk premium

---

<sup>18</sup>If the managerial production functions exhibit increasing returns to scale, then the total top managerial contribution can easily more than offset the middle managerial aggregate risk-sharing premium, and thus the top managerial sensitivity approaches the first best as the firm size increases.

is a component which also appears in the case of direct contracting. It is true that as the firm size increases, each middle managerial risk-sharing premium can become less and less important because of the diversification effect, and as a result, middle managerial effort incentives converge to those of direct contracting. However, although individual risk-sharing premia approach zero, the aggregate middle managerial risk-sharing premium, for sufficiently large  $N$ , may not always decrease to zero.<sup>19</sup> Furthermore, it can be shown that marginal aggregate middle managerial risk-sharing premium dominates in the limit marginal decreases in top managerial risk premium due to risk-sharing. Thus, this aggregate premium could still be an important cost for hierarchical contracting.

On the other hand, as mentioned before, a positive side of the aggregate middle managerial risk-sharing is that it can improve top managerial effort incentives, as it reduces the top managerial compensation risk. Furthermore, it can be shown that the marginal gain from this positive side dominates in the limit marginal aggregate middle managerial effort efficiency losses due to excessively high powered incentives. Consequently, as  $N$  increases sufficiently large, investors' decision on  $B$  boils down to a tradeoff between marginal aggregate middle managerial risk-sharing premium, and marginal total top managerial effort contribution. This tradeoff is also consistent with Rosen's (1982) intuition: top managerial effort makes an important contribution to the total output of the firm and hence the top managerial effort/decision contribution is expected to play a role in investors' decisions.

In the following subsection, we shall see that perhaps even more strikingly, when common uncertainty is present in all managerial outcomes, the middle managerial risk-sharing premium could grow far faster in the limit than the top managerial effort contribution, it is much more likely that optimal top managerial contract sensitivity approaches zero than it is in the case of independent outcomes.

## 4.2 Correlated Outcomes with Common Uncertainty

In this subsection, we consider the case where  $\sigma^c > 0$ . The fundamental difference between this common uncertainty case from the independent outcome case is that compensation risk caused by common uncertainty cannot be diversified away through risk-sharing within the firm.<sup>20</sup> Because of this common-uncertainty risk, the aggregate middle managerial risk-sharing

---

<sup>19</sup>It can be shown that under optimal hierarchical contracts, the aggregate risk-sharing premium approaches a positive amount as  $N \rightarrow \infty$ , either if  $\delta^i > 0$ ,  $i = 0, M$ , and  $K^M(\sigma^M)^2 < q^2$  or if  $\delta^i = 0$ ,  $i = 0, M$ , and  $K^M(\sigma^M)^2 < d^*$ .

<sup>20</sup>The common uncertainty risk in this paper still is a part of firm-specific risk. Although it is non-diversifiable within the firm, the common uncertainty risk may be diversifiable in capital markets. The common uncertainty may be viewed as company-wide firm-specific shocks. For example, operations of the whole firm can be affected by changes in market demand for products of the firm as a consequence of market competition against other competing firms, when the size of the market segment for similar products is more or less fixed. As a result, each

premium can explode as the firm size grows.

Let us introduce the following new notation:

$$\eta := \frac{(\sigma^c)^2}{(\sigma^c)^2 + (\sigma^0)^2}, \quad \text{and} \quad \varphi := \eta(\sigma^0)^2.$$

Then,  $\eta$  is the slope of regression of  $Y^i$  on  $Y^0$ , for  $i = 1, \dots, N$ . Moreover, we redefine  $\beta$ ,  $\gamma^0$  and  $\gamma^M$  as follows.

$$\beta = N^{\delta^M} c_\mu^M(\mu^M), \quad (20)$$

$$\gamma^0 = \frac{R^0 B - R^M \eta}{NR^0 B + R^M} + \eta(1 - \beta - (N-1)\gamma^M), \quad (21)$$

$$\gamma^M = \frac{\{R^0 B(N\varphi + (\sigma^M)^2) + R^M \varphi\}(1 - \beta) - R^M \varphi}{(N-1)R^0 B(N\varphi + (\sigma^M)^2) + R^M((N-1)\varphi + (\sigma^M)^2)}. \quad (22)$$

**Proposition 4.** *Suppose that the top managerial contract  $S^0(X) = A + BX$  is given. Under optimal middle managerial contracts, all middle managers exert the same levels of effort such that  $\mu^i = \mu^j = \mu^M$ , for  $i, j = 1, \dots, N$ , and each optimal middle managerial contract is given in the following form: For  $i = 1, \dots, N$ ,*

$$\begin{aligned} S^i(Y) &= \mathcal{W}_0^M - (\beta^i)^\top e + c^M(\mu^M) + (\beta^i)^\top Y \\ &\quad + \frac{R^M}{2} \left[ \left( \sum_{j=0}^N \beta^{ij} \right)^2 (\sigma^c)^2 + (\beta^{i0})^2 (\sigma^0)^2 + \sum_{j=1}^N (\beta^{ij})^2 (\sigma^M)^2 \right], \end{aligned} \quad (23)$$

where  $\beta^{ii} = \beta$ ;  $\beta^{i0} = \gamma^0$ ; and  $\beta^{ij} = \gamma^M$ , for  $j = 1, \dots, N$  and  $j \neq i$ .

In Proposition 4, the way the middle managerial performance measure is constructed is of particular interest. Note that if  $R^0 = 0$ , i.e., the top manager is risk neutral, then  $\gamma^0 = \bar{\gamma}^0$  and  $\gamma^M = \bar{\gamma}^M$ , where

$$\begin{aligned} \bar{\gamma}^0 &:= \eta\beta \left( -1 + (N-1) \frac{\varphi}{((N-1)\varphi + (\sigma^M)^2)} \right), \\ \bar{\gamma}^M &:= - \frac{\varphi\beta}{((N-1)\varphi + (\sigma^M)^2)}. \end{aligned}$$

In this case, we have

$$(\beta^i)^\top Y = \beta \left\{ Y^i - \frac{-1}{\beta} \left( \bar{\gamma}^0, \bar{\gamma}^M \mathbf{1}_{N-1}^\top \right) Y_{-i} \right\}.$$

---

managerial effort productivity can be affected by not only his/her own productivity shock but company-wide firm-specific shocks.

Also, note that  $\frac{-1}{\beta} (\bar{\gamma}^0, \bar{\gamma}^M \mathbf{1}_{N-1}^\top)$  is a vector of slopes of multiple regression of  $Y^i$  on  $Y_{-i}$ , which is also equal to  $(p_{-1})^\top (Q_{-i})^{-1}$ , as seen in Section 3. Recall that when the top manager is risk-neutral, he is not concerned with risk-sharing, but he is only interested in middle managerial effort incentives. Thus,  $Y^i - \frac{-1}{\beta} (\bar{\gamma}^0, \bar{\gamma}^M \mathbf{1}_{N-1}^\top) Y_{-i}$  can be interpreted as the optimal performance measure for middle manager  $i$ 's effort incentives, or as a sufficient statistic for manager  $i$ 's (optimal) effort level.

Thus, when  $R^0 > 0$ , in general the performance-related part of each middle managerial contract can be decomposed into two parts as follows.

$$(\beta^i)^\top Y = \beta \left\{ Y^i - \frac{-1}{\beta} (\bar{\gamma}^0, \bar{\gamma}^M \mathbf{1}_{N-1}^\top) Y_{-i} \right\} + (\gamma^0 - \bar{\gamma}^0, (\gamma^M - \bar{\gamma}^M) \mathbf{1}_{N-1}^\top) Y_{-i},$$

where

$$\begin{aligned} \gamma^0 - \bar{\gamma}^0 &= \frac{R^0 B + N R^0 B \eta}{N R^0 B + R^M} - \eta (N-1) (\gamma^M - \bar{\gamma}^M), \\ \gamma^M - \bar{\gamma}^M &= \frac{R^0 B (N\varphi + (\sigma^M)^2) \{ (N-1)\varphi + (\sigma^M)^2 (1-\beta) \}}{\{ (N-1) R^0 B (N\varphi + (\sigma^M)^2) + R^M ((N-1)\varphi + (\sigma^M)^2) \} ((N-1)\varphi + (\sigma^M)^2)}. \end{aligned}$$

Unlike the case of independent outcomes where they are used only for risk-sharing purposes,  $\gamma^0$  and  $\gamma^M$  are used for both incentives and risk-sharing. In particular, parts  $\bar{\gamma}^0$  and  $\bar{\gamma}^M$  estimate the common noise from  $Y_{-i}$ , a vector of other managerial performance outcomes, and the top manager uses the estimates, i.e.,  $\bar{\gamma}^0$  and  $\bar{\gamma}^M$ , to filter out common noise from  $Y^i$  in order to improve middle manager  $i$ 's effort incentives. The remaining parts, i.e.,  $\gamma^0 - \bar{\gamma}^0$  and  $(\gamma^M - \bar{\gamma}^M) \mathbf{1}_N$ , are for risk sharing, and are middle managerial shares of non-performance-related risks of  $Y^0$  and  $Y_{-i}$ , respectively. In fact,  $(\gamma^0 - \bar{\gamma}^0, (\gamma^M - \bar{\gamma}^M) \mathbf{1}_N) Y_{-i}$  is statistically orthogonal to the aforementioned sufficient statistic for middle manager  $i$ 's effort level.

As a special case, if  $\sigma^c = 0$ , then  $\bar{\gamma}^0 = \bar{\gamma}^M = 0$ , i.e., the effort-performance measure reduces to  $Y^i$ ; and  $\gamma^0$  and  $\gamma^M$  are solely used for risk-sharing purposes only. Thus,  $\gamma^0$  and  $\gamma^M$  in (21) and (22) with  $\sigma^c = 0$  become identical to those of the previous subsection for the independent outcome case.

Next, we examine the top managerial contract. First, let us define  $\tau$  as follows.

$$\tau = N\varphi + (\sigma^M)^2.$$

**Proposition 5.** *The optimal top managerial contract under hierarchical contracting is given in the following form.*

$$S^0(X) = \mathcal{W}_0^0 + c^0(\mu^0) - BE[X] + \frac{R^0}{2} B^2 \text{Var}(X) + BX,$$



where

$$B = \frac{1}{N^{-\delta^0} + \kappa(N)N^{1-\delta^M}} c_\mu^0(\mu^0), \quad (24)$$

$$\begin{aligned} E[X] = & (N^{-\delta^0} + \kappa(N)N^{1-\delta^M})\mu^0 + N \left[ N^{-\delta^M} \mu^M - \mathcal{W}_0^M - c^M(\mu^M) \right. \\ & \left. - \frac{R^M}{2} \left\{ (\gamma^0 + (N-1)\gamma^M + \beta)^2 (\sigma^c)^2 \right. \right. \\ & \left. \left. + (\gamma^0)^2 (\sigma^0)^2 + (\beta^2 + (N-1)(\gamma^M)^2) (\sigma^M)^2 \right\} \right], \quad (25) \end{aligned}$$

$$\begin{aligned} \text{Var}(X) = & \left\{ 1 + N - N(\gamma^0 + (N-1)\gamma^M + \beta) \right\}^2 (\sigma^c)^2 \\ & + \left\{ 1 - N\gamma^0 \right\}^2 (\sigma^0)^2 + N \left\{ 1 - (\beta + (N-1)\gamma^M) \right\}^2 (\sigma^M)^2, \quad (26) \end{aligned}$$

and  $\mu^0$  and  $\mu^M$  satisfy

$$\begin{aligned} 0 = & 1 - \beta - R^M \beta N^{2\delta^M} c_{\mu\mu}^M(\mu^M) \frac{\tau}{\tau - \varphi} (\sigma^M)^2 \\ & + R^0 B \left( \frac{\tau}{\tau - \varphi} \right) \left( \frac{R^M \{ (N-1)\varphi + (\sigma^M)^2 (1 - \beta) \} N^{2\delta^M} c_{\mu\mu}^M(\mu^M) (\sigma^M)^2}{(N-1)R^0 B \tau + R^M(\tau - \varphi)} \right). \quad (27) \end{aligned}$$

The first three terms of the LHS of condition (27) are familiar terms that appear when investors directly contract each middle manager.<sup>21</sup> Condition (27) immediately implies that  $0 < \beta < 1$ . Furthermore, by comparing (27) with (6), one can see that each middle manager works harder under hierarchical contracting than he/she does under direct contracting. This implication is parallel to that of Proposition 3.

Using Proposition 5, we now restate the investors' problem. Choose  $\mu^0$  and  $\mu^M$  to maximize

---

<sup>21</sup>For the three terms under direct contracting, see Eq.(6). Multiplying both sides by  $N^{\delta^M}$ , we have the LHS of Eq.(6) equal to the three terms.

the following expected net profit subject to (27):

$$\begin{aligned}
& E[X - S^0(X)] \\
&= N \left[ N^{-\delta^M} \mu^M - \mathcal{W}_0^M - c^M(\mu^M) - \frac{R^M}{2} \beta^2 \frac{\tau}{\tau - \varphi} (\sigma^M)^2 \right. \\
&\quad \left. - \frac{R^M}{2} \left\{ (\gamma^0 + (N-1)\gamma^M + \beta)^2 (\sigma^c)^2 \right. \right. \\
&\quad \left. \left. + (\gamma^0)^2 (\sigma^0)^2 + (N-1)(\gamma^M)^2 (\sigma^M)^2 - \beta^2 \left( \frac{\tau}{\tau - \varphi} - 1 \right) (\sigma^M)^2 \right\} \right] \\
&\quad + (N^{-\delta^0} + \kappa(N)N^{1-\delta^M}) \mu^0 - \mathcal{W}_0^0 - c^0(\mu^0) \\
&\quad - \frac{R^0}{2} B^2 \left[ \left\{ 1 + N - N(\gamma^0 + (N-1)\gamma^M + \beta) \right\}^2 (\sigma^c)^2 \right. \\
&\quad \left. + \left\{ 1 - N\gamma^0 \right\}^2 (\sigma^0)^2 + N \left\{ 1 - (\beta + (N-1)\gamma^M) \right\}^2 (\sigma^M)^2 \right], \tag{28}
\end{aligned}$$

where  $B$ ,  $\beta$ ,  $\gamma^0$  and  $\gamma^M$  are given in (24), (20), (21), and (22), respectively.

The above problem is clearly a lot more complex than that for the case of independent outcomes. However, limiting behaviors of the top managerial sensitivity with correlated outcomes are much simpler than those with independent outcomes.

**Theorem 2.** *Assume quadratic cost functions as in (15).*

*I. Limits of the top managerial sensitivity are as follows.*

- i. If  $\delta^0 > -\frac{1}{2}$ , then the top managerial sensitivity approaches zero.*
- ii. If  $\delta^0 = -\frac{1}{2}$  and if  $K^0 R^M (\sigma^c)^2 > (< \text{ resp.}) 1$ , then the top managerial sensitivity approaches zero (the first best, resp.).*
- iii. If  $\delta^0 < -\frac{1}{2}$ , then the top managerial sensitivity approaches the first best.*

*II. As the firm size grows, the optimal middle managerial sensitivity is always higher than, but approaches the second best sensitivity of the directing contracting case.*

Unlike those of Theorem 1, conditions of Theorem 2 for approaching either of zero or the first best in the limit are independent of  $q$  or  $d^*$ . Recall in Theorem 1 that both  $q$  and  $d^*$  are related to the bound of the marginal top managerial productivity, and that the limit of the top managerial sensitivity depends on a tradeoff between the two main economic factors, the marginal productivity of the top managerial effort and the aggregate middle managerial risk-sharing premium. The marginal productivity is affected by both  $q$  and the returns to scale of the top managerial effort production function.

In Theorem 2, the limit of the top managerial sensitivity still depends on the same tradeoff. In Part I of the theorem, the fact that the top managerial sensitivity is independent of  $q$  or  $d^*$  means that in the tradeoff, as long as  $q$  is bounded, the aggregate middle managerial risk-sharing premium is the dominant factor in determining the top managerial sensitivity in the limit.<sup>22</sup>

The reason is as follows. In the presence of  $\sigma^c$ , the common risk  $\sigma^c \xi^c$  cannot be diversified away through risk-sharing among managers. As a consequence, the aggregate middle-managerial risk-sharing premium for the common risk could explode as the firm size increases. Thus, when  $\delta^0 > -1/2$ , or when the firm does not have a sufficiently increasing returns to scale technology, the top managerial effort contribution can grow at a much slower speed than the aggregate common risk premium can, and thus decreasing the premium can become much more important than improving top managerial effort production. Consequently, the top managerial sensitivity approaches zero as the firm size grows. On the other hand, when  $\delta^0 < -1/2$ , the top managerial effort contribution can be exceptionally high such that it can explode even at a faster speed than the aggregate common risk premium can, and thus the top managerial sensitivity approaches the first best as the firm size grows.

## 5 Comparing Top Managerial Sensitivities of Direct and Hierarchical Contracts

In this section, we assume quadratic cost functions as in (15) and compare top managerial sensitivities under direct and hierarchical contracting arrangements. Recall that  $X$  is the net profit before top managerial compensation, and  $Y$  is a vector of all individual outcomes. In order to avoid confusion, in this section, we differentiate  $(X, Y)$ 's for the two different contracting regimes, by using new notation  $(X_D, Y_D)$  for direct contracting, and just  $(X, Y)$  for hierarchical contracting. Also recall that the top managerial sensitivities are computed based on  $Y_D$  for direct contracting and  $X$  for hierarchical contracting. Thus, direct comparison of the sensitivities for the two different contracting regimes may not be meaningful.

Nevertheless, in order to compare the two different sensitivities on an equal basis and to be consistent with frequent practices in the empirical literature on executive compensation, we convert the optimal vector of top managerial sensitivities to  $Y_D$  under direct contracting into another sensitivity to  $X_D$  which we define as a regression slope coefficient when the optimal top managerial compensation under direct contracting is regressed on  $X_D$ . Let  $B_D$  be the regression slope coefficient. In this section, we call  $B_D$  the top managerial sensitivity under

---

<sup>22</sup>The ratio of the sum of common risks in the total output to  $q$  or  $d^*$  diverges to infinity as  $N$  tends to infinity, and therefore,  $q$  or  $d^*$  does not influence the tradeoff.

direct contracting. Then we can compare the redefined direct-contracting sensitivity  $B_D$  with the hierarchical-contracting sensitivity  $B$ . We also write  $(\beta_D, \gamma_D^0, \gamma_D^M)$  and  $(\beta, \gamma^0, \gamma^M)$  for direct and hierarchical contracting, respectively, to denote middle managerial sensitivities to each middle manager's own outcome, the top managerial outcome, and other middle managerial outcomes.

For ease of exposition, we only focus on cases with nonincreasing returns to scale, which are more consistent with real life. To start the comparison of contract sensitivities for the two different contracting regimes, we first look at limiting behaviors of  $B_D$ .

**Proposition 6.** *If  $\delta^i \geq 0$ ,  $i = 0, M$ , then  $\lim_{N \rightarrow \infty} B_D = 0$ .*

When the firm has a nonincreasing returns to scale technology, the top managerial sensitivity under direct contracting approaches zero as the firm size increases. For intuition, one may imagine a special case where all outcomes are independent. Then  $S_D^0$  depends on  $Y_D^0$  only, where  $S_D^0$  and  $Y_D^0$  are, respectively, the top managerial salary and effort outcome under direct contracting, and intuitively,  $Var(X_D) \rightarrow \infty$  and  $Cov(Y_D^0, X_D) \rightarrow 0$ , as  $N \rightarrow \infty$ . Thus  $B_D = Cov(\beta_D^{00} Y_D^0, X_D) / Var(X_D) \rightarrow 0$ .

In the next subsection, we assume that all managerial effort outcomes are independent of each other. Then in the subsection following the next, we assume there is common uncertainty across all outcomes.

## 5.1 Independent Outcomes

With  $B_D$  properly defined as above, now we are ready to compare  $B$  and  $B_D$ .

**Proposition 7.** *Assume quadratic cost functions as in (15). Let  $B^*$  be the optimal top managerial sensitivity under hierarchical contracting.*

- I. *If  $\delta^i > 0$ , for  $i = 0, M$ , and  $K^0 R^M (\sigma^M)^2 - q^2 > 0$  ( $< 0$ , resp.), then  $B^* < (>, resp.) B_D$  for large  $N$ .*
- II. *If  $\delta^i = 0$  for  $i = 0, M$ , and  $R^M (\sigma^M)^2 - d^* > 0$  ( $< 0$ , resp.), then  $B^* < (>, resp.) B_D$  for large  $N$ .*

Proposition 7 tells us that if  $R^M (\sigma^M)^2$  is sufficiently high (low), then the top managerial sensitivity for a hierarchical contract is lower (higher) than that of its direct-contracting counterpart. We illustrate differences of the two contracting regimes in sensitivities using four tables and four figures. These tables and figures highlight differences in top managerial sensitivities as mainly described in Theorem 1 and Proposition 7, and the tables also provide numerical values of middle managerial sensitivities mainly based on Corollary 1 and Proposition 2.

$N$	$B_D$	$\beta_D$	$B$	$\beta$	$\gamma^0$	$\gamma^M$
1	0.189189	0.285714	0.179600	0.366928	0.152255	N.A.
5	0.088548	0.054805	0.023200	0.073398	0.020789	0.019672
10	0.046754	0.024617	0.005300	0.029406	0.005033	0.004910
15	0.031247	0.015278	0.002300	0.017434	0.002223	0.002189
20	0.023309	0.010866	0.001200	0.012013	0.001172	0.001159
25	0.018521	0.008335	0.000800	0.009106	0.000784	0.000778
30	0.015332	0.006708	0.000500	0.007194	0.000493	0.000489

N.A.: not applicable.

Table 1: Sensitivities with  $R^0 = R^M = 1$ ,  $K^0 = K^M = 0.001$ ,  $\sigma^0 = \sigma^M = 50$ ,  $\delta^0 = \delta^M = 0.6$ , and  $\kappa(N)N^{1-\delta^M} = 1 - \frac{1}{N}$ .

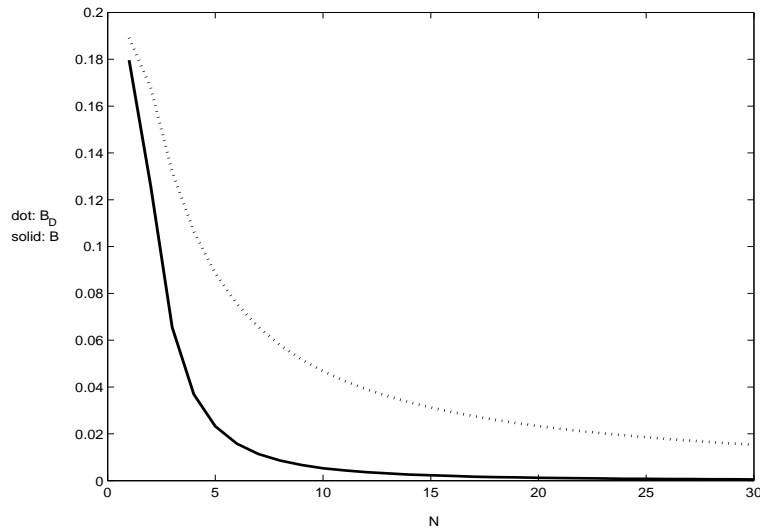


Figure 1: Top managerial sensitivity with  $R^0 = R^M = 1$ ,  $K^0 = K^M = 0.001$ ,  $\sigma^0 = \sigma^M = 50$ ,  $\delta^0 = \delta^M = 0.6$ , and  $\kappa(N)N^{1-\delta^M} = 1 - \frac{1}{N}$ .

Based on Theorem 1-I-i and Proposition 7-I with  $K^0 R^M (\sigma^M)^2 - q^2 > 0$ , Table 1 throws light on how top managerial sensitivities change as the number of middle managers increases. For this table, we assume that  $R^0 = R^M = 1$ ,  $K^0 = K^M = 0.001$ ,  $\sigma^0 = \sigma^M = 50$ ,  $\delta^0 = \delta^M = 0.6$ , and  $\kappa(N)N^{1-\delta^M} = 1 - \frac{1}{N}$ . The assumption  $\delta^0 = \delta^M = 0.6$  is to be consistent with an empirical finding by Baker and Hall [2004]. Note that as the number of middle managers increases from 1 to 30, top managerial sensitivities for direct contracts decrease from 0.189198 to 0.015332 as suggested by Proposition 6, whereas those for hierarchical contracting decrease from 0.1796 to 0.0005, which is consistent with Theorem 1-I-i. These two sets of numbers are consistent with Proposition 7-I. Note that the speed of decrease in the top managerial sensitivity is far faster under hierarchical contracting than it is under direct contracting. As repeatedly discussed

in previous sections, this dramatic decrease under hierarchical contracting occurs as investors try to discourage the top managerial risk-sharing behavior. In particular, the sensitivity for hierarchical contracting with  $N = 15$  is 0.0023, which is lower than Jensen and Murphy's [1990] \$3.25 for every \$1,000 change in the shareholders' wealth of large firms.<sup>23</sup>

Note also from the same table that sensitivities of middle managerial hierarchical contracts are always higher than those of their direct contracts, which is also a consequence of the top managerial risk-sharing behavior, as predicted by Theorem 1-II. However, the difference between hierarchical and direct contracts in middle managerial sensitivities narrows as the firm size grows, demonstrating the diversification effect in hierarchical risk sharing, as we noted in previous sections.

$N$	$B_D$	$\beta_D$	$B$	$\beta$	$\gamma^0$	$\gamma^M$
1	0.504505	0.285714	0.709100	0.525864	0.414897	N.A.
5	0.252213	0.054805	0.906300	0.202508	0.163843	0.156267
10	0.147252	0.024617	0.948400	0.110872	0.090462	0.088432
15	0.103258	0.015278	0.964200	0.075800	0.062355	0.061461
20	0.079282	0.010866	0.972500	0.057417	0.047555	0.047063
25	0.064246	0.008335	0.977700	0.046137	0.038428	0.038120
30	0.053954	0.006708	0.981200	0.038522	0.032238	0.032029

Table 2: Sensitivities with  $R^0 = R^M = 1$ ,  $K^0 = K^M = 0.001$ ,  $\sigma^0 = \sigma^M = 50$ ,  $\delta^0 = \delta^M = 0.6$ , and  $\kappa(N)N^{1-\delta^M} = 2 - \frac{1}{N}$ .

Table 2 is still related to Theorem 1-I-i and Proposition 7-I but with  $K^0 R^M (\sigma^M)^2 - q^2 < 0$ , and demonstrates a case where the sensitivity of the top managerial hierarchical contract approaches the first best, whereas that of the direct contract approaches zero at a relatively slow speed. Assumptions for Table 2 are the same as those for Table 1, except that in Table 2, the aggregate top managerial influence on middle managerial outcomes  $\kappa(N)N^{1-\delta^M}$  is increased to  $2 - \frac{1}{N}$ , i.e.,  $\kappa(N)N^{1-\delta^M} = 2 - \frac{1}{N}$  satisfying  $K^0 R^M (\sigma^M)^2 - q^2 < 0$ .

Comparing  $\gamma^0$  and  $\gamma^M$  in Tables 1 and 2, one can see that as the sensitivity increases, the top manager in Table 2 has stronger incentives to shift his compensation risk to middle managers than he does in Table 1. For example, when  $N = 1$ , the top manager in Table 2 shifts about 41% of his own performance risk to the middle manager, whereas he does only about 15% in Table 1. When  $N = 30$ , he has even stronger incentives for risk-sharing and shifts about 97% ( $\approx 0.032238 \times 30$ ) to 30 middle managers, whereas he does only 1.5% in Table 1.

<sup>23</sup>Later articles such as Hall and Leibman [1998] and Baker and Hall [2004] argue that Jensen-Murphy statistic tends to underestimate true sensitivities. In this paper, our numbers are just intended to illustrate the effect of hierarchical contracting on managerial contract sensitivities of firms of different sizes.

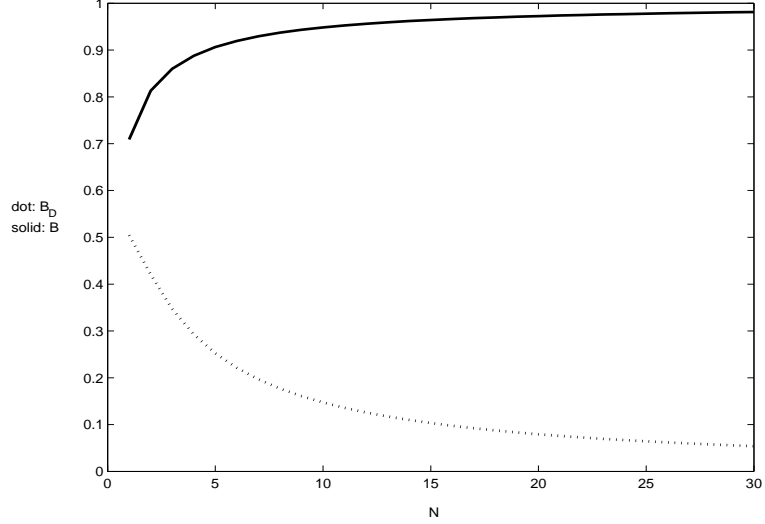


Figure 2: Top managerial sensitivity with  $R^0 = R^M = 1$ ,  $K^0 = K^M = 0.001$ ,  $\sigma^0 = \sigma^M = 50$ ,  $\delta^0 = \delta^M = 0.6$ , and  $\kappa(N)N^{1-\delta^M} = 2 - \frac{1}{N}$ .

$N$	$B_D$	$\beta_D$	$B$	$\beta$	$\gamma^0$	$\gamma^M$
1	0.077309	0.135135	0.050200	0.171122	0.047800	N.A.
2	0.114398	0.135135	0.063900	0.177843	0.056659	0.049380
5	0.085348	0.135135	0.025000	0.151807	0.022222	0.019277
10	0.053263	0.135135	0.007500	0.140359	0.006977	0.006040
15	0.038337	0.135135	0.003500	0.137624	0.003325	0.002877
20	0.029895	0.135135	0.002000	0.136574	0.001923	0.001664
25	0.024486	0.135135	0.001300	0.136077	0.001259	0.001089
30	0.020730	0.135135	0.000900	0.135791	0.000876	0.000758

Table 3: Sensitivities with  $R^0 = R^M = 1$ ,  $K^0 = K^M = 0.001$ ,  $\sigma^0 = \sigma^M = 80$ ,  $\delta^0 = \delta^M = 0$ , and  $\kappa(N)N^{1-\delta^M} = 1 - \frac{1}{N}$ .

Table 3 presents a numerical case of constant returns to scale when the outcome volatility is sufficiently high and the aggregate top managerial influence on middle managerial outcomes is sufficiently low. In particular, Table 3 is related to Theorem 1-I-ii and Proposition 7-II with  $K^0 R^M (\sigma^M)^2 - q^2 < 0$ . Since constant returns to scale imply that the middle managerial productivity is unaffected by the number of middle managers, or the firm size,  $\beta_D$  (for direct contracting) stays the same over  $N$ . Under hierarchical contracting, however, because of the top managerial motivation for risk-sharing, the middle managerial sensitivity is always larger than that of the direct contract, and decreases as  $N$  increases. Again, the decrease is a result of the previously-mentioned diversification effect on risk sharing.

The top managerial sensitivities in Table 3 change in the same fashion in the limit as do

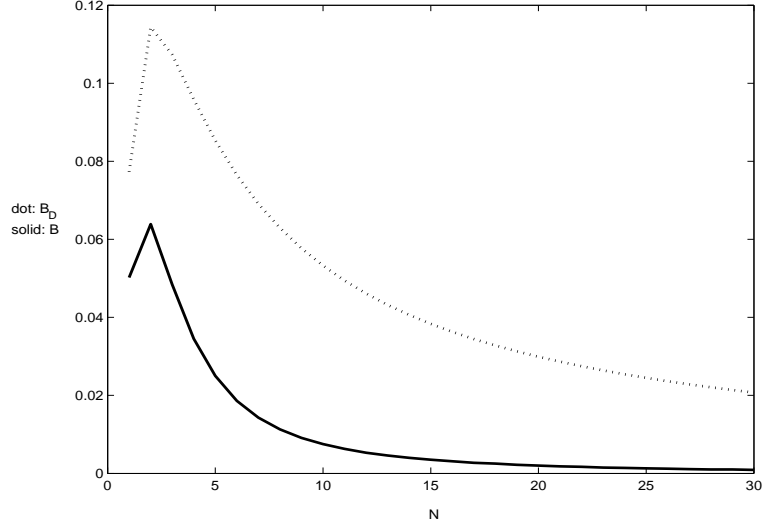


Figure 3: Top managerial sensitivity with  $R^0 = R^M = 1$ ,  $K^0 = K^M = 0.001$ ,  $\sigma^0 = \sigma^M = 80$ ,  $\delta^0 = \delta^M = 0$ , and  $\kappa(N)N^{1-\delta^M} = 1 - \frac{1}{N}$ .

those in Table 1. However, since managers in Table 3 are more productive than those in Table 1, sensitivities in Table 3 are higher in general than those in Table 1. One exceptional behavior of  $B$  is noted at  $N = 2$ : that is,  $B$  temporarily increases when  $N$  increases from 1 to 2, and then decreases monotonically. This spike in Figure 3 occurs because the top managerial productivity temporarily dominates the aggregate middle managerial risk-sharing premium. The temporary domination is partly due to a temporary surge in the top managerial productivity in  $\kappa(N)N^{1-\delta^M}$  from 0 to  $1/2$  as  $N$  increases from 1 to 2. After  $N$  greater than 2, the productivity grows at a much slower rate than the risk-sharing premium does.

$N$	$B_D$	$\beta_D$	$B$	$\beta$	$\gamma^0$	$\gamma^M$
1	0.189189	0.285714	0.179600	0.366928	0.152255	N.A.
5	0.219557	0.285714	0.905500	0.373399	0.163817	0.122758
10	0.142848	0.285714	0.960900	0.333153	0.090574	0.066414
15	0.104658	0.285714	0.975600	0.318130	0.062402	0.045382
20	0.082407	0.285714	0.982300	0.310323	0.047578	0.034453
25	0.067914	0.285714	0.986100	0.305545	0.038441	0.027763
30	0.057740	0.285714	0.988600	0.302319	0.032246	0.023247

Table 4: Sensitivities with  $R^0 = R^M = 1$ ,  $K^0 = K^M = 0.001$ ,  $\sigma^0 = \sigma^M = 50$ ,  $\delta^0 = \delta^M = 0$ , and  $\kappa(N)N^{1-\delta^M} = 1 - \frac{1}{N}$ .

Table 4 is also for a case of constant returns to scale when the outcome volatility is sufficiently low such that the overall top managerial productivity dominates the aggregate middle



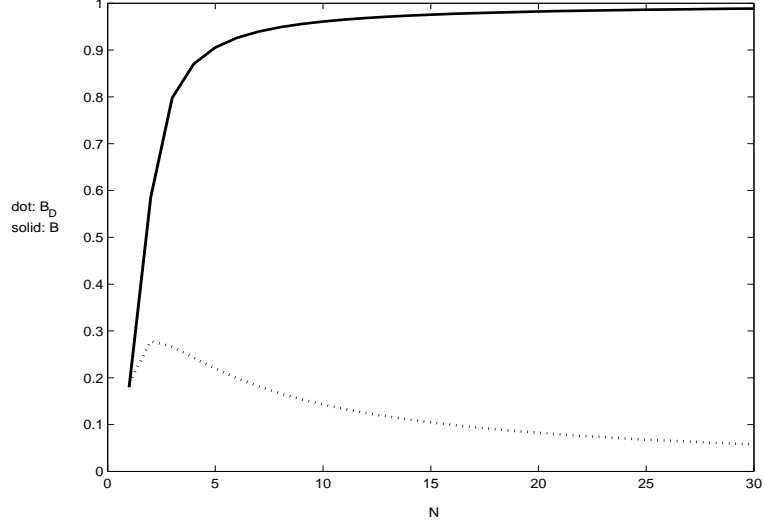


Figure 4: Top managerial sensitivity with  $R^0 = R^M = 1$ ,  $K^0 = K^M = 0.001$ ,  $\sigma^0 = \sigma^M = 50$ ,  $\delta^0 = \delta^M = 0$ , and  $\kappa(N)N^{1-\delta^M} = 1 - \frac{1}{N}$ .

managerial risk-sharing premium. In particular, Table 2 is related to Theorem 1-I-ii and Proposition 7-II but with  $K^0 R^M (\sigma^M)^2 - q^2 > 0$ .

Note that  $B$  quickly jumps from 0.1796 to 0.9055 as  $N$  changes from 1 to 5. The size of the jump is noticeable when compared with the effect on  $B$  in Table 2 for the same change in  $N$ . The reason is that when the production technology generates constant returns to scale across managers, increasing the top managerial sensitivities becomes less costly in terms of risk-sharing premia than it does when the technology is decreasing returns to scale. Furthermore, in the case of Table 4, the top manager also aggressively shifts his compensation risks to middle managers, as can be inferred from much higher values of  $\beta$ ,  $\gamma^0$  and  $\gamma^M$  than their counterparts in Table 3.

## 5.2 Correlated Outcomes with Common Uncertainty

Let us now assume  $\sigma^c > 0$  and the firm has a nonincreasing returns to scale technology. The top managerial sensitivity under hierarchical contracting  $B$  can be compared with that under direct contracting  $B_D$  as follows.

**Proposition 8.** *Assume quadratic cost functions as in (15). Suppose that  $\delta^i \geq 0$ ,  $i = 0, M$ . If  $[q > 0]$  or  $[q = 0 \text{ and } 1 > \delta^M \geq \delta^0 > 0]$ , then  $B^* < B_D$  for large  $N$ .*

Proposition 8 tells us that, when production functions exhibit nonincreasing returns to scale, regardless of values in  $q$  and  $d$ , the top managerial sensitivity under hierarchical con-

tracting is less than that under direct contracting as the firm size increases sufficiently large. Next, we try to illustrate main points of Proposition 8 and Theorem 2, using Tables 5 and 6.

$N$	$B_D$	$\beta_D$	$\gamma_D^0$	$\gamma_D^M$	$B$	$\beta$	$\gamma^0$	$\gamma^M$
1	0.1351	0.3149	-0.1134	N.A.	0.1171	0.3657	0.0109	N.A.
5	0.0323	0.0732	-0.0108	-0.0108	0.0062	0.0786	-0.0044	-0.0049
10	0.0103	0.0351	-0.0030	-0.0030	0.0010	0.0360	-0.0020	-0.0020
15	0.0049	0.0224	-0.0013	-0.0013	0.0003	0.0227	-0.0010	-0.0010
20	0.0029	0.0161	-0.0007	-0.0007	0.0001	0.0162	-0.0006	-0.0006
25	0.0019	0.0125	-0.0005	-0.0005	0.0001	0.0126	-0.0004	-0.0004
30	0.0013	0.0101	-0.0003	-0.0003	0.0001	0.0102	-0.0002	-0.0002

Table 5: Sensitivities with  $R^0 = R^M = 1$ ,  $K^0 = K^M = 0.001$ ,  $\sigma^c = 30$ ,  $\sigma^0 = \sigma^M = 40$ ,  $\delta^0 = \delta^M = 0.6$ , and  $\kappa(N)N^{1-\delta^M} = 1 - \frac{1}{N}$ .

$N$	$B_D$	$\beta_D$	$\gamma_D^0$	$\gamma_D^M$	$B$	$\beta$	$\gamma^0$	$\gamma^M$
1	0.1351	0.3149	-0.1134	N.A.	0.1171	0.3657	0.0109	N.A.
5	0.0673	0.3526	-0.0520	-0.0520	0.0198	0.3626	-0.0286	-0.0358
10	0.0236	0.3655	-0.0310	-0.0310	0.0032	0.3673	-0.0269	-0.0281
15	0.0117	0.3710	-0.0221	-0.0221	0.0010	0.3716	-0.0208	-0.0212
20	0.0070	0.3740	-0.0172	-0.0172	0.0005	0.3743	-0.0165	-0.0167
25	0.0046	0.3760	-0.0140	-0.0140	0.0002	0.3761	-0.0138	-0.0138
30	0.0033	0.3773	-0.0119	-0.0119	0.0001	0.3774	-0.0117	-0.0118

Table 6: Sensitivities with  $R^0 = R^M = 1$ ,  $K^0 = K^M = 0.001$ ,  $\sigma^c = 30$ ,  $\sigma^0 = \sigma^M = 40$ ,  $\delta^0 = \delta^M = 0$ , and  $\kappa(N)N^{1-\delta^M} = 1 - \frac{1}{N}$ .

Table 5 shows a case of decreasing returns to scale in the presence of common uncertainty, and Table 6 a case of constant returns to scale. In both cases, the top managerial sensitivities under hierarchical contracts approach zero at faster speeds than do those under direct contracts, which is consistent with Theorem 2 and Proposition 7. Moreover, the top managerial sensitivities under both direct and hierarchical contracts in Table 6 decrease at slower speeds than do those in Table 5, because the top managerial effort productivity is higher under a constant-return-to-scale technology than it is under a decreasing-return-to-scale technology. Theorem 2 suggests that the same patterns of limiting behavior of sensitivities persist until  $\delta^0$  hits  $-1/2$ .

Moreover, in Table 5,  $\gamma_D^0$  and  $\gamma_D^M$  are negative, as regression slopes on  $Y_{-i}$  are used to improve middle managerial incentives. In the same table,  $\gamma^0$  and  $\gamma^M$  are still mostly negative, although they are less negative than  $\gamma_D^0$  and  $\gamma_D^M$ . Under hierarchical contracting, when  $\gamma^0$  and  $\gamma^M$  are negative, it means that adjustment of each middle managerial performance for

common noise dominates the top managerial risk-sharing concern, and as a result, each middle managerial performance will be explicitly discounted based on other managerial performances. Unless the top managerial risk aversion is sufficiently high enough,  $\gamma^0$  and  $\gamma^M$  can stay negative.

Both Tables 5 and 6 show similar patterns in various sensitivities over  $N$ , except that in Table 6, both  $\beta_D$  and  $\beta$  are in general increasing. The reason is that as  $N$  increases, the common noise is more and more filtered out and thus the volatility of the performance measure for each middle managerial performance decreases, and thus  $\beta_D$  increase. Even when  $\beta_D$  increases,  $\beta$  may not increase when  $N$  is small, because with a small  $N$ , additional risk imposed on each middle manager (for top manager's risk-sharing) may not be sufficiently diversified away and thus the volatility of the performance measure may temporarily increase, as can be noted from  $\beta$  decreasing from 0.3657 to 0.3626 in Table 6, while  $\beta_D$  is increasing from 0.3149 to 0.3526, as  $N$  changes from 1 to 5.

## 6 Conclusion

We have examined the structure of optimal contracts of both the top and middle managers under a hierarchical contracting environment where investors contract the top manager and the top manager is given discretion to subcontract middle managers. As well-recognized in the literature, such discretion to the top manager may be given when he has a technology to monitor outcomes of other managers.

We have compared the optimal structure of hierarchical contracts with that of direct contracts under which investors directly contract each middle manager. We have argued that under hierarchical contracting where investors are risk neutral and all managers are risk averse, each optimal contract between the top and middle managers is based on three performance measures: individual middle managerial effort outcome, top managerial effort outcome, and the aggregate outcome of the firm. Note that the first measure is a sufficient statistic for the individual middle manager, if all outcomes are independent of each other. The last two performance measures are included as the top manager tries to share his own compensation risks with middle managers, even when these two measures have almost nothing to do with individual managerial effort performance. This observation may provide a rationale to justify common managerial compensation practices such as company-wide bonus programs including stock option plans for nonexecutive employees.

We also have compared the sensitivities of both top and middle managerial contracts under the direct and hierarchical contracting environments. We have argued that under hierarchical contracting, the top managerial sensitivity approaches either as high as the first best or as low as zero, as the firm size grows. In particular, under some reasonable conditions such as

nonincreasing returns to scale in labor and correlated managerial effort outcomes, sensitivities of top managerial hierarchical contracts of sufficiently large firms can be far lower than that of the direct contract can.

We have also shown that under hierarchical contracting, middle managerial contract sensitivities are higher than, but approach, those of direct middle managerial contracts, as the firm size grows. The reason is that under hierarchical contracting, middle managerial contracts are determined based on both incentives and risk-sharing as the risk averse top manager tries to shift his own compensation risk to middle managers, whereas under direct contracting, middle managerial contracts are determined solely based on incentives.

Finally, a remark on promising future research avenues is in order. In this paper, we have focused on optimal contracts under the assumption that the firm is hierarchically organized, and the top manager can costlessly monitor middle managerial outcomes. It should be clear even without relying on mathematics that when it is too costly for the principal to monitor individual managerial outcomes, hiring a monitor as top manager can improve the principal's wealth, and thus the hierarchical organizational form of the firm can naturally arise. It is also clear that in our model setting, there can exist the optimal size of the firm when the production function exhibits sufficiently decreasing returns to scale in the number of middle managers. Nevertheless, issues of optimal hierarchical organizational form, and optimal monitoring policy in the presence of monitoring costs can still arise. We believe this paper can serve as a benchmark for these important issues, and we leave them for future research.

# Appendix

## A Proof of Proposition 1

Note that manager  $i$ 's expected utility is

$$E[-\exp\{-R^i(S^i - c^i(\mu^i))\}] = -\exp\{-R^i(\alpha^i + (\beta^i)^\top e - c^i(\mu^i) - \frac{R^i}{2}(\beta^i)^\top DD^\top \beta^i)\}, \quad i = 0, 1, \dots, N.$$

Thus, the first order conditions (FOCs) of all agents are

$$\beta^{00} = N^{\delta^0} \left( c_\mu^0(\mu^0) - (\beta_{-0})^\top \theta \right),$$

$$\beta^{ii} = N^{\delta^i} c_\mu^i(\mu^i), \quad i = 1, \dots, N$$

and their participation constraints are

$$\alpha^i + (\beta^i)^\top e - c^i(\mu^i) - \frac{R^i}{2} (\beta^i)^\top DD^\top \beta^i = \mathcal{W}_0^i, \quad i = 0, 1, \dots, N.$$

Using the FOCs, we have

$$\begin{aligned} (\beta^0)^\top DD^\top \beta^0 &= (\beta_{-0})^\top Q_{-0} \beta_{-0} + 2\beta^{00} (p_{-0})^\top \beta_{-0} + [(\sigma^c)^2 + (\sigma^0)^2] (\beta^{00})^2 \\ &= (\beta_{-0})^\top F^\top \beta_{-0} + 2N^{\delta^0} c_\mu^0(\mu^0) G^\top \beta_{-0} + [(\sigma^c)^2 + (\sigma^0)^2] N^{2\delta^0} (c_\mu^0(\mu^0))^2, \end{aligned}$$

and for  $i = 1, \dots, N$ ,

$$(\beta^i)^\top DD^\top \beta^i = (\beta_{-i})^\top Q_{-i} \beta_{-i} + 2N^{\delta^i} c_\mu^i(\mu^i) (p_{-i})^\top \beta_{-i} + [(\sigma^c)^2 + (\sigma^i)^2] (N^{\delta^i} c_\mu^i(\mu^i))^2.$$

Hence, the investors' problem is to choose  $\beta_{-i}$  and  $\mu^i$ ,  $i = 0, 1, \dots, N$ , to maximize

$$\begin{aligned} &E \left[ \mathbf{1}_{N+1}^\top Y - \sum_{i=0}^N S^i \right] \\ &= [N^{-\delta^0} + \sum_{i=1}^N \kappa^i(N) N^{-\delta^i}] \mu^0 - c^0(\mu^0) - \frac{R^0}{2} [(\beta_{-0})^\top F^\top \beta_{-0} \\ &\quad + 2N^{\delta^0} c_\mu^0(\mu^0) G^\top \beta_{-0} + [(\sigma^c)^2 + (\sigma^0)^2] N^{2\delta^0} (c_\mu^0(\mu^0))^2] \\ &\quad + \sum_{i=1}^N \left[ N^{-\delta^i} \mu^i - c^i(\mu^i) \right. \\ &\quad \left. - \frac{R^i}{2} [(\beta_{-i})^\top Q_{-i} \beta_{-i} + 2N^{\delta^i} c_\mu^i(\mu^i) (p_{-i})^\top \beta_{-i} + [(\sigma^c)^2 + (\sigma^i)^2] (N^{\delta^i} c_\mu^i(\mu^i))^2] \right]. \end{aligned}$$

Note that the above problem is concave in  $\beta_{-i}$  since  $F$  is positive definite. To see this, Let  $x = (x_1, \dots, x_N)^\top$  be an arbitrary  $N$ -dimensional vector. Then,

$$\begin{aligned} x^\top Fx &= (\sigma^c)^2 \left( \sum_{i=1}^N (1 - \kappa^i(N) N^{\delta^0 - \delta^i}) x_i \right)^2 + (\sigma^0)^2 \left( \sum_{i=1}^N \kappa^i(N) N^{\delta^0 - \delta^i} x_i \right)^2 \\ &\quad + \sum_{i=1}^N (\sigma^i)^2 x_i^2 + (\sigma^c)^2 \sum_{i \neq j=1, \dots, N} (\kappa^j(N) N^{\delta^0 - \delta^j} - \kappa^i(N) N^{\delta^0 - \delta^i}) x_i x_j. \end{aligned}$$

Since the last term of the right hand side (RHS) of the above equality is zero,  $x^\top Fx \geq 0$  and the equality holds if and only if  $x$  is a zero vector. Thus,  $F$  is positive definite.

Also note that the FOCs with respect to  $\beta_{-i}$ 's are

$$\beta_{-0} = -N^{\delta^0} c_\mu^0(\mu^0) F^{-1} G,$$

and

$$\beta_{-i} = -N^{\delta^i} c_\mu^i(\mu^i) (Q_{-i})^{-1} p_{-i}, \quad i = 1, \dots, N.$$

Substituting these back into the investors' problem, it becomes to choose  $\mu^i$ ,  $i = 0, 1, \dots, N$  to maximize

$$\begin{aligned} &E \left[ \mathbf{1}_{N+1}^\top Y - \sum_{i=0}^N S^i \right] \\ &= [N^{-\delta^0} + \sum_{i=1}^N \kappa^i(N) N^{-\delta^i}] \mu^0 - c^0(\mu^0) - \frac{R^0}{2} (N^{\delta^0} c_\mu^0(\mu^0))^2 [(\sigma^c)^2 + (\sigma^0)^2 - G^\top F^{-1} G] \\ &\quad + \sum_{i=1}^N \left[ N^{-\delta^i} \mu^i - c^i(\mu^i) \right. \\ &\quad \left. - \frac{R^i}{2} (N^{\delta^i} c_\mu^i(\mu^i))^2 [(\sigma^c)^2 + (\sigma^i)^2 - (p_{-i})^\top (Q_{-i})^{-1} p_{-i}] \right]. \end{aligned}$$

The FOCs for this maximization problem are Eq's (1) and (2). Then the assertion of the proposition immediately follows from these FOCs and the FOCs with respect to  $\beta_{-i}$ ,  $i = 0, 1, \dots, N$ , together with the middle managerial participation constraints.  $\square$

## B Proof of Corollary 1

Note that under the assumption of identical middle managers, the  $N \times N$  matrix  $F$  becomes

$$F = \begin{bmatrix} s & t & \dots & \dots & t \\ t & s & t & \dots & t \\ \dots & \dots & \dots & \dots & \dots \\ t & \dots & \dots & \dots & s \end{bmatrix},$$

where

$$\begin{aligned} s &= (\sigma^c)^2 + (\sigma^M)^2 - 2(\sigma^c)^2 \kappa(N) N^{\delta^0 - \delta^M} + [(\sigma^c)^2 + (\sigma^0)^2] (\kappa(N) N^{\delta^0 - \delta^M})^2 \\ t &= (\sigma^c)^2 - 2(\sigma^c)^2 \kappa(N) N^{\delta^0 - \delta^M} + [(\sigma^c)^2 + (\sigma^0)^2] (\kappa(N) N^{\delta^0 - \delta^M})^2. \end{aligned}$$

Then, it can be shown (by mathematical induction) that

$$F^{-1} = \begin{bmatrix} x & y & \dots & \dots & y \\ y & x & y & \dots & y \\ \dots & \dots & \dots & \dots & \dots \\ y & \dots & \dots & \dots & x \end{bmatrix},$$

where

$$\begin{aligned} x &= \frac{s + (N - 2)t}{(s - t)(s + (N - 1)t)}, \\ y &= \frac{-t}{(s - t)(s + (N - 1)t)}. \end{aligned}$$

Moreover, the  $N$  dimensional vector  $G$  becomes

$$G = \left( (\sigma^c)^2 - [(\sigma^c)^2 + (\sigma^0)^2] \kappa(N) N^{\delta^0 - \delta^M} \right) \mathbf{1}_N.$$

Hence we have

$$\begin{aligned} F^{-1}G &= \left( (\sigma^c)^2 - [(\sigma^c)^2 + (\sigma^0)^2] \kappa(N) N^{\delta^0 - \delta^M} \right) (x + (N - 1)y) \mathbf{1}_N \\ &= \frac{(\sigma^c)^2 - [(\sigma^c)^2 + (\sigma^0)^2] \kappa(N) N^{\delta^0 - \delta^M}}{(\sigma^M)^2 + N \left[ (\sigma^c)^2 (1 - \kappa(N) N^{\delta^0 - \delta^M})^2 + (\sigma^0)^2 (\kappa(N) N^{\delta^0 - \delta^M})^2 \right]} \mathbf{1}_N, \end{aligned}$$

and

$$G^\top F^{-1}G = \frac{N \left( (\sigma^c)^2 - [(\sigma^c)^2 + (\sigma^0)^2] \kappa(N) N^{\delta^0 - \delta^M} \right)^2}{(\sigma^M)^2 + N \left[ (\sigma^c)^2 (1 - \kappa(N) N^{\delta^0 - \delta^M})^2 + (\sigma^0)^2 (\kappa(N) N^{\delta^0 - \delta^M})^2 \right]}.$$

By using the same method as above, one can show, for  $i = 1, \dots, N$ ,

$$(Q_{-i})^{-1} p_{-i} = \frac{(\sigma^c)^2}{(\sigma^M)^2 [(\sigma^c)^2 + (\sigma^0)^2] + (N - 1)(\sigma^c)^2 (\sigma^0)^2} \left[ (\sigma^M)^2 \quad (\sigma^0)^2 \quad \dots \quad (\sigma^0)^2 \right]^\top.$$

and

$$(\sigma^c)^2 + (\sigma^M)^2 - (p_{-i})^\top (Q_{-i})^{-1} p_{-i} = \frac{(\sigma^M)^2 \left( (\sigma^M)^2 [(\sigma^c)^2 + (\sigma^0)^2] + N(\sigma^c)^2 (\sigma^0)^2 \right)}{(\sigma^M)^2 [(\sigma^c)^2 + (\sigma^0)^2] + (N - 1)(\sigma^c)^2 (\sigma^0)^2}.$$

Substituting the above equations for  $F^{-1}G$ ,  $G^\top F^{-1}G$ ,  $(Q_{-i})^{-1} p_{-i}$ , and  $(p_{-i})^\top (Q_{-i})^{-1} p_{-i}$  into relevant equations in Proposition 1, we obtain the Corollary.  $\square$

## C Proof of Corollary 2

Let us rewrite Eq.(13) as follows.

$$1 - \beta - a\beta + \left( \frac{aB(1 - \beta)}{(N - 1)B + r} \right) = 0,$$

where

$$a := R^M N^{2\delta^M} K^M (\sigma^M)^2, \quad b := a + 1, \quad \text{and} \quad r := \frac{R^M}{R^0}.$$

Then, we have

$$1 - \beta = \frac{a((N - 1)B + r)}{(bN - 1)B + rb}, \quad \text{or} \quad \beta = \frac{(N - 1 + a)B + r}{(bN - 1)B + rb}.$$

- By differentiating the above equation for  $\beta$ ,

$$\frac{\partial \beta}{\partial B} = \frac{ra^2}{[(bN - 1)B + rb]^2} > 0.$$

- $\frac{\partial \gamma^M}{\partial B} > 0$  follows from

$$\gamma^M = \frac{B(1 - \beta)}{(N - 1)B + r} = \frac{aB}{(bN - 1)B + rb}.$$

- $\frac{\partial \gamma^0}{\partial B} > 0$  trivially holds.
- $Var(X)$  can be rewritten as

$$Var(X) = \left( \frac{r}{NB + r} \right)^2 (\sigma^0)^2 + N \left( \frac{ar}{(bN - 1)B + rb} \right)^2 (\sigma^M)^2,$$

which implies  $\frac{\partial Var(X)}{\partial B} < 0$ .

## D Proof of Lemma 1

Let  $\Phi'(B; N)$  be the derivative of  $\Phi(B; N)$  with respect to  $B$ . Calculation shows

$$\begin{aligned} \Phi'(B; N) &= \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} (1 - B) \\ &\quad - \frac{R^M r(N + r)B}{(NB + r)^3} (\sigma^0)^2 - \frac{ra^3 N \{bN - 1 + rb\} B}{N^{2\delta^M} K^M \{(bN - 1)B + rb\}^3}. \end{aligned} \quad (\text{A.1})$$

If  $B \geq 1$ , then  $\Phi'(B; N) < 0$  which implies  $\Phi(B; N) < \Phi(1; N)$  for  $B > 1$ . On the other hand, condition (10) implies an optimal  $B$  must be nonnegative. Therefore, investors maximize



$\Phi(B; N)$  only over the interval  $[0, 1]$ . Since the interval  $[0, 1]$  is compact and  $\Phi(B; N)$  is continuous on it, a maximum of  $\Phi(B; N)$  exists in  $[0, 1]$ . Since  $\Phi'(0; N) = \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} > 0$ ,  $\Phi'(1; N) = -\frac{R^M r}{(N+r)^2}(\sigma^0)^2 - \frac{ra^3 N}{N^{2\delta^M} K^M \{bN-1+rb\}^2} < 0$ , and  $\Phi(B; N)$  is continuously differentiable over  $[0, 1]$ , a maximum can occur in  $(0, 1)$ .  $\square$

## E Proof of Theorem 1

Let us start with the following lemma which helps identify the approximate location of the global maximum when the firm size is sufficiently large.

**Lemma A.1.** *Suppose that the contract sensitivity  $B$  is set to either zero or one, and that the top manager optimally contracts middle managerial contracts. Then investors' expected wealth levels evaluated at  $B$  equal to zero and one are compared with each other in the limit as follows:*

- i. (Decreasing returns to scale.) Suppose that  $\delta^i > 0$ ,  $i = 0, M$ . Then,  $\lim_{N \rightarrow \infty} [\Phi(0; N) - \Phi(1; N)] > 0$  ( $=, < 0$ , resp.) iff  $K^0 R^M (\sigma^M)^2 - q^2 > 0$  ( $=, < 0$ , resp.).*
- ii. (Constant returns to scale.) Suppose that  $\delta^i = 0$ ,  $i = 0, M$ . Then,  $\lim_{N \rightarrow \infty} [\Phi(0; N) - \Phi(1; N)] > 0$  ( $=, < 0$ , resp.) iff  $R^M (\sigma^M)^2 - d^* > 0$  ( $=, < 0$ , resp.).*
- iii. (Increasing returns to scale.) If  $\delta^i < 0$ ,  $i = 0, M$ , then  $\lim_{N \rightarrow \infty} [\Phi(0; N) - \Phi(1; N)] = -\infty$ .*

**Proof:** Note that investors' maximand (18) can be rewritten as follows:

$$\begin{aligned} \Phi(B; N) &= \frac{N}{2bN^{2\delta^M} K^M} + \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} \left( B - \frac{B^2}{2} \right) \\ &\quad - \frac{R^M}{2} (N+r) \left( \frac{B}{NB+r} \right)^2 (\sigma^0)^2 - \frac{aN(bN-1+rb)}{2bN^{2\delta^M} K^M} \left( \frac{aB}{(bN-1)B+rb} \right)^2 \end{aligned} \quad (\text{A.2})$$

Thus,

$$\begin{aligned} \Phi(0; N) - \Phi(1; N) &= -\frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{2K^0} \\ &\quad + \frac{R^M}{2} \left( \frac{1}{N+r} \right) (\sigma^0)^2 + \frac{a^2}{2bN^{2\delta^M} K^M} \left( \frac{aN}{bN-1+rb} \right). \end{aligned} \quad (\text{A.3})$$

Recall that  $a = R^M N^{2\delta^M} K^M (\sigma^M)^2$ , and  $b = a + 1$ , and let  $d := R^M (\sigma^M)^2$ . Then, one can

show,

$$\begin{aligned} \text{if } \delta^0, \delta^M > 0, \quad \lim_{N \rightarrow \infty} [\Phi(0; N) - \Phi(1; N)] &= \frac{K^0 R^M (\sigma^M)^2 - q^2}{2K^0}, \\ \text{if } \delta^i = \delta^M = 0, \quad \lim_{N \rightarrow \infty} [\Phi(0; N) - \Phi(1; N)] &= \frac{K^0 (K^M)^2 d^3 - (1+q)^2 \left\{ (K^M)^2 d^2 + 2K^M d + 1 \right\}}{2K^0 (1+K^M d)}, \\ \text{if } \delta^i, \delta^M < 0, \quad \lim_{N \rightarrow \infty} [\Phi(0; N) - \Phi(1; N)] &= -\infty. \end{aligned}$$

Clearly, the equation,  $K^0 (K^M)^2 d^3 - (1+q)^2 \left\{ (K^M)^2 d^2 + 2K^M d + 1 \right\} = 0$ , has a unique solution  $d^*$ , and  $K^0 (K^M)^2 d^3 - (1+q)^2 \left\{ (K^M)^2 d^2 + 2K^M d + 1 \right\} > 0 (=, <, resp.)$  if  $d > (=, <, resp.) d^*$ . Hence the lemma is proved.  $\square$

Now, we are ready to prove the theorem. We prove Part II first.

*Part II: Middle managerial contract sensitivity.*

Let  $\beta_D$  and  $\beta_H$  be middle managerial contracting sensitivities for the direct and hierarchical contracting cases, respectively. Then by (6),  $\beta_D$  satisfies  $1 - \beta_D - a\beta_D = 0$ , i.e.,  $\beta_D = \frac{1}{b}$ , whereas, by (17),  $\beta_H = \frac{(N-1+a)B+r}{(bN-1)B+rb}$ . Hence, we have  $\beta_H - \beta_D = \frac{a^2 B}{b\{(bN-1)B+rb\}} > 0$ . However, regardless of values of  $\delta^i$ ,  $i = 0, M$ , we have

$$0 < \frac{a^2 B}{b\{(bN-1)B+rb\}} < \frac{a^2}{b(bN-1+rb)} \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Therefore,  $\beta_H - \beta_D \rightarrow 0$ , as  $N \rightarrow \infty$ .  $\square$

*Part I: The top managerial contract sensitivity.* As stated in the Theorem, we have three cases, i, ii and iii.

*Case i:*  $\delta^i > 0$ ,  $i = 0, M$ . For this case, we have two subcases: subcase i-a with  $q = 0$  and subcase i-b with  $q > 0$ . The proof of *Case i* is different across these two subcases, because when  $q = 0$ , the top managerial production function is fixed, whereas when  $q > 0$ , the top managerial productivity grows with  $N$ .

*Subcase i-a:* We assume  $q = 0$ . Let  $L := N^{-\delta^0}$ . If  $B \in [\frac{1}{N}\sqrt{L}, 1]$ , then (A.2) implies

$$\begin{aligned} & \frac{1}{L^2} [\Phi(0; N) - \Phi(B; N)] \\ &= -\frac{\left(\frac{N^{-\delta^0}}{L} + \frac{\kappa(N)N^{1-\delta^M}}{L}\right)^2}{K^0} \left(B - \frac{B^2}{2}\right) + \frac{R^M}{2} \frac{1}{L^2} (N+r) \left(\frac{B}{NB+r}\right)^2 (\sigma^0)^2 \\ & \quad + \frac{1}{L^2} \frac{aN(bN-1+rb)}{2bN^{2\delta^M}K^M} \left(\frac{aB}{(bN-1)B+rb}\right)^2 \\ &> -\frac{\left(\frac{N^{-\delta^0}}{L} + \frac{\kappa(N)N^{1-\delta^M}}{L}\right)^2}{2K^0} + \frac{1}{L^2} \frac{aN(bN-1+rb)}{2bN^{2\delta^M}K^M} \left(\frac{a\frac{1}{N}\sqrt{L}}{(bN-1)\frac{1}{N}\sqrt{L}+rb}\right)^2. \end{aligned}$$

Note that the RHS of the inequality is positive for a large  $N$ , because the first term of the RHS is bounded and its second term diverges to  $\infty$ . Thus,  $\Phi(0; N) > \Phi(B; N)$  for all  $B \in [\frac{1}{N}\sqrt{L}, 1]$  for large  $N$ . Therefore, by Lemma 1, we must have

$$B_N^* \in \left(0, \frac{1}{N}\sqrt{L}\right), \quad (\text{A.4})$$

which immediately implies  $B_N^* \rightarrow 0$  as  $N \rightarrow \infty$ . Hence *subcase i-a* is proved.

*Subcase i-b:* Now, we assume  $q > 0$ . In this subcase, the principal's objective function becomes much more complex than it is in subcase i-a. We need Lemma A.2 and Corollary A.1 for some insight into the shape of the principal's objective function.

**Lemma A.2.** *Let  $z_1, z_2, z_3, z_4, z_5$  be positive constants such that  $2 > z_1 > 1 > z_2 > z_3 > z_4$ ,  $\frac{2}{3} < z_2 < 4z_3 - 1$ ,  $3z_3 < 1 + z_4$ , and  $3z_4 + z_5 < 1$ .<sup>24</sup> Suppose either that  $\delta^i > 0$ ,  $i = 0, M$ , and  $q > 0$ , or that  $\delta^i = 0$ ,  $i = 0, M$ , and  $q \geq 0$ . Then, there exists  $J$  such that for all  $N \geq J$ ,  $\Phi(B, N)$  satisfies the following properties: (1)  $\Phi''(B, N) < 0$  on  $[0, \frac{1}{Nz_1}]$ , (2)  $\Phi'(B, N) < 0$  on  $[\frac{1}{Nz_1}, \frac{1}{Nz_2}]$ , (3)  $\Phi''(B, N) > 0$  on  $[\frac{1}{Nz_2}, \frac{1}{Nz_3}]$ , (4)  $\Phi'(B, N) > 0$  on  $[\frac{1}{Nz_3}, 1 - \frac{1}{Nz_5}]$ , (5)  $\Phi''(B, N) < 0$  on  $[1 - \frac{1}{Nz_5}, 1]$ , and (6)  $\Phi'(B, N) < 0$  on  $[1, \infty]$ , where  $\Phi'(B; N)$  and  $\Phi''(B; N)$  are, respectively, the first and second derivatives of  $\Phi(B; N)$  with respect to  $B$ .*

**Proof:** Since the proof of property (6) is given in the proof of Lemma 1, we only prove properties (1) to (5). Calculation shows

$$\begin{aligned} \Phi''(B; N) &= -\frac{\left(N^{-\delta^0} + \kappa(N)N^{1-\delta^M}\right)^2}{K^0} + \frac{R^M r(N+r)(2NB-r)}{(NB+r)^4} (\sigma^0)^2 \\ & \quad + \frac{ra^3 N(bN-1+rb)\{2(bN-1)B-rb\}}{N^{2\delta^M} K^M \{(bN-1)B+rb\}^4}. \end{aligned}$$

<sup>24</sup>The existence of such constants can be easily verified.

(1) Note that since  $z_1 > 1$ , for  $B \in [0, \frac{1}{N^{z_1}}]$  and for a large  $N$ , we have  $2NB - r < 2N \frac{1}{N^{z_1}} - r < 0$  and  $2(bN - 1)B - rb < 2(bN - 1) \frac{1}{N^{z_1}} - rb < b(2N \frac{1}{N^{z_1}} - r) < 0$ . Hence,  $\Phi''(B, N) < 0$  on  $[0, \frac{1}{N^{z_1}}]$  for a large  $N$ .

(2) For  $B \in [\frac{1}{N^{z_1}}, \frac{1}{N}]$ , (A.1) implies

$$\begin{aligned} \Phi'(B; N) &< \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} \\ &\quad - \frac{R^M r(N+r) \frac{1}{N^{z_1}} (\sigma^0)^2}{(N \frac{1}{N} + r)^3} - \frac{ra^3 N \{bN - 1 + rb\} \frac{1}{N^{z_1}}}{N^{2\delta^M} K^M \{(bN - 1) \frac{1}{N} + rb\}^3}. \end{aligned}$$

Since  $z_1 < 2$ , the RHS of this inequality diverges to  $-\infty$ . Thus,  $\Phi'(B, N) < 0$  on  $[\frac{1}{N^{z_1}}, \frac{1}{N}]$  for a large  $N$ . On the other hand, for  $B \in [\frac{1}{N}, \frac{1}{N^{z_2}}]$ , again (A.1) implies

$$\begin{aligned} \Phi'(B; N) &< \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} \\ &\quad - \frac{R^M r(N+r) \frac{1}{N} (\sigma^0)^2}{(N \frac{1}{N^{z_2}} + r)^3} - \frac{ra^3 N \{bN - 1 + rb\} \frac{1}{N}}{N^{2\delta^M} K^M \{(bN - 1) \frac{1}{N^{z_2}} + rb\}^3}. \end{aligned}$$

Since  $z_2 > \frac{2}{3}$ , the RHS of this inequality diverges to  $-\infty$ . Thus,  $\Phi'(B, N) < 0$  on  $[\frac{1}{N}, \frac{1}{N^{z_2}}]$  for a large  $N$ . Therefore,  $\Phi'(B, N) < 0$  on  $[\frac{1}{N^{z_1}}, \frac{1}{N^{z_2}}]$  for a large  $N$ .

(3) For  $B \in [\frac{1}{N^{z_2}}, \frac{1}{N^{z_3}}]$ , (A.1) implies

$$\begin{aligned} \Phi''(B; N) &> - \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} + \frac{R^M r(N+r)(2N \frac{1}{N^{z_2}} - r) (\sigma^0)^2}{(N \frac{1}{N^{z_3}} + r)^4} \\ &\quad + \frac{ra^3 N (bN - 1 + rb) \{2(bN - 1) \frac{1}{N^{z_2}} - rb\}}{N^{2\delta^M} K^M \{(bN - 1) \frac{1}{N^{z_3}} + rb\}^4}. \end{aligned}$$

Since  $z_3 > \frac{1+z_2}{4}$ , the RHS of this inequality diverges to  $\infty$ . Thus  $\Phi''(B; N) > 0$  on  $[\frac{1}{N^{z_2}}, \frac{1}{N^{z_3}}]$  for a large  $N$ .

(4) For  $B \in [\frac{1}{N^{z_3}}, \frac{1}{N^{z_4}}]$ ,

$$\begin{aligned} \Phi'(B; N) &> \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} \left(1 - \frac{1}{N^{z_4}}\right) \\ &\quad - \frac{R^M r(N+r) \frac{1}{N^{z_4}} (\sigma^0)^2}{(N \frac{1}{N^{z_3}} + r)^3} - \frac{ra^3 N \{bN - 1 + rb\} \frac{1}{N^{z_4}}}{N^{2\delta^M} K^M \{(bN - 1) \frac{1}{N^{z_3}} + rb\}^3}. \end{aligned}$$

Since  $3z_3 < 1 + z_4$ , the RHS converges to a positive number. Thus  $\Phi'(B; N) > 0$  on  $[\frac{1}{N^{z_3}}, \frac{1}{N^{z_4}}]$  for a large  $N$ . On the other hand, for  $B \in [\frac{1}{N^{z_4}}, 1 - \frac{1}{N^{z_5}}]$ ,

$$\begin{aligned} \Phi'(B; N) &> \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} \frac{1}{N^{z_5}} \\ &\quad - \frac{R^M r(N+r) (1 - \frac{1}{N^{z_5}}) (\sigma^0)^2}{(N \frac{1}{N^{z_4}} + r)^3} - \frac{ra^3 N \{bN - 1 + rb\} (1 - \frac{1}{N^{z_5}})}{N^{2\delta^M} K^M \{(bN - 1) \frac{1}{N^{z_4}} + rb\}^3}. \end{aligned}$$

Thus,

$$N^{z_5} \Phi'(B; N) > \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} - \frac{R^M r(N+r)(N^{z_5} - 1)}{(N^{\frac{1}{N^{z_4}}} + r)^3} (\sigma^0)^2 - \frac{ra^3 N \{bN - 1 + rb\} (N^{z_5} - 1)}{N^{2\delta^M} K^M \{(bN - 1)^{\frac{1}{N^{z_4}}} + rb\}^3}.$$

Since  $3z_4 + z_5 < 1$ , the RHS approaches a positive number as  $N \rightarrow \infty$ . Thus  $\Phi'(B; N) > 0$  on  $[\frac{1}{N^{z_4}}, 1 - \frac{1}{N^{z_5}}]$  for a large  $N$ . Therefore,  $\Phi'(B; N) > 0$  on  $[\frac{1}{N^{z_3}}, 1 - \frac{1}{N^{z_5}}]$  for a large  $N$ .

(5) Finally, for  $B \in [1 - \frac{1}{N^{z_5}}, 1]$ ,

$$\begin{aligned} \Phi''(B; N) < & -\frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} + \frac{R^M r(N+r)(2N-r)}{(N(1 - \frac{1}{N^{z_5}}) + r)^4} (\sigma^0)^2 \\ & + \frac{ra^3 N (bN - 1 + rb) \{2(bN - 1) - rb\}}{N^{2\delta^M} K^M \{(bN - 1)(1 - \frac{1}{N^{z_5}}) + rb\}^4}. \end{aligned}$$

The RHS approaches a negative number. Thus,  $\Phi''(B; N) < 0$  on  $[1 - \frac{1}{N^{z_5}}, 1]$ .  $\square$

The next corollary immediately follows from Lemmas 1 and A.2.

**Corollary A.1.** *Suppose either that  $\delta^i > 0$ ,  $i = 0, M$  and  $q > 0$ , or that  $\delta^i = 0$ ,  $i = 0, M$  and  $q = 0$ . Then,  $N \geq J$  implies*

$$B_N^* \in \left(0, \frac{1}{N^{z_1}}\right) \cup \left(1 - \frac{1}{N^{z_5}}, 1\right),$$

where  $z_1, z_5$  and  $J$  are as in Lemma A.2.

Now we are ready to prove the *subcase i-b*. If  $N \geq J$  and  $B \in [1 - \frac{1}{N^{z_5}}, 1]$ , then, by concavity (see Lemma A.2), we have

$$\begin{aligned} \Phi(B; N) &\leq \Phi(1; N) + \Phi'(1; N)(B - 1) \leq \Phi(1; N) - \Phi'(1; N) \frac{1}{N^{z_5}} \\ &= \Phi(1; N) + \left[ \frac{R^M r}{(N+r)^2} (\sigma^0)^2 + \frac{ra^3 N}{N^{2\delta^M} K^M \{bN - 1 + rb\}^2} \right] \frac{1}{N^{z_5}}. \end{aligned}$$

Therefore,

$$\begin{aligned} \Phi(0; N) - \Phi(B; N) &\geq \Phi(0; N) - \Phi(1; N) \\ &\quad - \left[ \frac{R^M r}{(N+r)^2} (\sigma^0)^2 + \frac{ra^3 N}{N^{2\delta^M} K^M \{bN - 1 + rb\}^2} \right] \frac{1}{N^{z_5}}. \quad (\text{A.5}) \end{aligned}$$

If  $N \geq J$  and  $B \in [0, \frac{1}{N^{z_5}}]$ , then, by concavity (see Lemma A.2),

$$\begin{aligned}\Phi(B; N) &\leq \Phi(0; N) + \Phi'(0; N)B \leq \Phi(0; N) + \Phi'(0; N)\frac{1}{N^{z_1}} \\ &= \Phi(0; N) + \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} \frac{1}{N^{z_1}}.\end{aligned}$$

Therefore,

$$\Phi(1; N) - \Phi(B; N) \geq \Phi(1; N) - \Phi(0; N) - \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{K^0} \frac{1}{N^{z_1}}. \quad (\text{A.6})$$

Suppose that  $\delta^i > 0$ ,  $i = 0, M$ , and  $q > 0$ . If  $N \geq J$  and  $B \in [1 - \frac{1}{N^{z_5}}, 1]$ , then, (A.4) implies that the RHS of (A.5) tends to  $\frac{K^0 R^M (\sigma^M)^2 - q^2}{2K^0}$  as  $N \rightarrow \infty$ . Hence, if  $K^0 R^M (\sigma^M)^2 - q^2 > 0$ , then there exists  $J' \geq J$  such that  $N \geq J'$  implies  $\Phi(0; N) > \Phi(B; N)$  for all  $B \in [1 - \frac{1}{N^{z_5}}, 1]$ . Thus, by Corollary A.1, for a large  $N$ ,

$$B_N^* \in \left(0, \frac{1}{N^{z_1}}\right), \quad (\text{A.7})$$

which implies  $B_N^* \rightarrow 0$  as  $N \rightarrow \infty$  if  $K^0 R^M (\sigma^M)^2 - q^2 > 0$ . On the other hand, if  $N \geq J$  and  $B \in [0, \frac{1}{N^{z_5}}]$ , then (A.4) implies the RHS of (A.6) tends to  $-\frac{K^0 R^M (\sigma^M)^2 - q^2}{2K^0}$  as  $N \rightarrow \infty$ . An argument similar to the above shows that  $B_N^* \in (1 - \frac{1}{N^{z_5}}, 1)$  for a large  $N$ . That is,  $B_N^* \rightarrow 1$  as  $N \rightarrow \infty$  if  $K^0 R^M (\sigma^M)^2 - q^2 < 0$ . This completes the proof of subcase *i-b*.

*Case ii.* Suppose  $\delta^i = 0$ ,  $i = 0, M$  and  $q \geq 0$ . Then, by the essentially same arguments as in the above subcase *i-b*, (A.4) together with (A.5), and (A.6) imply that if  $R^M (\sigma^M)^2 > d^*$ ,  $B_N^* \in \left(0, \frac{1}{N^{z_1}}\right)$  for a large  $N$ , and that if  $R^M (\sigma^M)^2 < d^*$ ,  $B_N^* \in (1 - \frac{1}{N^{z_5}}, 1)$  for a large  $N$ . Hence  $B_N^* \rightarrow 0$  if  $R^M (\sigma^M)^2 > d^*$ , and  $B_N^* \rightarrow 1$  if  $R^M (\sigma^M)^2 < d^*$ .

*Case iii.* Finally, we consider the case where  $\delta^i < 0$ ,  $i = 0, M$ , and  $q > 0$ . Using (A.2), we have

$$\begin{aligned}\Phi(1; N) - \Phi(B; N) &= \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{2K^0} (1 - B)^2 \\ &\quad - \frac{R^M}{2} (N + r) \left[ \left( \frac{1}{N + r} \right)^2 - \left( \frac{B}{NB + r} \right)^2 \right] (\sigma^0)^2 \\ &\quad - \frac{aN(bN - 1 + rb)}{2bN^{2\delta^M} K^M} \left[ \left( \frac{a}{bN - 1 + rb} \right)^2 - \left( \frac{aB}{(bN - 1)B + rb} \right)^2 \right].\end{aligned}$$

Thus, if  $B \in [0, 1 - N^{\delta^0}]$ , then, we have

$$\begin{aligned} & \Phi(1; N) - \Phi(B; N) \\ & > \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{2K^0}(1-B)^2 - \frac{R^M}{2} \frac{1}{N+r}(\sigma^0)^2 - \frac{a^3 N}{2bN^{2\delta^M} K^M (bN - 1 + rb)} \\ & > \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{2K^0} N^{2\delta^0} - \frac{R^M}{2} \frac{1}{N+r}(\sigma^0)^2 - \frac{a^3 N}{2bN^{2\delta^M} K^M (bN - 1 + rb)}. \end{aligned}$$

The last two terms in the RHS of the second inequality approach zero, because  $\delta^i < 0$ ,  $a = R^M N^{2\delta^M} K^M (\sigma^M)^2$ , and  $b = a + 1$ . Since the first term in the RHS of the second inequality converges to  $\frac{1}{2K^0}$ , we have  $\Phi(1; N) > \Phi(B; N)$  for all  $B \in [0, 1 - N^{\delta^0}]$  if  $N$  is large. Thus, by Lemma 1,  $B_N^* \in (1 - N^{\delta^0}, 1)$  for a large  $N$ . Therefore  $B_N^* \rightarrow 1$  as  $N \rightarrow \infty$ .  $\square$

## F Proof of Propositions 4 and 5

By symmetry, we may assume that  $\alpha^i, \beta^{i0}, \beta^{ij}$  and  $\beta^{ii}$  are respectively identical for  $i, j = 1, \dots, N$  and  $i \neq j$ . In particular, for  $i, j = 1, \dots, N$  and  $i \neq j$ , let  $\alpha^i \equiv \alpha$ ,  $\beta^{i0} \equiv \gamma^0$ ,  $\beta^{ij} \equiv \gamma^M$ ,  $\beta^{ii} \equiv \beta$ . Then, we obtain

$$\begin{aligned} \text{Var}((\beta^i)^\top Y) &= \left( \sum_{j=0}^N \beta^{ij} \right)^2 (\sigma^c)^2 + (\beta^{i0})^2 (\sigma^0)^2 + \sum_{j=1}^N (\beta^{ij})^2 (\sigma^M)^2 \\ &= \left( \gamma^0 + (N-1)\gamma^M + \beta \right)^2 (\sigma^c)^2 \\ &\quad + (\gamma^0)^2 (\sigma^0)^2 + (N-1)(\gamma^M)^2 (\sigma^M)^2 + \beta^2 (\sigma^M)^2, \quad i = 1, \dots, N. \end{aligned}$$

For  $i = 1, \dots, N$ , middle manager  $i$ 's problem is as follows: given  $S^i = \alpha + (\beta^i)^\top Y$ , choose  $\mu^i$  to maximize

$$\begin{aligned} & E[-\exp\{-R^M(\alpha^i + (\beta^i)^\top Y - c^M(\mu^i))\}] \\ &= -\exp\left\{-R^M\left(\alpha^i + (\beta^i)^\top e - c^M(\mu^i) - \frac{R^M}{2}\text{Var}((\beta^i)^\top Y)\right)\right\}. \end{aligned}$$

The first order condition is  $\beta^{ii} = N^{\delta^M} c_\mu^M(\mu^i)$ . Since  $\beta^{ii} \equiv \beta$ , we have  $\mu^i \equiv \mu^M$  and  $\beta = N^{\delta^M} c_\mu^M(\mu^M)$ , which is the same as (20). On the other hand, the middle managerial participation constraint implies that, for  $i = 1, \dots, N$ ,

$$S^i(Y) = \mathcal{W}_0^M - (\beta^i)^\top e + c^M(\mu^M) + \frac{R^M}{2}\text{Var}((\beta^i)^\top Y) + (\beta^i)^\top Y.$$

By substituting  $S^i$  into  $X$ , we have

$$X = \mathbf{1}_{N+1}^\top Y - \sum_{i=1}^N \left( \mathcal{W}_0^M - (\beta^i)^\top e + c^M(\mu^M) + \frac{R^M}{2}\text{Var}((\beta^i)^\top Y) + (\beta^i)^\top Y \right).$$

Thus,  $E[X]$  and  $Var(X)$  are as shown in (25) and (26), respectively. Given  $S^0 = A + BX$ , the top manager's problem is to choose  $\mu^0$ ,  $\gamma^0$ ,  $\gamma^M$ , and  $\mu^M$  to maximize

$$\begin{aligned} & E \left[ -\exp \left\{ -R^0 \{A + BX - c^0(\mu^0)\} \right\} \right] \\ &= E \left[ -\exp \left\{ -R^0 \left( A - c^0(\mu^0) + BE[X] - \frac{R^0}{2} B^2 Var(X) \right) \right\} \right]. \end{aligned}$$

The first order condition with respect to  $\mu^0$  is given by (24). Solving the first order conditions with respect to  $\gamma^0$  and  $\gamma^M$  simultaneously, yields (21) and (22). Substituting (20), (21), and (22), some calculation shows that the first order condition with respect to  $\mu^M$  is (27). Finally, by applying the participation constraint of the top manager to  $S^0 = A + BX$ , we have the top managerial optimal contract as stated in Proposition 5.  $\square$

## G An Outline of the Proof of Theorem 2

The proof of Theorem 2 requires much more complex and tedious algebraic manipulations than that of Theorem 1 does. Here we provide only an outline of the proof.<sup>25</sup>

With quadratic cost functions as in (15), condition (27) implies

$$\beta = \frac{\tau(N - 1 + a)B + (\tau - \varphi)r}{\tau(bN - 1)B + (\tau b - \varphi)r}, \quad (\text{A.8})$$

where  $a := R^M N^{2\delta^M} K^M (\sigma^M)^2$ ,  $b := 1 + a$ , and  $r := \frac{R^M}{R^0}$ . Substituting (A.8) into (28), we reduce the investor's expected profit to  $\Phi(B; N)$ , a function of a single decision variable  $B$ .

*Part 1.* Consider limits of the top managerial sensitivity. We first show that for a large  $N$ ,  $\Phi'(B; N) < 0$  for all  $B \geq 1$ . Thus, if  $N$  is large enough, the maximizer  $B_N^*$  of  $\Phi(B; N)$  must lie in  $[0, 1)$ .

The proof that  $B_N^*$  approaches zero is done by finding a positive constant  $z$ , under the conditions stated in the theorem, such that  $\Phi(0; N) > \Phi(B; N)$  for all  $B \in [\frac{1}{N^z}, 1]$  when  $N$  is

<sup>25</sup>A detailed proof is available upon request.



sufficiently large. Namely, we have

$$\begin{aligned}
& \Phi(0; N) - \Phi(B; N) \tag{A.9} \\
& > -\frac{R^M(\sigma^M)^2\varphi N}{2(\tau - \varphi)} \left( \frac{N - 1 + a}{bN - 1} \right)^2 - \frac{(N^{-\delta^0} + \kappa(N)N^{1-\delta^M})^2}{2K^0} \\
& \quad - \frac{R^M}{2} N \left( \frac{(1 - \eta)(\sigma^M)^2}{b\tau - \varphi} \right)^2 (\sigma^c)^2 \\
& \quad + \frac{R^M}{2} N \left[ \frac{\frac{1}{N^z}}{N\frac{1}{N^z} + r} + \eta \frac{N\frac{1}{N^z}}{N\frac{1}{N^z} + r} + (1 - \eta) \frac{\tau(bN - 1)\frac{1}{N^z} + r(\sigma^M)^2}{\tau(bN - 1)\frac{1}{N^z} + (b\tau - \varphi)r} \right]^2 (\sigma^c)^2 \\
& \quad - \frac{R^M}{2} N \left( \frac{\eta(\sigma^M)^2}{b\tau - \varphi} \right)^2 (\sigma^0)^2 - \frac{R^M}{2} N(N - 1) \left( \frac{\varphi}{b\tau - \varphi} \right)^2 (\sigma^M)^2. \tag{A.10}
\end{aligned}$$

The second and fourth terms of the right-hand side of inequality (A.10) stand out as  $N \rightarrow \infty$ . The second term is the top managerial productivity (equal to effort productivity minus his cost of effort), and the fourth term is the effect of the common risk on the investors' profit. Since, under the stated assumptions, the fourth term dominates the second term as well as the other terms and approaches zero,  $B_N^*$  converges to 0 as  $N \rightarrow \infty$ .

The proof that  $B_N^*$  approaches 1 is done as follows. First, we show that there exists a positive constant  $\epsilon$  such that if  $N$  is large enough, then for all  $B \in [0, \epsilon]$ ,

$$\Phi(1; N) > \Phi(B; N).$$

Then, we show that there exists a positive constant  $z_1$  such that  $\Phi'(B; N) > 0$  on  $[\epsilon, 1 - \frac{1}{n^{z_1}}]$  if  $N$  is large enough.

*Part 2.* Consider the middle managerial sensitivity. Note that by (6), middle managerial contract sensitivity  $\beta_D$  in the direct contracting case satisfies

$$1 - \beta_D - a\beta_D \frac{\tau}{\tau - \varphi} = 0,$$

so that  $\beta_D = \frac{\tau - \varphi}{b\tau - \varphi}$ , whereas by (A.8), in the hierarchical contracting case,  $\beta_H$ , is

$$\beta_H = \frac{\tau(N - 1 + a)B + (\tau - \varphi)r}{\tau(bN - 1)B + (\tau b - \varphi)r}.$$

Hence, we have

$$\beta_H - \beta_D = \frac{a\tau(a\tau + \varphi(N - 1))B}{(b\tau - \varphi)\{\tau(bN - 1)B + (\tau b - \varphi)r\}} > 0.$$

However, we have

$$0 < \frac{a\tau(a\tau + \varphi(N-1))B}{(b\tau - \varphi)\{\tau(bN-1)B + (\tau b - \varphi)r\}} < \frac{a\tau(a\tau + \varphi(N-1))}{(b\tau - \varphi)\{\tau(bN-1) + (\tau b - \varphi)r\}} \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

so that

$$\beta_H - \beta_D \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Therefore, the assertions of the theorem follow.  $\square$

The following corollary is an outcome from the detailed proof of Theorem 2, and turns out to be useful in proving propositions appearing in the rest of the paper.

**Corollary A.2.** 1. If  $\delta^M \geq \delta^0 > 0$  and  $q = 0$ , then we have for a large  $N$ ,

$$B_N^* \in [0, \frac{1}{N^{z_0}}), \quad \text{for any } z_0 \text{ such that}$$

$$z_0 < 2 + \min \left[ 2\delta^M, 1 + 2\delta^0, \frac{5}{2} + \min(\delta^0, 2\delta^M) \right] \quad (\text{A.11})$$

2. If  $\delta^M \geq \delta^0 > 0$  and  $q > 0$ , then, we have for a large  $N$ ,

$$B_N^* \in [0, \frac{1}{N^{z_0}}), \quad \text{for any } z_0 \text{ such that } \frac{3}{2} < z_0 < 2 + \min(2\delta^M, 1). \quad (\text{A.12})$$

3. If  $\delta^M = \delta^0 = 0$ , then we have for a large  $N$ ,

$$B_N^* \in [0, \frac{1}{N^{z_0}}), \quad \text{for any } z_0 \text{ such that } \frac{3}{2} < z_0 < 2. \quad (\text{A.13})$$

**Proof:** The proof of each case is done by finding a constant  $z_0$  such that  $\Phi'(B; N) < 0$  for all  $B \in [\frac{1}{N^{z_0}}, \frac{1}{N^z}]$  if  $N$  is large.

## H Proof of Proposition 6

By definition, we have

$$B_D = \frac{\text{Cov}(S_D^0(Y), X_D)}{\text{Var}(X_D)}.$$

However,

$$\begin{aligned} \text{Cov}(S_D^0(Y), X_D) &= (\beta_D^{00} + N\beta_D^{0M}) \left\{ 1 + N - N \frac{(1-\eta)(\sigma^M)^2}{b\tau - \varphi} \right\} (\sigma^c)^2 + \beta_D^{00} \left( 1 + \frac{N\eta(\sigma^M)^2}{b\tau - \varphi} \right) (\sigma^0)^2 \\ &\quad + N\beta_D^{0M} \left( 1 - \frac{(\sigma^M)^2}{b\tau - \varphi} \right) (\sigma^M)^2, \\ \text{Var}(X_D) &= \left\{ 1 + N - N \frac{(1-\eta)(\sigma^M)^2}{b\tau - \varphi} \right\}^2 (\sigma^c)^2 + \left( 1 + \frac{N\eta(\sigma^M)^2}{b\tau - \varphi} \right)^2 (\sigma^0)^2 \\ &\quad + N \left( 1 - \frac{(\sigma^M)^2}{b\tau - \varphi} \right)^2 (\sigma^M)^2, \end{aligned}$$

where

$$\begin{aligned}
\beta_D^{00} &= \frac{1 + N^{\delta^0} \kappa(N) N^{1-\delta^M}}{\text{DEN}} \times \left[ (\sigma^M)^2 + N(\sigma^c)^2 \left( 1 - \kappa(N) N^{\delta^0 - \delta^M} \right) \right], \\
N\beta_D^{0M} &= -\frac{1 + N^{\delta^0} \kappa(N) N^{1-\delta^M}}{\text{DEN}} \times \left[ N(\sigma^c)^2 - N \left( (\sigma^c)^2 + (\sigma^0)^2 \right) \kappa(N) N^{\delta^0 - \delta^M} \right], \\
\text{DEN} &= (\sigma^M)^2 + N \left[ (\sigma^c)^2 \left( 1 - \kappa(N) N^{\delta^0 - \delta^M} \right)^2 + (\sigma^0)^2 \left( \kappa(N) N^{\delta^0 - \delta^M} \right)^2 \right] \\
&\quad + R^0 N^{2\delta^0} K^0 \left[ \left( (\sigma^c)^2 + (\sigma^0)^2 \right) (\sigma^M)^2 + N(\sigma^c)^2 (\sigma^0)^2 \right],
\end{aligned}$$

and thus

$$\beta^{00} + N\beta^{0M} = \frac{1 + N^{\delta^0} \kappa(N) N^{1-\delta^M}}{\text{DEN}} \times \left[ (\sigma^M)^2 + (\sigma^0)^2 N^{\delta^0} \kappa(N) N^{1-\delta^M} \right].$$

We have the following four cases depending on  $\sigma^c$  and  $q$ .

1. Assume  $\sigma^c > 0$ .

(a) Suppose  $q = 0$ . Then, we have

$$\begin{aligned}
\beta_D^{00} &= \frac{(\sigma^M)^2 + N(\sigma^c)^2}{\text{DEN}}, \\
N\beta_D^{0M} &= -\frac{N(\sigma^c)^2}{\text{DEN}}, \\
\text{DEN} &= (\sigma^M)^2 + N(\sigma^c)^2 + R^0 N^{2\delta^0} K^0 \left[ \left( (\sigma^c)^2 + (\sigma^0)^2 \right) (\sigma^M)^2 + N(\sigma^c)^2 (\sigma^0)^2 \right].
\end{aligned}$$

Hence, we have

$$\beta^{00} + N\beta^{0M} = \frac{(\sigma^M)^2}{\text{DEN}}.$$

It is straightforward to show that

$$N^{2+2\delta^0} B_D \quad \text{approaches a positive constant.} \tag{A.14}$$

Thus  $B_D \rightarrow 0$  as  $N \rightarrow \infty$ .

(b) Suppose  $q > 0$ . It is also straightforward to show that

$$N^2 B_D \quad \text{approaches a positive constant.} \tag{A.15}$$

Thus  $B_D \rightarrow 0$  as  $N \rightarrow \infty$ .

2. Assume  $\sigma^c = 0$ . Then, we have

$$\begin{aligned}\beta_D^{00} &= \frac{1 + N^{\delta^0} \kappa(N) N^{1-\delta^M}}{\text{DEN}} (\sigma^M)^2, \\ N\beta_D^{0M} &= \frac{1 + N^{\delta^0} \kappa(N) N^{1-\delta^M}}{\text{DEN}} N(\sigma^0)^2 \kappa(N) N^{\delta^0 - \delta^M}, \\ \text{DEN} &= (\sigma^M)^2 + N(\sigma^0)^2 \left( \kappa(N) N^{\delta^0 - \delta^M} \right)^2 \\ &\quad + R^0 N^{2\delta^0} K^0 (\sigma^0)^2 (\sigma^M)^2,\end{aligned}$$

and that

$$\beta^{00} + N\beta^{0M} = \frac{1 + N^{\delta^0} \kappa(N) N^{1-\delta^M}}{\text{DEN}} \times \left[ (\sigma^M)^2 + (\sigma^0)^2 N^{\delta^0} \kappa(N) N^{1-\delta^M} \right].$$

(a) Suppose that  $q = 0$ . Then one can easily see that

$$N^{1+2\delta^0} B_D \text{ approaches a positive constant.} \quad (\text{A.16})$$

Thus,  $B_D \rightarrow 0$  as  $N \rightarrow \infty$ .

(b) Suppose that  $q > 0$ . Then it can be easily checked that

$$NB_D \text{ approaches a positive constant.} \quad (\text{A.17})$$

Thus  $B_D \rightarrow 0$  as  $N \rightarrow \infty$ .

Therefore, the assertion follows.  $\square$

## I Proof of Proposition 7

I. Consider the case where  $\delta^i > 0$ , for  $i = 0, M$ .

(a) If  $K^0 R^M (\sigma^M)^2 - q^2 < 0$ , then it is clear that  $B^* > B_D$  for a large  $N$ , because  $B^*$  approaches the first best while  $B_D$  approaches 0.

(b) Consider the case where  $K^0 R^M (\sigma^M)^2 - q^2 > 0$ .

i. Suppose that  $q = 0$ . Then by (A.4), we have

$$B^* \in \left( 0, \frac{1}{N^{1+\frac{\delta^0}{2}}} \right).$$

However, by (A.1), for  $B \in \left[ \frac{1}{N^z}, \frac{1}{N^{1+\frac{\delta^0}{2}}} \right]$  where  $z$  is any positive constant such that  $1 + \frac{\delta^0}{2} < z < 3 + 2\delta^0$ , we have

$$\Phi'(B; N) \leq \frac{N^{-2\delta^0}}{K^0} - \frac{ra^3 N \{bN - 1 + rb\} \frac{1}{N^z}}{N^{2\delta^M} K^M \left\{ (bN - 1) \frac{1}{N^{1+\frac{\delta^0}{2}}} + rb \right\}^3}.$$

The RHS of this inequality diverges to  $-\infty$ . Thus,  $\Phi(B; N)$  decreases on the interval  $\left[\frac{1}{Nz}, \frac{1}{N^{1+\frac{\delta^0}{2}}}\right]$ , which implies, for a large  $N$ ,

$$B^* \in \left(0, \frac{1}{Nz}\right),$$

where  $z$  is any positive constant such that  $1 + \frac{\delta^0}{2} < z < 3 + 2\delta^0$ . Hence, by (A.16), we get  $B^* < B_D$  for a large  $N$ .

- ii. Suppose that  $q > 0$ . Recall that  $z_1 > 1$  in (A.7). Thus, the assertion follows by (A.7) and (A.17).

II. The proof of the assertion for the case where  $\delta^i = 0$  for  $i = 0, M$  is similar to that of part I. □

## J Proof of Proposition 8

1. Consider the case where  $q > 0$ . If  $\delta^i > 0, i = 0, M$ , then, by (A.12) and (A.15), it is clear that  $B^* < B_D$  for a large  $N$ .

Suppose that  $\delta^i = 0, i = 0, M$ . Let  $z_0$  be given as in (A.13). Also let the positive constant in (A.15) which  $N^2 B_D$  approaches be  $\lambda$ . Then one can write

$$B_D = \frac{\lambda}{N^2} + \frac{\varepsilon_N}{N^2},$$

where  $\lim_{N \rightarrow \infty} \varepsilon_N = 0$ . For  $B \in [B_D, \frac{1}{Nz_0}]$ , it can be shown that  $\frac{\Phi'(B; N)}{N}$  is less than or equal to a negative constant for a large  $N$  so that  $\Phi'(B; N) < 0$  on  $[B_D, \frac{1}{Nz_0}]$  if  $N$  is large. Hence, we have  $B^* < B_D$  for a large  $N$ .

2. Consider the case where  $q = 0$  and  $1 > \delta^M \geq \delta^0 > 0$ . Note that in (A.11),

$$\min \left[ 2\delta^M, 1 + 2\delta^0, \frac{5}{2} + \min(\delta^0, 2\delta^M) \right] = \min \left[ 2\delta^M, 1 + 2\delta^0 \right].$$

- (a) If  $\min \left[ 2\delta^M, 1 + 2\delta^0 \right] > 2\delta^0$ , (or equivalently,  $\delta^M > \delta^0$ ), then we can choose  $z_0$  in (A.11) so that  $z_0 > 2 + 2\delta^0$ . Hence by (A.14), for a large  $N$ ,  $B^* < B_D$ .

- (b) If  $\min \left[ 2\delta^M, 1 + 2\delta^0 \right] = 2\delta^0$ , (or equivalently,  $\delta^0 = \delta^M$ ), then it is straightforward to show that the positive constant in (A.14) which  $N^{2+2\delta^0} B_D$  approaches is equal to  $\frac{1}{R^0 K^0 (\sigma^c)^2}$ . Thus, we can write

$$B_D = \frac{1}{R^0 K^0 (\sigma^c)^2 N^{2+2\delta^0}} + \frac{\varepsilon_N}{N^{2+2\delta^0}},$$

where  $\lim_{N \rightarrow \infty} \varepsilon_N = 0$ . For  $B \in [B_D, \frac{1}{N^{z_0}}]$ , where  $2 \leq z_0 < 2 + 2\delta^0$ , it can be shown that  $N^{2\delta^0-1} \times \Phi'(B; N)$  is less than or equal to a negative constant for a large  $N$ . Hence,  $B^* < B_D$  for a large  $N$ .  $\square$

## K Agency problem under hierarchial contracting in continuous time

In the text of the paper, all results are based on the assumption that all contracts are linear in outcomes. In fact, this assumption is just for the sake of simplicity and without loss of generality. We provide a brief explanation of a continuous-time version of our hierarchical contracting model under which all linear optimal contracts derived in the text are optimal indeed. The proof of the linearity result is similar to Koo, Shim, and Sung (2005), and thus it is omitted.<sup>26</sup>

We take the unit-interval  $[0, 1]$  as a time horizon. The production of the team output can be described as follows: The effort of manager  $i$ , for  $i = 0, 1, \dots, N$ , results in outcome  $Y_t^i$  at time  $t \in [0, 1]$ , dynamics of which are given by

$$dY_t^i = (\mu_t^i + \kappa^i(N)\mu_t^0) N^{-\delta^i} dt + \sigma^c dB_t^c + \sigma^i dB_t^i,$$

where  $\kappa^0 = 0$ ,  $\kappa^i \geq 0$  for  $i = 1, \dots, N$ ;  $\sigma^c$ , and  $\sigma^i$  for  $i = 0, 1, \dots, N$  are constant diffusion coefficients; and  $B_t = (B_t^c, B_t^0, B_t^1, \dots, B_t^N)^\top$  is an  $N + 2$ -vector of independent standard Brownian motions defined on a probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ . We assume that the filtration  $(\mathcal{F}_t)$  is the completion of the filtration generated by the Brownian motion  $B_t$ .

Let us use the following notation:

$$Y_t = (Y_t^0, \dots, Y_t^N)^\top \quad \mu_t = (\mu_t^0, \dots, \mu_t^N)^\top, f_t = (f_t^0, \dots, f_t^N)^\top,$$

We assume that middle managerial salary functions  $\mathcal{F}_t$ -measurable random variables of the following class:<sup>27</sup> for  $i = 1, \dots, N$ ,

$$S^i(Y) = F^i(Y) + \int_0^1 \alpha^i(t, Y) dt + \int_0^1 (\beta^i(t, Y))^\top dY_t, \quad (\text{A.18})$$

where both  $\alpha^i$  and  $\beta^i(t, Y) \equiv (\beta^{i0}(t, Y), \dots, \beta^{iN}(t, Y))^\top$  for  $i = 0, 1, \dots, N$  are bounded  $\mathcal{F}_t$ -predictable, and  $F^i$  is bounded and  $\mathcal{F}_1$ -measurable. On the other hand, the top managerial salary  $S^0$  is based on the profit process  $X_1 \equiv \sum_{i \in A} Y_1^i - \sum_{i \in A^c} S^i(Y)$ . That is,  $S^0$  is  $X_1$ -measurable and  $S^0(X_1)$ .

<sup>26</sup>A detailed proof is available upon request.

<sup>27</sup>See Schättler and Sung (1993) for a detailed explanation of admissible salary functions.

The principal is risk neutral and managers' preferences can be characterized by exponential utility functions with the coefficient of absolute risk aversion is equal to  $R^i$  for  $i \in A$ . We assume that the diffusion rates  $\sigma^c, \sigma^a$  for  $a \in A$  are common knowledge to the principal and managers.

The principal's problem for hierarchical contracting is stated as follows:

**Problem 1.** Choose  $\{(\mu^i, S^i); i = 0, \dots, N\}$  to maximize

$$E [X_1 - S^0(X)] \quad \text{subject to}$$

$$(1) \quad X_1 = \sum_{i=0}^N Y_1^i - \sum_{i=1}^N S^i(Y)$$

$$(2) \quad \{\mu^0, (\mu^i, S^i; i = 1, \dots, N)\}$$

$$\in \arg \max_{\{\hat{\mu}^0, (\hat{\mu}^i, \hat{S}^i; i=1, \dots, N)\}} E \left[ -\exp \left\{ -R^0 \left\{ S^0(X_1) - \int_0^1 c^0(\hat{\mu}_t^0) dt \right\} \right\} \right]$$

$$\text{s.t. (i)} \quad dY_t^i = (\hat{\mu}_t^i + \kappa^i(N)\hat{\mu}_t^0) N^{-\delta^i} dt + \sigma^c dB_t^c + \sigma^i dB_t^i, \quad \forall i,$$

$$X_1 = \sum_{i=0}^N Y_1^i - \sum_{i=1}^N S^i(Y),$$

$$(ii) \quad \forall i = 1, \dots, N,$$

$$\hat{\mu}^i \in \arg \max_{\{\bar{\mu}_t^i\}} E \left[ -\exp \left\{ -R^i \left\{ \hat{S}^i(Y) - \int_0^1 c^i(\bar{\mu}_t^i) dt \right\} \right\} \right]$$

$$\text{s.t.} \quad dY_t^i = (\bar{\mu}_t^i + \kappa^i(N)\bar{\mu}_t^0) N^{-\delta^i} dt + \sigma^c dB_t^c + \sigma^i dB_t^i,$$

$$dY_t^j = (\hat{\mu}_t^j + \kappa^j(N)\hat{\mu}_t^0) N^{-\delta^j} dt + \sigma^c dB_t^c + \sigma^j dB_t^j, \quad \forall j \neq i$$

$$(iii) \quad \forall i = 1, \dots, N,$$

$$E \left[ -\exp \left\{ -R^i \left\{ \hat{S}^i - \int_0^1 c^i(\hat{\mu}_t^i) dt \right\} \right\} \right] \geq -\exp \{-R^i \mathcal{W}_0^i\}$$

$$(3) \quad E \left[ -\exp \left\{ -R^0 \left\{ S^0(X) - \int_0^1 c^0(\mu_t^0) dt \right\} \right\} \right] \geq -\exp \{-R^0 \mathcal{W}_0^0\}.$$

For the problem stated above, one can show that all contracts for both the top and all middle managers are linear in  $X_1$  and  $Y_1$ , respectively, by modifying proofs given in our companion paper, Koo, Shim and Sung [2006]. Furthermore, optimal contracts for the above problem are identical to those in the text.

## References

1. Aggarwal, R., and A. Samwick, "Performance Incentives within Firms: The Effect of Management Responsibility," *Journal of Finance* 58, 2003, 1613–1649.
2. Aghion, P. and J. Tirole, "Formal and Real Authority in Organizations," *Journal of Political Economy*, 105, 1997, 1-29.
3. Antle, R., and A. Smith, "An empirical investigation of the relative performance evaluation of corporate executives," *Journal of Accounting Research* 24, 1–39.
4. Baker G. P. and B. Hall, "CEO Incentives and Firm Size," *Journal of Labor Economics* 22, 2004, 767–798.
5. Bitler, M., T. Moskowitz, and A. Vissing-Jorgensen, "Testing Agency Theory with Entrepreneurial Effort and Wealth," *Journal of Finance* 60, 2005, 539–576.
6. Bolton, Patrick and Mathias Dewatripont, "The Firm as a Communication Network," *Quarterly Journal of Economics* 108, November 1994, 809–839.
7. Coase, R. H. "The Nature of the Firm," *Economica* 30, November 1937, 386–405.
8. Calvo, G. and S. Wellisz, "Supervision, Loss of Control, and Optimum Size of the Firm" *Journal of Political Economy* 86, April 1978, 943–952.
9. Calvo, G. and S. Wellisz, "Hierarchy, Ability, and Income Distribution," *Journal of Political Economy* 87, October 1979, 991–1010.
10. Core, J. E. and W. Guay, "Stock option plans for non-executive employees ," *Journal of Financial Economics* 61, 2001, 253–287.
11. Core, J. E. and W. Guay, "When efficient contracts require risk-averse executives to hold equity: Implications for option valuation, for relative performance evaluation, and for the corporate governance debate ," working paper, 2003, The Wharton School, University of Pennsylvania.
12. Dye, R. A., "Relative Performance Evaluation and Project Selection," *Journal of Accounting Research* 30, Spring 1992, 27–52.
13. De Motta, A., "Managerial Incentives and Internal Capital Markets," *Journal of Finance* 58, 2003, 1193–1220.



14. Gibbons, R. and K. Murphy, "Relative performance evaluation for chief executive officers," *Industrial and Labor Relations Review* 43, 1990, 30–51.
15. Groves, T. "Incentives in Teams," *Econometrica* 41, 617–631.
16. Hall, B. J. and J. Liebman, "Are CEOs really paid like bureaucrats," *Quarterly Journal of Economics* 113, August 1998, 653–691.
17. Hart, O., and J. Moore, "On the design of Hierarchies: Communication versus Specialization," *Journal of Political Economy*, 113, 2005, 675-702.
18. Haubrich, J. G., "Risk Aversion, Performance Pay, and the Principal-Agent Problem," *Journal of Political Economy*, 102(2), 1994, 258-276.
19. Jensen, Michael C., and Kevin J. Murphy: "Performance Pay and Top-Management Incentives," *Journal of Political Economy*, Vol. 98, 1990, 225-264.
20. Koo, H., G. Shim, and J. Sung, "Optimal Multi-Agent Performance Measures for Team Contracts," 2006, working paper.
21. Lazear, E. and S. Rosen, "Rank-order Tournaments as Optimum Labor Contracts," *Journal of Political Economy* 89, 1981, 841–864.
22. Melumad, N.D., D. Mookherjee, and S. Reichelstein, "Hierarchical Decentralization of Incentive Contracts," *Rand Journal of Economics*, 26, 1995, 654-672.
23. Mookherjee, D., "Optimal Incentive Schemes with Many Agents," *Review of Economic Studies*, 51, 1984, 433–446.
24. Marschak, Thomas and Stefan Reichelstein, "Network Mechanisms, Informational Efficiency, and Hierarchies," *Journal of Economic Theory* 79, 1998, 106–141.
25. Mirrlees, J.A. "The Optimal Structure of Incentives and Authority within an Organization," *Bell Journal of Economics*, 1976, 105-131.
26. Mookherjee, D., "Optimal Incentive Schemes with Many Agents," *Review of Economic Studies* 51, 1984, 433–446.
27. Nalebuff, B. and J. Stiglitz, "Prizes and Incentives: toward a general theory of compensation and competition," *Bell Journal of Economics*, 1983, 21–43.
28. Qian, Y. "Incentives and Loss of Control in an Optimal Hierarchy," *Review of Economic Studies*, 61, 1994, 527–544.

29. Radner, R., "The Organization of Decentralized Information Processing," *Econometrica* 61, September 1993, 1109–1146.
30. Radner, R., "Hierarchy: The Economics of Managing," *Journal of Economic Literature* 30, September 1992, 1382–1415.
31. Rosen, S., "Authority, Control and the Distribution of Earnings," 13, 1982, 311-323.
32. Schaefer, S. "The Dependence of Pay-performance Sensitivity on the Size of the Firm," *Review of Economics and Statistics*, 1998, 436–443.
33. Schättler, H., and J. Sung, "The First-Order Approach to Continuous-Time Principal-Agent Problem with Exponential Utility," *Journal of Economic Theory*, 61, 1993, 331-371.
34. Stein, J. C., "Information Production and Capital Allocation: Decentralized vs. Hierarchical Firms," *Journal of Finance* 57, 2002, 1981-1921.
35. Sung, J., "Linearity with Project Selection and Controllable Diffusion Rate in Continuous-Time Principal-Agent Problems," *Rand Journal of Economics*, 26, 1995, 720-743.
36. Sung, J., "Optimal Contracts under Adverse Selection and Moral Hazard; a Continuous-Time Approach," *Review of Financial Studies*
37. Williamson, O.E. "Hierarchical Control and Optimum Firm Size," *Journal of Political Economy* 75, April 1967, 123–138.