Mechanism Design with Collusive Supervision: A Three-tier Agency Model with a Continuum of Types, including Applications to Organizational Design◆

Yutaka Suzuki

Faculty of Economics, Hosei University
4342 Aihara, Machida-City, Tokyo 194-0298, Japan
E-Mail: yutaka@hosei.ac.jp

Revised, December 22, 2009

ABSTRACT

We apply the "Monotone Comparative Statics" method à la Topkis (1978), Edlin and Shannon (1998), and Milgrom and Segal (2002)'s generalized envelope theorem to the three-tier agency model with hidden information and collusion à la Tirole (1986, 1992), thereby providing a framework that can address the issues treated in the existing literature, e.g., Kofman and Lawarree (1993)'s auditing application, *in a much simpler fashion.* Using its tractable framework, we examine some interesting extensions, such as the effect of introducing another supervisor, the problem resulting from a lack of the principal's commitment, and the effect of incorporating behavioral elements into the model. In addition, we derive some clear and robust implications applicable to corporate governance reform, such as a choice between the companies with auditors vs. committees as a top management organization.

Key Words: Mechanism Design, Collusion, Supervision, Monotone Comparative Statics, Corporate Governance, Behavioral Economics

JEL Classification: D82, D86

◆This is an extended, full paper version of Suzuki (2008). I would like to thank the audiences at Hosei University, Association for Public Economic Theory (PET08), Seoul, and the 1st Annual UECE Lisbon Meetings: Game Theory and Applications (2009), Lisbon, for their useful comments. This research was supported by Grant-in-Aid for Scientific Research by Japan Society for the Promotion of Science(C) No. 20530162 (2008-2010).

1. Introduction

Recently, auditing has rapidly been increasing in importance in Japan, as well as in the U.S. and Western countries, to meet the needs of corporate governance. Corporate scandals such as those that rocked Yamaichi Securities, Daiwa Bank, Snow Brand Milk Products, and Kanebo in Japan and Enron and WorldCom in the U.S. are examples of firms that failed to build up the effective corporate governance, and collusive supervision (auditing) and revelation of false information was a common occurrence. Auditors (supervisors) usually have greater access to accurate information on the agents, but are subject to collusive pressure (the collusive offer) from the auditees (agents). The means by which adequate supervision (auditing) is used to enhance the efficiency of corporate governance and by which collusive supervision (auditing) can be deterred are important parts of corporate governance reform.

In a typical framework of the top management organization of Japanese firms, a shareholders' meeting elects a director (or a Board of Directors) and an auditor who audits the execution of the management work and makes a report at the shareholders' meeting. With this auditing system, which has been legally amended several times, it is often said that auditors have access to a great deal of information inside the firm, including the ability of top managers to perform their jobs, while on the other hand it is doubtful that the auditor can objectively supervise the management while maintaining his independence. Indeed, there is a notion that collusive auditing often exists where an auditor and a manager collude to manipulate information. Thus, corporations should optimally utilize the auditing information in order to increase the shareholders' interests, with an arrangement that the auditor and the manager do not collude. Many Japanese firms, such as Toyota and Canon, do preserve and try to improve this traditional Japanese auditing system. However, some companies with auditors, falling into low performance under collusive auditing, tend to move to those with committees, where the monitoring of the manager is tightened and the independence of supervision is ensured by employing outside directors as a majority of the committee members. Our paper can be viewed as an analysis of a top management organization in a hidden information setting.

Literature exists that deals with the issues associated with corporate governance and auditing in a three-tier agency model with collusion, developed by Tirole (1986, 1992) and Laffont and Tirole (1991), Laffont and Martimort (1997) etc. In particular, Kofman and Lawarree (1993) applied a three-tier agency model—consisting of the two-type (productivity) agent, the internal and external auditors (supervisors), and the principal—to the issue of auditing and collusion.[1] However, this is a rather complicated model whose structure involves a Kuhn- Tucker problem with many IC (Incentive Compatibility) and IR (Individual Rationality) constraints, and is not a simple mathematical model. This mathematical complexity of this model is a disadvantage.

We introduce here the outcomes of "Monotone Comparative Statics" à la Topkis (1978), Milgrom and Roberts (1990), Edlin and Shannon (1998), and Milgrom and Segal (2002) into the analysis of corporate governance in a three-tier agency model with a continuum of types. Our paper provides a framework that can address the issues treated in the existing literature *in a much simpler fashion*, and is indeed beneficial in that we can obtain some clear and robust implications for corporate governance reform.

The basic tradeoff in our model is the benefit from the reduction in information rent by adding the auditor (supervisor) versus the resource cost of adding him into the hierarchy, and this bottom line is preserved through the extension and generalization of the model. The optimal collusion-proof contract in the Principal-Supervisor-Agent three-tier regime has the property whereby (1) *Efficiency at the top* (the highest type) and (2) *Downward distortion* for all other types, and the downward distortion is mitigated at the optimum, in comparison with the Principal-Agent two-tier regime. The optimal solution allows a simple comparative statics, which shows that downward

[1] Bolton and Dewatripont (2005)'s recent textbook presents a simple version of the collusion models (Tirole (1986), Kofman and Lawarree (1993)).

distortions from the first best output levels diminish when the accuracy of supervision increases and the efficiency of collusion declines. This is a specific contribution to the literature. Whether the principal indeed has an incentive to introduce a supervisor—that is, selects a three-tier hierarchy—depends on the balance between the net benefits from both the improvement of marginal incentives and the reduction in information rent and the resource cost of the auditor (supervisor). We obtain these results by constructing a three-tier model with a mathematically more tractable structure, which exploits the outcome of "Monotone Comparative Statics" à la Topkis (1978) and Edlin and Shannon (1998), and Milgrom and Segal (2002)'s generalized envelope theorem.

Though we basically consider a situation where the principal can *commit* to the collusion-proof contract, that is, 'full commitment', we analyze as an extension what happens when the principal cannot fully commit to the mechanism and renegotiation is unavoidable. When the principal commits herself to the supervisor reward scheme, but does not commit to the one for the agent, she will be tempted to modify the initial contract (or the outcome) unilaterally, using the information revealed by the supervisor. This situation is similar to the ratchet problem and the renegotiation problem caused by the lack of the principal's commitment in the dynamics of the incentive contracts, studied early by Laffont-Tirole (1988) and Dewatripont (1988) etc. If the agent anticipates such a modification, since he can benefit from a failure by the supervisor to report his type truthfully, he will offer the supervisor the transfer (side payment) equivalent to his information rent. Thus, the principal must pay to the supervisor in opposition to the collusive offer by the agent. Hence, the principal can strictly improve his payoff ex post, but must bear the ex ante incentive cost. In this situation, we analyze whether the principal can do better in the equilibrium without her some commitment than if she can fully commit.

As another extension, behavioral elements à la Fahr and Schmidt (1999) are incorporated into the model, and their effects on the optimal solution are examined in the principal-supervisor-agent hidden information model with collusion. Concretely, we assume that the agent and the supervisor enjoy "private benefits" or "psychological benefits" from output achievement, which is non-monetary and non-transferable, such as job satisfaction and pride/self esteem generated through more output. It is then found that the behavioral elements can reduce the monetary reward for inducing the true information. Hence, the virtual surplus for each type is increased by the reduction of information rent (an incentive cost for inducing a truthful information revelation). Thus, the optimal solution with behavioral elements becomes greater than the one with no behavioral elements. This will have an important implication for the organizational design.

In summary, we apply the monotone comparative statics method to the three-tier agency model with hidden information and collusion, thereby providing a framework that can address the issues treated in the existing literature *in a much simpler fashion.* By using its tractable framework, we examine some interesting extensions, such as the effect of introducing another supervisor, a problem created by the lack of the principal's commitment, and the effect of incorporating behavioral elements. Finally, we derive some clear and robust implications that can be applicable to corporate governance reform and organizational design.

2. Principal-Agent Hidden Information Model with a Continuum of Types

2.1 Setting

We consider two players: a principal (P) and an agent (A). The principal owns the firm and hires the manager (agent) to run it. θ is the manager's ability to run the firm and $C(X,\theta)$ is the effort cost for the manager of type θ to attain the output X. For each θ, $C(X,\theta)$ satisfies $C(X,\theta)>0$, $\partial C(X,\theta)/\partial X>0, \partial^2 C(X,\theta)/\partial X^2>0, \forall X\in\mathbb{R}_+$. W is the wage payment the agent receives, and so his utility is $W-C(X,\theta)$. We normalize the agent's reservation utility as 0. The timing of the game is as follows. Prior to contracting, θ is determined randomly by nature and is known only to the manager (agent). The principal proposes a take-it-or-leave-it contract offer to the manager. The contract is written as $W(X)$, where X is the output level by the manager and W is the wage he receives if he generates X. If the manager accepts the offer, a contract is signed and the principal is fully committed. If he rejects the offer, the game ends.

2.2 Preliminary: Single Crossing Property (SCP) and Monotonicity of Agent's Choice

Faced with a wage scheme $W(X)$, the agent of type θ will choose

$$X \in \arg\max_{X \in \mathrm{X}} \left[W(X) - C(X, \theta)\right]$$

Analysis is dramatically simplified when the Agent's types can be ordered so that higher types choose a higher output when faced with any wage. We identify when solutions to the parameterized maximization program $\max\limits_{X \in \mathrm{X}} U(X, \theta) := W(X) - C(X, \theta)$ are strictly increasing in the parameter θ. A key property to ensure monotone comparative statics is the following:

Definition 1 *A function $U : X \times \theta \to \mathbb{R}$ where $X, \theta \subset \mathbb{R}$ has the **Single Crossing Property (SCP)** if $U_X(X, \theta)$ exists and is strictly increasing in $\theta \in \Theta$.*[2]

$U(X, \theta) = W(X) - C(X, \theta)$ has SCP if $U_X(X, \theta) = W_X(X) - C_X(X, \theta)$ exists and is strictly increasing in $\theta \in \Theta$ for all $X \in \mathrm{X}$. In this case, $U(X, \theta)$ satisfies SCP when the marginal cost of output $C_X(X, \theta)$ is decreasing in type θ, i.e., higher types always have gentler indifference curves. SCP implies that large increases in X are less costly for higher parameters θ.

Theorem 1 (Edlin and Shannon 1998)
Let $\theta'' > \theta'$, $X' \in \arg\max\limits_{X \in \mathrm{X}} U(X, \theta')$, and $X'' \in \arg\max\limits_{X \in \mathrm{X}} U(X, \theta'')$. Then, if U has SCP, and either X' or X'' is in the interior of X, then $X'' > X'$.

We can apply Theorem 1 to the agent's choice when facing a wage scheme $W(\cdot)$, assuming that the agent's cost $C(X, \theta)$ satisfies SCP. To ensure full separation of types, we need to assume that the wage $W(\cdot)$ is differentiable. Then, $U(X, \theta)$ will satisfy SCP, and Theorem 1 implies that interior output choices are strictly increasing in types, i.e., we have *full separation*.

2.3 The Full information Benchmark

As a benchmark, we consider the case in which the Principal observes the Agent's type θ. Given θ, she offers the bundle (X, W) to solve:

$$\max_{(X,W) \in \mathrm{X} \times \mathbb{R}} X - W(X) \text{ s.t. } W(X) - C(X, \theta) \geqslant 0 \text{ (IR)}$$

(IR) is the Agent's Individual Rationality constraint, and (IR) binds at a solution. Hence, the Principal eventually solves: $\max\limits_{X \in \mathrm{X}} X - C(X, \theta)$ This is exactly the Total Surplus maximization. Let $X^{FB}(\theta)$ denote a solution, which we call the First Best (FB) solution. Using Theorem 1, we check whether our assumptions ensure that $X^{FB}(\theta)$ is strictly increasing in type θ. If $C(X, \theta)$ satisfies SCP, which implies that Total Surplus $X - C(X, \theta)$ satisfies SCP, and if $X^{FB}(\theta)$ is in the interior for each θ, we can conclude that $X^{FB}(\theta)$ is strictly increasing in θ.

Now we consider a different contract from the contract $W : X \to \mathbb{R}$ which we have considered so far, where the agent is asked to announce his type $\hat{\theta}$, and receives payment $W\left(\hat{\theta}\right)$ in exchange for

[2] Edlin and Shannon (1998) introduced this SCP under the name of "increasing marginal returns".

an output $X\left(\hat{\theta}\right)$ on the basis of his announcement $\hat{\theta}$. This is called a *Direct Revelation Contract.* According to the *Revelation Principle*, any contract $W : X \to \mathbb{R}$ can be replaced with a *Direct Revelation Contract* that has an equilibrium in which all types receive the same bundles as in the original contract $W : X \to \mathbb{R}$.

2.4 Solution with a Continuum of Types

Let the type space be continuous: $\Theta = \left[\underline{\theta}, \overline{\theta}\right]$, with the cumulative distribution function $F(\cdot)$, and with a strictly positive density $f(\theta) = F'(\theta)$. In addition to previous assumptions, we assume that $C(X, \theta)$ is continuously differentiable in θ for all X, and $C_\theta(X, \theta)$ is bounded uniformly across (X, θ). The principal's problem is:

$$\max_{\langle X(\cdot), W(\cdot)\rangle} \int_{\underline{\theta}}^{\overline{\theta}} [X(\theta) - W(\theta)] f(\theta)\, d\theta$$

$$\begin{array}{llll} \text{s.t.} & W(\theta) - C(X(\theta), \theta) \geqslant W(\hat{\theta}) - C\left(X\left(\hat{\theta}\right), \theta\right) & (\mathrm{IC}_{\theta\hat{\theta}}) & \forall \theta, \hat{\theta} \in \Theta \\ & W(\theta) - C(X(\theta), \theta) \geqslant 0 & (\mathrm{IR}_\theta) & \forall \theta \in \Theta \end{array}$$

Just as in the two-type case, out of all the participation constraints, only the lowest type's IR binds.

Lemma 1 *At a solution $(X(\cdot), W(\cdot))$, all IR_θ with $\theta > \underline{\theta}$ are not binding, and only $IR_{\underline{\theta}}$ is binding.*

As for the analysis of ICs with a continuum of types, Mirrlees (1971) introduced a widely used way to reduce the number of incentive constraints by replacing them with the corresponding First-Order Conditions. The "trick" is as follows.

(IC) can be written as $\theta \in \arg\max_{\hat{\theta}\in\Theta} U\left(\hat{\theta}, \theta\right)$, where $U\left(\hat{\theta}, \theta\right) = W(\hat{\theta}) - C\left(X\left(\hat{\theta}\right), \theta\right)$ is the utility that the agent of type θ receives by announcing that his type is $\hat{\theta}$. If $\theta \in \left(\underline{\theta}, \overline{\theta}\right)$ and $U\left(\hat{\theta}, \theta\right)$ is differentiable in $\hat{\theta}$, then the first order condition $\partial U\left(\hat{\theta}, \theta\right) \Big/ \partial\hat{\theta}\Big|_{\hat{\theta}=\theta} = 0$ is necessary for the above optimality. We define the Agent's equilibrium utility (the value):

$$U(\theta) \equiv U(\theta, \theta) = W(\theta) - C(X(\theta), \theta)$$

Note that this utility depends on θ in two ways – through the agent's true type and through his announcement. Differentiating with respect to θ, we have $U'(\theta) = U_{\hat{\theta}}(\theta, \theta) + U_\theta(\theta, \theta)$, where the first derivative of U is with respect to the agent's announcement (the first argument) and the second derivative is with respect to the agent's true type (the second argument). Since the first derivative equals zero by $\partial U\left(\hat{\theta}, \theta\right) \Big/ \partial\hat{\theta}\Big|_{\hat{\theta}=\theta} = 0$, we have $U'(\theta) = U_\theta(\theta, \theta)$. This condition is nothing but the well known *Envelope Theorem*: the full derivative of the value of the agent's maximization problem with respect to the parameter – his type – equals to the partial derivative holding the agent's optimal announcement fixed. More concretely,

$$\frac{dU\left(\hat{\theta}, \theta\right)}{d\theta} = \frac{\partial\left[W(\hat{\theta}) - C\left(X\left(\hat{\theta}\right), \theta\right)\right]}{\partial\hat{\theta}} \times \frac{d\hat{\theta}}{d\theta} + \frac{\partial\left[-C\left(X\left(\hat{\theta}\right), \theta\right)\right]}{\partial\theta}$$

Since $\partial\left[W(\hat{\theta}) - C\left(X\left(\hat{\theta}\right), \theta\right)\right] \Big/ \partial\hat{\theta} = 0$ at $\hat{\theta} = \theta$ (the agent's optimal announcement is *Truth Telling*), we have the *envelope condition*:

$$U'(\theta) = \frac{dU(\theta, \theta)}{d\theta} = -\frac{\partial C(X(\theta), \theta)}{\partial\theta}.$$

By integrating it, we have the important formula:

$$U(\theta) = U(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} \frac{\partial C(X(\tau), \tau)}{\partial \tau} d\tau \qquad \textbf{(ICFOC)}$$

(ICFOC) demonstrates that with a continuum of types, *incentive compatibility constraints* pin down up to a constant plus all types' utilities for a given output rule $X(\cdot)$. This remarkable result is derived from the generalized Envelope Theorem by Milgrom and Segal (2002).

Intuitively, (ICFOC) incorporates local incentive constraints, ensuring that the Agent does not gain by slightly misrepresenting θ. By itself, it does not ensure that the Agent cannot gain by misrepresenting θ by a large amount. For example, (ICFOC) is consistent with the truthful announcement $\hat{\theta} = \theta$ being a local maximum, but not a global one. It is even consistent with truthful announcement being a local minimum.

Fortunately, these situations can be ruled out. For this purpose, recall that by SCP, Topkis (1978) and Edlin and Shannon (1998) establish that the agent's output choices from any tariff (and therefore in any incentive compatible contract) are nondecreasing in type. Thus, any piecewise differentiable IC contract must satisfy that $X(\cdot)$ is nondecreasing (**M**).

It turns out that under SCP, ICFOC in conjunction with (M) do ensure that truthtelling is a global maximum, i.e., all ICs are satisfied:

Lemma 2 $(X(\cdot), W(\cdot))$ *is **Incentive Compatible** if and only if **both (ICFOC) and (M) hold**, where $U(\theta) = W(\theta) - C(X(\theta), \theta)$. In summary, "Incentive Constraints $\Leftrightarrow$ First Order Condition (ICFOC) + Monotonicity (M)"*

Proof See, Appendix 1

Given (ICFOC), we can express transfers: $\underbrace{W(\theta)}_{\text{Wage Payment}} = \underbrace{C(X(\theta), \theta)}_{\text{Effort Cost}} + \underbrace{U(\theta)}_{\substack{\text{Information Rent} \\ \text{given for type } \theta}}$

3. Collusion and Supervision

3.1 Introduction of a Supervisor and the Collusion-proof Problem

Now, we introduce a supervisor into the model. The principal can have access, at a cost z, to a supervisor who can, for each θ, provide a proof of this fact with probability p, and with $1-p$, is unable to obtain any information.[3] We assume that proofs of θ cannot be falsified. In other words, θ is hard information. On the other hand, the agent can potentially benefit from a failure by the supervisor to truthfully report that his type is θ when the supervisor observed the signal θ. A self-interested supervisor colludes with the agent only if he benefits from such behavior. We assume the following collusion technology: if the agent offers the supervisor a transfer (side payment) t, he benefits up to kt, where $k \in [0,1]$. That is, only a fraction, $k \in [0,1]$, of the agent's bribe ends up in the supervisor's hands. The idea is that transfers of this sort may be hard to organize and subject to resource losses. We follow the literature in assuming that side-contracts of this sort are *enforceable* (See, e.g., Tirole 1992).

To avoid collusion, the principal will have to offer the supervisor a reward $W_s(\theta)$ for providing θ, such that the following *coalition incentive compatibility constraint* is satisfied.

$$W_s(\theta) \geqslant kU(\theta) = k\left[U(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} \frac{\partial C(X(\tau), \tau)}{\partial \tau} d\tau\right]$$

[3] We assume that the agent correctly knows whether the supervisor is informed of his type information θ or not. This is the same assumption as the earlier literature, e.g., Tirole (1986).

Indeed, once the information θ is obtained, the principal will reduce the Agent θ's payment $W(\theta)$ to effort cost $C(X(\theta),\theta)$, and not pay the information rent $U(\theta)$ to the agent θ. The agent is thus ready to pay the supervisor an amount of $U(\theta)$, and the value of this side payment to the supervisor is $kU(\theta)$, where $k \in [0,1]$. Therefore, hiring a supervisor and eliciting his information requires the principal to pay $W_s(\theta) = kU(\theta), \forall\theta$ to the supervisor if the (hard) information of θ is provided. Substituting $W_s(\theta) = kU(\theta)$ into the Principal's objective function, the virtual surplus for type θ in the Principal-Supervisor-Agent regime is $X(\theta) - C(X(\theta),\theta) - [(1-p)+pk]\,U(\theta)$

Hence, the program of designing the optimal collusion-proof contract can be rewritten as

$$\max_{X(.),U(.)} \int_{\underline{\theta}}^{\overline{\theta}} \left[\underbrace{X(\theta) - C(X(\theta),\theta)}_{\text{Total Surplus}} - [(1-p)+pk] \underbrace{U(\theta)}_{\substack{\text{Information}\\ \text{Rent}}} \right] f(\theta)\, d\theta - z$$

$$\begin{array}{lll} \text{s.t.} & dX(\theta)/d\theta \geqslant 0\text{: } X(\theta) \text{ is nondecreasing} & \text{(M)} \\ & U(\theta) = U(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} \frac{\partial C(X(\tau),\tau)}{\partial \tau} d\tau & \text{(ICFOC)} \\ & U(\underline{\theta}) = W(\underline{\theta}) - C(X(\underline{\theta}),\underline{\theta}) \geqslant \underline{u}(\text{Const.}) & (\text{IR}_{\underline{\theta}}) \end{array}$$

Note that the objective function takes the familiar form of the expected difference between total surplus and the Agent's information rent.

3.2 Solving the Relaxed Problem

Thus, the problem can be rewritten as

$$\max_{X(.)} \int_{\underline{\theta}}^{\overline{\theta}} \left[X(\theta) - C(X(\theta),\theta) - [(1-p)+pk] \left(U(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} \frac{\partial C(X(\tau),\tau)}{\partial \tau} d\tau \right) \right] f(\theta) d\theta - z$$

$$\text{s.t. } dX(\theta)/d\theta \geqslant 0 \ (M) \ \forall\theta$$

where $\int_{\underline{\theta}}^{\overline{\theta}} \left[U(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} \frac{\partial C(X(\tau),\tau)}{\partial \tau} d\tau \right] f(\theta)\, d\theta$ can be called the expected information rents.

Lemma 3

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[U(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} \frac{\partial C(X(\tau),\tau)}{\partial \tau} d\tau \right] f(\theta)\, d\theta = U(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta} \frac{1-F(\theta)}{f(\theta)} f(\theta)\, d\theta$$

Proof See, Appendix 2

Substituting these expected information rents into the principal's program, and ignoring the constant $U(\underline{\theta})$, the program becomes

$$\max_{X(.)} \int_{\underline{\theta}}^{\overline{\theta}} \left[X(\theta) - C(X(\theta),\theta) + [(1-p)+pk] \frac{\partial C(X(\theta),\theta)}{\partial \theta} \frac{1-F(\theta)}{f(\theta)} \right] f(\theta) d\theta - z$$

$$\text{s.t. } dX(\theta)/d\theta \geqslant 0 \ (M) \ \forall\theta$$

We ignore the Monotonicity Constraint (M) and solve the resulting *relaxed program*. Thus, the principal maximize the expected value of the expression within the square brackets, which is called the *virtual surplus*, and denoted by $J(X,\theta)$. This expected value is maximized by simultaneously maximizing virtual surplus for (almost) every type θ, i.e.,

$$X^S(\theta) \in \arg\max_{X(\cdot)} X(\theta) - C(X(\theta),\theta) + [(1-p)+pk] \left[\frac{1-F(\theta)}{f(\theta)} \right] \frac{\partial C(X(\theta),\theta)}{\partial \theta}$$

This defines the optimal output rule $X^S(\cdot)$ for the relaxed program. The principal's choice of $X^S(\theta)$ can be understood as a trade-off between maximizing the total surplus for type θ and reducing the information rents of all types above θ, just as in the two-type case. Indeed, **(ICFOC)** says that output choice X for type θ results in additional information rent $-\partial C(X(\theta),\theta)/\partial\theta$ for all types above θ.

In particular, for the highest type $\overline{\theta}$, there are no higher types, i.e., $F(\overline{\theta}) = 1$ and the principal just maximizes total surplus, choosing $X^S(\overline{\theta}) = X^{FB}(\overline{\theta})$. In words, we have *efficiency at the top*. For all other types, the principal will distort output to reduce information rents. To see the direction of distortion, consider the parameterized maximization program

$$\max_{X\in \mathrm{X}} \Psi(X,\gamma) = X(\theta) - C(X(\theta),\theta) + \gamma\left[\frac{1-F(\theta)}{f(\theta)}\right]\frac{\partial C(X(\theta),\theta)}{\partial\theta}$$

Here $\gamma = 0$ corresponds to surplus-maximization (first-best), and $\gamma = 1\,(p = 0, k \in [0,1])$ corresponds to the principal's (relaxed) second best program with only one agent.

Note that $\frac{\partial\Psi(X,\gamma)}{\partial X\partial\gamma} = \left[\frac{1-F(\theta)}{f(\theta)}\right]\frac{\partial^2 C(X(\theta),\theta)}{\partial X\partial\theta} < 0$ for $\theta < \overline{\theta}$ since the agent's value $U(X,\theta) = W(X) - C(X,\theta)$ has the single crossing property (SCP), that is, $\partial^2 U(X,\theta)/\partial X\partial\theta = -\partial^2 C(X,\theta)/\partial X\partial\theta > 0$. Therefore, $\Psi(X,\gamma)$ has SCP in $(X,-\gamma)$, and by Theorem 1 (Edlin and Shannon), we have $X^*(\gamma = 1) \Leftrightarrow X(\theta) < X^{FB}(\theta) \Leftrightarrow X^*(\gamma = 0)$ for all $\theta < \overline{\theta}$. In words, the principal makes all types other than the highest type underproduce in order to reduce the information rents of types above them. Similarly, by introducing the supervisor, which basically corresponds to $0 < \gamma < 1$, we have

$$X^*(\gamma = 1) \Leftrightarrow X(\theta) < X^*(\gamma \in (0,1)) \Leftrightarrow X^S(\theta) \leqslant X^*(\gamma = 0) \Leftrightarrow X^{FB}(\theta).$$

Hence, in the Principal-Supervisor-Agent regime, the principal can induce more marginal incentives than the second best regime with only one agent through the reduction in total and marginal information rents paid to the supervisor and the agent θ, in other words, reducing the implementation costs for any $X < X^S(\overline{\theta}) = X^{FB}(\overline{\theta})$. Figure 1 depicts this result.

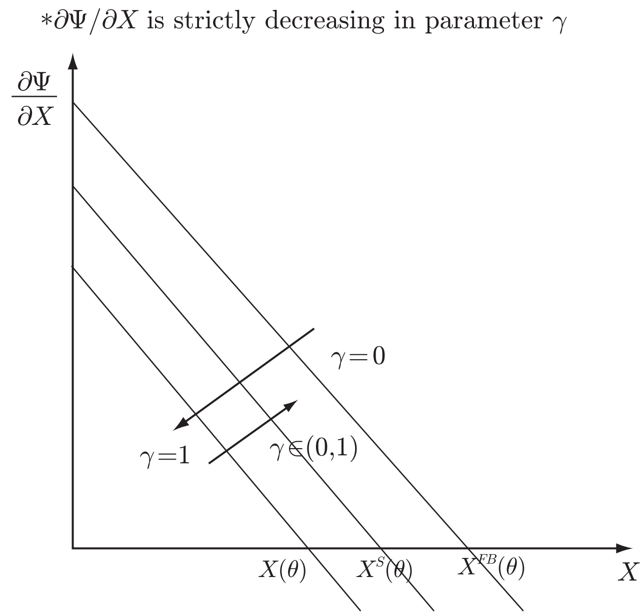

Figure 1

Thus, we obtain the following proposition.

Proposition 1 *In the Principal-Supervisor-Agent regime with a continuum of types, the optimal collusion-proof contract has the property that*

(1) Efficiency at the top (the highest type $\overline{\theta}$) $X(\overline{\theta}) = X^{FB}(\overline{\theta})$
(2) Downward distortion for all other types $\theta \in [\underline{\theta}, \overline{\theta})$ is mitigated, that is,

$$X(\theta) \underbrace{\leqslant}_{\substack{\text{Equality} \\ \text{holds at } k=1}} X^S(\theta) \underbrace{\leqslant}_{\substack{\text{Equality holds} \\ \text{either at } p=1,k=0 \\ \text{or } \theta=\overline{\theta}}} X^{FB}(\theta).$$

Now, remember that we ignored the monotonicity constraint (M) and solved the *relaxed program.* So, we need to check that the solution $X^S(\theta)$ indeed satisfies the monotonicity constraint (M), that is, the output rule $X^S(\theta)$ is nondecreasing. We can check it using Theorem 1. To simplify expressions, define $h(\theta) \equiv f(\theta)/[1 - F(\theta)] > 0$, which is called the *hazard rate* of type θ. Then, the principal's program can be rewritten as

$$\max_{X \in \mathrm{X}} J(X, \theta) = X - C(X, \theta) + \frac{[(1-p) + pk]}{h(\theta)} \frac{\partial C(X, \theta)}{\partial \theta}$$

By Topkis (1978) and Theorem 1, assuming that $C(X, \theta)$ is sufficiently smooth, a sufficient condition for $X^S(\theta)$ to be nondecreasing in θ is for the following derivative to be strictly increasing in θ:

$$\frac{\partial J(X, \theta)}{\partial X} = 1 - \frac{\partial C(X, \theta)}{\partial X} + \frac{[(1-p) + pk]}{h(\theta)} \frac{\partial^2 C(X, \theta)}{\partial X \partial \theta} \tag{*}$$

Since $-C(X, \theta)$ satisfies SCP, the second term is strictly increasing in θ, and the first term does not depend on θ. The only problematic term, therefore, is the third term. Our result is ensured when the third term is nondecreasing in θ. Since $1/h(\theta)$ is positive and $\partial^2 C(X, \theta)/\partial X \partial \theta$ is negative, this is ensured when $\partial^2 C(X, \theta)/\partial X \partial \theta$ is nondecreasing. That is, we have

Proposition 2 *A sufficiency condition for the optimal collusion-proof solution $X^S(\theta)$ to satisfy the monotonicity constraint (M) is that the following conditions hold.*
1. *$\partial^2 C(X, \theta)/\partial X \partial \theta$ is nondecreasing in θ.*
2. *The hazard rate $h(\theta)$ is nondecreasing.*

Example: The first assumption is satisfied e.g., in the following cost function forms:

$$C(X, \theta) = (X - \theta)^{\alpha} \text{ and } C(X, \theta) = (X/\theta)^{\alpha}, \alpha \geqslant 2$$

The second condition is called the "Monotone Hazard Rate Condition" and satisfied by many familiar probability distributions. Now, we can present the following proposition on the comparative statics.

Proposition 3 *Suppose that the sufficiency condition in proposition 2 holds. Then, the optimal collusion-proof solution $X^S(\theta)$ is nondecreasing in the parameter p, and nonincreasing in the parameter k.*

Proof: From the equation $(*)$, the derivative $J_X(X, \theta)$ is nondecreasing in the parameter p, because the derivative of $J_X(X, \theta)$ in the parameter p is $-1 + k \leqslant 0$ for $k \in [0, 1]$, multiplied by the negative terms. Hence, from the Theorem 1, the optimal solution $X^S(\theta)$ is nondecreasing in the parameter p. Particularly, $X^S(\theta)$ is strictly increasing in p for $k \in [0, 1)$ from Theorem 1. The latter part can also be proved in the same way: The derivative $J_X(X, \theta)$ is nonincreasing in the parameter k for $p \in [0, 1]$, and thus the optimal solution $X^S(\theta)$ is nonincreasing in the parameter k. ■

Figure 2 depicts this result.

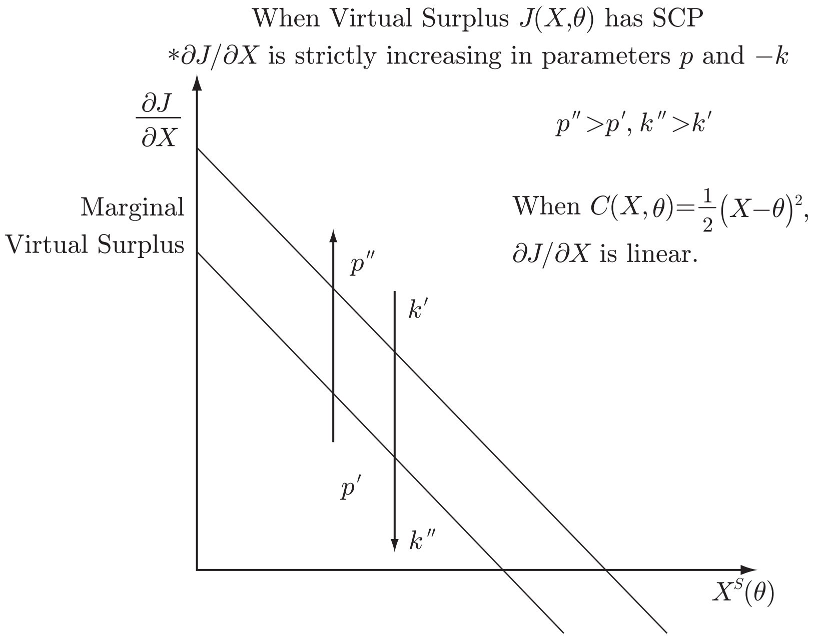

Figure 2

This result could be said to demonstrate the advantage of our approach, because the extensions of the Tirole (1986) model, such as Laffont and Tirole (1991), Kofman and Lawarree (1993), Laffont and Martimort (1997), and Suzuki (1999), often have the complicated structure of a Kuhn-Tucker problem with many IC and IR constraints, and so the global characterization of the optimal solutions as well as the robust comparative statics are difficult to obtain, and only a local characterization of the solution and comparative statics is possible in the above collusion literature, while on the other hand, we can readily perform a robust (monotone) comparative statics, and the rationale of the results is clear and intuitive.

We present economic insight on corporate governance. Under collusive supervision (auditing), that is, $p \downarrow$ and $k \uparrow$, the optimal collusion-proof solution (output) $X^S(\theta)$ by the agent (manager) becomes lower , as does the principal's (shareholder's) payoff. Such lower performance firms should move to some organizational form achieving $p \uparrow$ and $k \downarrow$. Hence, a company with committees could be said to be one of the desirable forms, in that it tightens the monitoring of the agent (manager) $p \uparrow$ and ensures the independence of supervision $k \downarrow$ by employing outside directors as a majority of committee members.

4. Improvement by Adding Another Supervisor

Here, another supervisor is introduced, who is honest (not strategic), but only with a smaller probability $p' \, (\leqslant p)$ can observe the signal θ. We assume for simplicity that the states which he can observe are included in the ones which the main supervisor can observe, and that it is a common knowledge. In this setting, when the main supervisor tries to tell a lie (hides information θ) collusively, the sub-supervisor observes the signal θ with probability p', and reports it to the principal *at no incentive cost*, since he is honest (not strategic). In this case, the main supervisor can not obtain any positive information rent. Thus, the expected gain for the main supervisor when he observes the signal θ will be reduced to $(p - p')\, kU(\theta)$. Bringing in an additional supervisor can help, even if it costs z', provided he is honest: the sub-supervisor can work as a checking device for collusion and reduce the information rent of the main-supervisor. Due to the reduction of the expected information rent, the marginal incentive of the agent will also be increased in equilibrium.

Let us formally check this argument. The principal maximizes the virtual surplus $J(X,\theta)$,

$$\max_{X\in \mathrm{X}} J(X,\theta) = X(\theta) - C(X(\theta),\theta) + \frac{[(1-p)+(p-p')k]}{h(\theta)}\frac{\partial C(X,\theta)}{\partial \theta}$$

The first order condition for the optimum is,

$$\frac{\partial J(X,\theta)}{\partial X} = 1 - \frac{\partial C(X,\theta)}{\partial X} + \frac{[(1-p)+(p-p')k]}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X \partial \theta} = 0$$

Since $(1-p)+(p-p')k \leqslant (1-p)+pk, \forall p' \in [0,p]$, we have the following proposition on the comparison of the equilibrium incentives.

Proposition 4 *Supposing that $X^{S'}(\theta)$ is the solution of this regime, we obtain:*

$$X(\theta) \leqslant X^{S}(\theta) \leqslant X^{S'}(\theta) \leqslant X^{FB}(\theta) \ \textit{ for all } \theta \in [\underline{\theta}, \overline{\theta}]$$

If the expected reduction in the information rent $p'kU(\theta)$ is greater than the resource cost z' of the sub-supervisor, the principal has indeed an incentive to introduce a sub-supervisor into the organization. That is, another auditor can serve as an incentive mechanism not only for the main-auditor (main-supervisor) but also for the management (agent). This simple argument gives a rationale for the auditing system consisting of main- and sub-supervisors (auditors) found in corporate governance reform.

5. A Problem from Lack of Commitment

So far, we have considered a situation where the principal can *commit* to the collusion-proof contract. That is, 'full commitment'. Here, we examine more explicitly the timing of the game. The principal has access to the supervisor, who chooses a message m $\in \{\emptyset, \theta\}$, where $\emptyset$ means that he did not obtain any information. If the principal receives the message from the supervisor that the type information is θ, the principal will have an incentive to modify the original contract. The principal can raise her payoff by *eliminating the downward distortions in all other types* than the highest one $\overline{\theta}$. Namely, instead of $\{X(\theta), W(\theta)\}$, she will offer the efficient (first best) contract $\{X^{FB}(\theta), W^{FB}(\theta)\}$, and the information rent $U(\theta)$ will be exploited by the principal. In summary, the principal commits herself to the reward scheme for the supervisor, but does not commit to the one for the agent. Thus, she is tempted to modify the initial contract (or the outcome $\{X(\theta), W(\theta)\}$) unilaterally, using the information revealed by the supervisor.[4]

If the agent of type θ anticipates this modification, since he can benefit from a failure by the supervisor to report his type θ truthfully, he will offer the supervisor the transfer (side payment) $t = U(\theta)$, the amount equivalent to his information rent, of which the supervisor benefits up to kt, where $k \in [0,1]$. Thus, the principal must pay $W_S(\theta) = kU(\theta)$ to the supervisor in opposition to the collusive offer by the agent, in order to elicit true information. In summary, the principal can strictly improve his payoff ex-post by changing $X(\theta)$ into $X^{FB}(\theta)$, but must bear the ex-ante incentive cost $kU(\theta)$. This is the trade-off for the principal when the supervisor obtains the proof of true information, with probability p.

Only when the supervisor cannot obtain any information for θ with probability $1-p$, does the principal commit herself to the initial scheme $\{X(\theta), W(\theta)\}\forall\theta$, and the same trade-off between the total surplus and the information rent emerges.

[4] This idea is similar to the ratchet effect and the renegotiation problem caused by lack of a principal's commitment in the dynamics of incentive contracts, which were studied earlier by Laffont-Tirole (1988), and Dewatripont (1988) etc.

The virtual surplus for type θ in the Principal-Supervisor-Agent regime is written as

$$(1-p)\left[X(\theta)-C\left(X(\theta),\theta\right)\right]+\underbrace{p}_{\substack{\theta \text{ is} \\ \text{revealed}}}\times\left[\underbrace{X^{FB}(\theta)-C\left(X^{FB}(\theta),\theta\right)}_{\text{(Ex post) First Best Allocative Efficiency}}\right]-\left[(1-p)+pk\right]U(\theta)$$

Eventually, in this regime, the principal maximizes the *virtual surplus* $J(X,\theta)$,

$$\max_{X\in \mathrm{X}} J(X,\theta)=(1-p)\left[X(\theta)-C\left(X(\theta),\theta\right)\right]+\frac{\left[(1-p)+pk\right]}{h(\theta)}\frac{\partial C(X,\theta)}{\partial\theta}$$

The first order condition for the optimum is,

$$\begin{aligned}\frac{\partial J(X,\theta)}{\partial X}=&(1-p)\left[1-\frac{\partial C(X,\theta)}{\partial X}\right]+\frac{\left[(1-p)+pk\right]}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta}=0\\ \Leftrightarrow\ &\underbrace{1-\frac{\partial C(X,\theta)}{\partial X}}_{\text{Marginal Total Surplus}}+\underbrace{\frac{\left[1+\frac{p}{1-p}k\right]}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta}}_{\text{Marginal Information Rent}}=0\end{aligned}$$

Noting that the marginal information rent for each $\theta\in\left[\underline{\theta},\overline{\theta}\right)$ becomes larger than any other former regimes, we have the following proposition on the comparison of equilibrium incentives.

Proposition 5 *Supposing that $X^{NC}(\theta)$ is the solution (in the no-information phase Ø) of this 'No-Commitment' regime, we obtain:*

$$X^{NC}(\theta)\leqslant X(\theta)\leqslant X^{S}(\theta)\leqslant X^{S'}(\theta)\leqslant X^{FB}(\theta)\ \text{ for all } \theta\in\left[\underline{\theta},\overline{\theta}\right]$$

Figure 3 depicts this result.

*No Commitment Regime $X^{NC}(\theta)$ vs. Collusion-proof, Commitment Regime $X^{S}(\theta)$

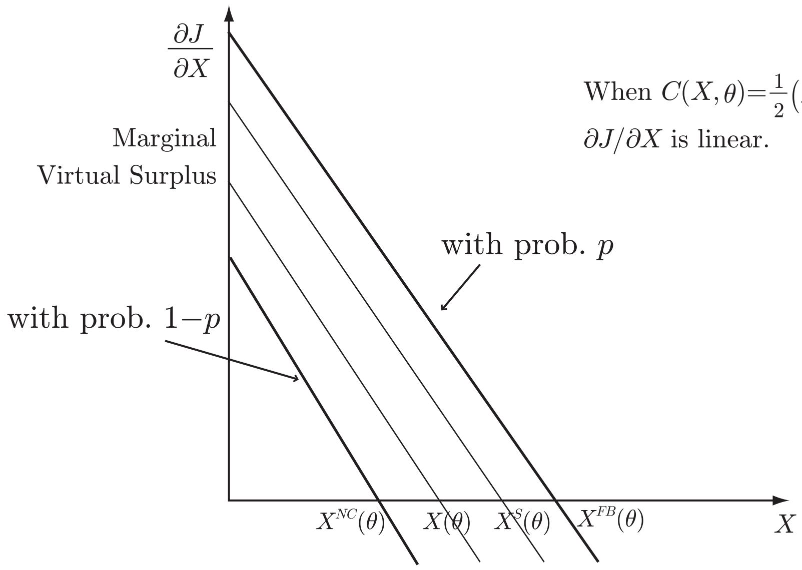

Figure 3

Note that the comparison between the payoffs for the principal is ambiguous. It depends on the relative size in the three terms of the 'No-Commitment' regime (NC) and the 'Principal-Supervisor-Agent' regime (S) with full commitment.

The expected payoff for the principal in the NC regime is

$$(1-p)\int_{\underline{\theta}}^{\overline{\theta}}\left[X^{NC}(\theta)-C\left(X^{NC}(\theta),\theta\right)\right]f(\theta)d\theta+p\times\int_{\underline{\theta}}^{\overline{\theta}}\left[X^{FB}(\theta)-C\left(X^{FB}(\theta),\theta\right)\right]f(\theta)d\theta$$
$$+\left[(1-p)+pk\right]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C\left(X^{NC}(\theta),\theta\right)}{\partial\theta}f(\theta)d(\theta)$$

The expected payoff for the principal in Principal-Supervisor-Agent commitment regime (S) is

$$(1-p)\int_{\underline{\theta}}^{\overline{\theta}}\left[X^{S}(\theta)-C\left(X^{S}(\theta),\theta\right)\right]f(\theta)d\theta+p\times\int_{\underline{\theta}}^{\overline{\theta}}\left[X^{S}(\theta)-C\left(X^{S}(\theta),\theta\right)\right]f(\theta)d\theta$$
$$+\left[(1-p)+pk\right]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C\left(X^{S}(\theta),\theta\right)}{\partial\theta}f(\theta)d(\theta)$$

The comparison in relative size of the three terms is summarized in the following chart.

No Commitment Regime (NC): $X^{NC}(\theta)$	Collusion-proof, Commitment Regime: $X^{S}(\theta)$
$(1-p)\int_{\underline{\theta}}^{\overline{\theta}}\left[X^{NC}(\theta)-C\left(X^{NC}(\theta),\theta\right)\right]f(\theta)d\theta$	$(1-p)\int_{\underline{\theta}}^{\overline{\theta}}\left[X^{S}(\theta)-C\left(X^{S}(\theta),\theta\right)\right]f(\theta)d\theta$ (Bigger expected total surplus w.p. $1-p$)
$p\int_{\underline{\theta}}^{\overline{\theta}}\left[X^{FB}(\theta)-C\left(X^{FB}(\theta),\theta\right)\right]f(\theta)d\theta$ (Bigger ex post expected total surplus w.p. p)	$p\int_{\underline{\theta}}^{\overline{\theta}}\left[X^{S}(\theta)-C\left(X^{S}(\theta),\theta\right)\right]f(\theta)d\theta$
$\left[(1-p)+pk\right]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C\left(X^{NC}(\theta),\theta\right)}{\partial\theta}f(\theta)d(\theta)$	$\left[(1-p)+pk\right]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C\left(X^{S}(\theta),\theta\right)}{\partial\theta}f(\theta)d(\theta)$ (Bigger expected information rent)

Thus, we obtain the following proposition:

Proposition 6 *When the relative gain from the second term in the 'No Commitment' regime overcomes the relative loss from the sum of the first and third terms, the principal can do better in the equilibrium without her some commitment (NC), than if she can fully commit (S).*

From the above trade-off table, we see that as p becomes bigger, the advantage in the ex-post expected total surplus becomes bigger in the "No-Commitment" regime. Note that in Figure 3 the outer First Best line is realized with probability p. This may be consistent with a situation, where in companies with committees, the committee (the supervisor in our model) accurately grasps the state (type information) of the agent (operating officer) with a high probability and the first best scheme is imposed for the agent.

6. Incorporating Behavioral Elements into the Model

In this section, we incorporate behavioral elements à la Fahr and Schmidt (1999) into the model.[5] We assume that the agent and the supervisor enjoy "private benefits" or "psychological benefits"

[5] The impact of Fehr and Schmidt (1999) has been tremendous. It is a logical approach to the behavioral economics and explains multitude of evidence. Suzuki (2007) considers a setting where the existence of behavioral elements with a zero-sum structure leads to a strong incentive for vertical collusion in the principal-supervisor-two agent hierarchy, and analyzes the optimal (incomplete) contract design problem.

from output achievement, which is non-monetary and non-transferable, such as job satisfaction and pride/self esteem generated through more output. As a behavioral assumption, we formulate these "private benefits" or "psychological benefits" in a simple way such that

$$s_i\left(X(\theta)\right) = s_i X(\theta), i = A, S, \text{where } 0 \leqslant s_s \leqslant s_A$$

Note that $s_i(i = A, S)$ is a constant psychological benefit per unit output for agent (A) or supervisor (S). Then the payoff for type θ when he made an announcement of $\hat{\theta}$ is

$$U\left(\hat{\theta},\theta\right) = \underbrace{W\left(\hat{\theta}\right)}_{\substack{\text{Monetary}\\ \text{wage payment}}} + \underbrace{s_A\left(X\left(\hat{\theta}\right)\right)}_{\substack{\text{Non-monetary}\\ \text{psychological benefit}}} - C\left(X\left(\hat{\theta}\right),\theta\right)$$

Equilibrium utility (value) after truth telling (IC) constraint is imposed is

$$U(\theta) = W(\theta) + s_A\left(X\left(\theta\right)\right) - C\left(X\left(\theta\right),\theta\right)$$

¿From the envelope theorem, we have

$$\frac{d}{d\theta}\left[\max_{\hat{\theta}} W\left(\hat{\theta}\right) + s_A\left(X\left(\hat{\theta}\right)\right) - C\left(X\left(\hat{\theta}\right),\theta\right)\right] = -\left.\frac{\partial C\left(X\left(\hat{\theta}\right),\theta\right)}{\partial\theta}\right|_{\hat{\theta}=\theta}$$

That is, $\frac{dU(\theta)}{d\theta} = -\frac{\partial C(X(\theta),\theta)}{\partial\theta}$

Hence, we can express the monetary transfer $W(\theta)$ as follows.

$$\underbrace{W(\theta)}_{\text{Wage Payment}} = \underbrace{C(X(\theta),\theta)}_{\text{Effort Cost}} + \underbrace{U(\theta)}_{\substack{\text{Information Rent}\\ \text{given for type } \theta}} - \underbrace{s_A X(\theta)}_{\substack{\text{Behavioral Element}\\ \text{for type } \theta}}$$

Similarly, in order to avoid collusion, the principal offers the supervisor a monetary reward $W_s\left(\theta\right)$ for providing θ, such that the following *coalition incentive compatibility constraint* is satisfied.

$$\underbrace{W_s\left(\theta\right)}_{\text{Wage Payment}} + \underbrace{s_s X\left(\theta\right)}_{\substack{\text{Behavioral Element}\\ \text{for supervisor}}} \geqslant \underbrace{kU(\theta)}_{\substack{\text{Information Rent}\\ \text{given for supervisor}}} = k\underbrace{\left[U\left(\underline{\theta}\right) - \int_{\underline{\theta}}^{\theta}\frac{\partial C\left(X\left(\tau\right),\tau\right)}{\partial\tau}d\tau\right]}_{U(\theta)}$$

Therefore, eliciting the supervisor's information requires the principal to pay $W_s\left(\theta\right) = kU\left(\theta\right) - s_s X\left(\theta\right), \forall\theta$ to him if the (hard) information of θ is provided. Thus, we see that **behavioral elements can reduce the monetary reward** for providing the information θ.

Substituting $W(\theta) = C\left(X\left(\theta\right),\theta\right) + U(\theta) - s_A X(\theta)$ and $W_s\left(\theta\right) = kU\left(\theta\right) - s_s X\left(\theta\right)$ into the principal's objective function, the virtual surplus for type θ in the Principal-Supervisor-Agent regime is

$$\begin{aligned}
&X(\theta) - C\left(X\left(\theta\right),\theta\right) - (1-p)\left[U(\theta) - s_A X(\theta)\right] - p\left[kU(\theta) - s_s X(\theta)\right]\\
&= \underbrace{X(\theta) - C\left(X(\left(\theta\right),\theta\right)}_{\text{Total Surplus}} - \left[(1-p) + pk\right]\underbrace{U(\theta)}_{\substack{\text{Information}\\ \text{Rent}}} + \underbrace{\left[(1-p)s_A + ps_s\right]X(\theta)}_{\substack{\text{Gain from reduction of Information Rent}\\ \text{by behavioral benefits}}}
\end{aligned}$$

Hence, the program of designing the optimal collusion-proof contract *with behavioral elements* can be rewritten as

$$\begin{aligned}
&\max_{X(\cdot)}\int_{\underline{\theta}}^{\overline{\theta}}\Bigg[\left[1 + (1-p)s_A + ps_s\right]X(\theta) - C\left(X\left(\theta\right),\theta\right)\\
&\quad - \left[(1-p) + pk\right]\left(U\left(\underline{\theta}\right) - \int_{\underline{\theta}}^{\theta}\frac{\partial C\left(X\left(\tau\right),\tau\right)}{\partial\tau}d\tau\right)\Bigg]f(\theta)d\theta - z\\
&\quad \text{s.t.} \qquad dX(\theta)/d\theta \geqslant 0 \quad (M) \quad \forall\theta
\end{aligned}$$

From the lemma 2, the program becomes

$$\begin{aligned}\max_{X(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} \Bigg[& [1+(1-p)s_A+ps_s]\, X(\theta) - C\left(X\left(\theta\right),\theta\right) \\ & + [(1-p)+pk]\, \frac{\partial C\left(X\left(\theta\right),\theta\right)}{\partial\theta}\frac{1-F(\theta)}{f(\theta)}\Bigg] f(\theta)d\theta - z \\ \text{s.t.} \qquad & dX(\theta)/d\theta \geqslant 0 \quad (M) \quad \forall\theta\end{aligned}$$

We ignore the Monotonicity Constraint (M) and solve the *relaxed program.* The principal maximizes the expected value of the *modified virtual surplus*, denoted by $J^B\left(X,\theta\right)$. This expected value is maximized by simultaneously maximizing the modified virtual surplus for (almost) every type θ, i.e.,

$$X^B\left(\theta\right) \in \arg\max_{X(\cdot)} [1+(1-p)s_A+ps_s]\, X\left(\theta\right) - C\left(X\left(\theta\right),\theta\right) + [(1-p)+pk]\left[\frac{1-F\left(\theta\right)}{f\left(\theta\right)}\right]\frac{\partial C\left(X\left(\theta\right),\theta\right)}{\partial\theta}$$

This defines the optimal output rule $X^B\left(\cdot\right)$ for the program.

The principal's program can then be rewritten as

$$\max_{X\in \mathrm{X}} J^B\left(X,\theta\right) = [1+(1-p)s_A+ps_s]\, X - C\left(X,\theta\right) + \frac{[(1-p)+pk]}{h(\theta)}\frac{\partial C(X,\theta)}{\partial\theta}$$

where $h(\theta)$ is the hazard rate.

We take the derivative:

$$\frac{\partial J^B(X,\theta)}{\partial X} = [1+(1-p)\, s_A+ps_s] - \frac{\partial C\left(X,\theta\right)}{\partial X} + \frac{[(1-p)+pk]}{h\left(\theta\right)}\frac{\partial^2 C\left(X,\theta\right)}{\partial X\partial\theta} \qquad (**)$$

Proposition 7 *The optimal solution $X^B\left(\theta\right)$ with behavioral elements is greater than the solution $X^S\left(\theta\right)$ in the principal-supervisor-agent framework with no behavioral elements, that is, $X^B\left(\theta\right) \geqslant X^S\left(\theta\right)$. The principal's profit also increases in equilibrium with behavioral elements.*

Proof: Since $\frac{\partial J^s(X,\theta)}{\partial X} = 1 - \frac{\partial C(X,\theta)}{\partial X} + \frac{[(1-p)+pk]}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta} = 0$ at $X = X^S\left(\theta\right)$, we see from the above $(**)$ that $\frac{\partial J^B(X,\theta)}{\partial X} = (1-p)s_A + ps_s \geqslant 0$ at $X = X^S\left(\theta\right)$. Therefore, $X^S\left(\theta\right)$ cannot be optimal when the behavioral elements are introduced, except for $s_A = s_s = 0$. A small increase in X would increase $J^B\left(X,\theta\right)$, and we have $X^B\left(\theta\right) \geqslant X^S\left(\theta\right)$. ■

Next, we examine the comparative statics on the optimal solution $X^B\left(\theta\right)$, and we have:

Corollary *The optimal solution $X^B\left(\theta\right)$ with behavioral elements is increasing in the parameter s_A and s_s, and decreasing in the parameter k, but ambiguous whether the optimal solution is increasing in p.*

Proof: From $(**)$, the derivative $J^B_X\left(X,\theta\right)$ is increasing in s_A and s_s (behavioral elements). That is, $J^B\left(X,\theta\right)$ has SCP in $(X; s_A, s_s)$. Similarly, the derivative $J^B_X\left(X,\theta\right)$ is *decreasing* in k.
But, it is ambiguous whether the derivative $J^B_X\left(X,\theta\right)$ is increasing in p as follows.

$$J^B_{Xp}\left(X,\theta\right) = \underbrace{[s_s - s_A]}_{\substack{\text{Net negative effect from} \\ \text{the disappearance of behavioral} \\ \text{benefit on the agent side}}} - \underbrace{\frac{(1-k)}{h(\theta)}\underbrace{\frac{\partial^2 C\left(X,\theta\right)}{\partial X\partial\theta}}_{<0}}_{\substack{\text{Marginal gain from} \\ \text{the reduction of information rent}}}$$

It depends on which term is dominant between the first and the second one. ■

7. Conclusion

Recently, auditing to meet the needs of corporate governance has rapidly been increasing in importance in Japan, as well as in the U.S. and Western countries. Given this trend, we were motivated to build a theoretical model to examine how supervision (auditing) could be utilized in order to enhance the effectiveness of corporate governance and to deter collusive supervision (auditing).

We introduced the outcomes of "Monotone Comparative Statics" à la Topkis (1978) and Edlin and Shannon (1998), and Milgrom and Segal (2002)'s generalized envelope theorem into a familiar screening (self selection) model with a continuum of types, and constructed a three-tier agency model with a mathematically tractable structure. This should be an advantage in modeling in comparison with the collusion literature e.g., Kofman and Lawarree (1993)'s auditing application of the three-tier agency model à la Tirole (1986, 1992). The basic trade-off involved in adding the auditor (supervisor) into the hierarchy is the benefit obtained by the discrete reduction in information rent and the improvement of marginal incentives (outputs) versus the resource cost of the auditor (supervisor). This bottom line was consistently preserved through the model.

Throughout the basic model of the paper we considered a situation where the principal can *commit* to a collusion-proof contract, that is, 'full commitment'. We used the revelation principle, solving programs in which the principal always prevents collusion between the auditor (supervisor) and the manager (agent). In the optimal contract, nobody colludes: this is called the collusion-proof principle. However, this does not imply an obvious inconsistency with reality, where collusive supervision (auditing) often makes headlines, as stated in the introduction. The revelation principle and the collusion-proof principle are solution techniques which facilitate characterization of the optimal contract.[6]

We then showed as an extension what happens when the principal cannot fully commit to the mechanism and the renegotiation is unavoidable. When the principal commits herself to the reward scheme for the supervisor, but does not commit to the one for the agent, she is tempted to modify the initial contract (or the outcome) unilaterally, using the information revealed by the supervisor. The situation is similar to the ratchet problem and the renegotiation problem caused by lack of the principal's commitment in the dynamics of incentive contracts, studied early by Laffont-Tirole (1988), and Dewatripont (1988) etc. If the agent anticipates such a modification, since he can benefit from a failure by the supervisor to report his type truthfully, he will offer the supervisor the transfer (side payment) equivalent to his information rent. Thus, the principal must pay the supervisor in opposition to the collusive offer by the agent. Thus, the principal can strictly improve his payoff ex post, but must bear the ex ante incentive cost. However, the comparison between the payoffs for the principal is ambiguous. It depends on the relative sizes of several terms between a 'No-Commitment' regime and a 'Collusion-proof, full commitment' one. We showed that the principal can do better in the equilibrium without her some commitment than if she can fully commit.

As another extension, we incorporated the behavioral elements à la Fahr and Schmidt (1999) into the model, and examined their effects on the optimal solution in the principal-supervisor-agent hidden information model with collusion. We assumed that the agent and the supervisor enjoy non-monetary and non-transferable "private benefits" or "psychological benefits" from output achievement, such as job satisfaction and pride/self esteem generated through more output. Then, we found that behavioral elements could reduce the monetary reward for inducing true information. Hence, the virtual surplus for each type is increased by the reduction of information rent (an incentive cost for inducing a truthful information revelation). Thus, the optimal solution with behavioral elements became greater than the one with no behavioral elements. This will have important implications for organizational design.

[6] Indeed, if we consider an *incomplete* grand contract situation like Tirole (1992), Laffont and Tirole (1991), and Suzuki (2007), *equilibrium collusion* can improve efficiency. Such models indeed could be usefully applied, in such fields as political economy, regulation, and authority delegation in organizations.

Hence, we can say that the overall contribution of our paper is to apply the monotone comparative statics method to the three-tier agency model with hidden information and collusion, thereby providing a framework that can address the issues treated in the existing literature *in a much simpler fashion*. By using its tractable framework, we examined some interesting extensions, such as the effect of introducing another supervisor, a problem resulting by lack of a principal's commitment, and the effect of incorporating behavioral elements. Finally, we derived some clear and robust implications applicable to corporate governance reform and organizational design.

References

1. Bolton, P and M. Dewatripont (2005) *Contract Theory* MIT Press
2. Dewatripont, M (1989) "Renegotiation and Information Revelation over Time: The Case of Optimal Labor Contracts," *Quarterly Journal of Economics*, Vol.104. No3, August, 589-619.
3. Fehr, E. and K. Schmidt (1999) "A Theory of Fairness, Competition and Cooperation" *Quarterly Journal of Economics*, Vol.114, No3, 817-868.
4. Edlin, A and Shannon, C (1998) "Strict Monotonicity in Comparative Statics." *Journal of Economic Theory*, 81, July, 201-219.
5. Kofman, F, and Lawarree, J, (1993) "Collusion in Hierarchical Agency", *Econometrica*, Vo1.61, No3, May, 629-656.
6. Laffont, J and D. Martimort (1997) "Collusion under Asymmetric Information" *Econometrica*, Vol.65, No4, 875-911.
7. Laffont, J-J. and J. Tirole (1988) "The Dynamics of Incentive Contracts," *Econometrica*, Vol. 56 No.5, 1153-75, September
8. Laffont, J. and J. Tirole (1991) "The Politics of Government Decision-Making: A Theory of Regulatory Capture," *Quarterly Journal of Economics*. 106, 1089-1127.
9. Milgrom, P, and Segal, I (2002) "Envelope Theorems for Arbitrary Choice Sets," *Econometrica* 70(2), March, 583-601
10. Mirrlees, J.A. (1971) "An Exploration in the Theory of Optimum Income Taxation" *Review of Economic Studies*, 38, 175-208.
11. Suzuki, Y., (2007) "Collusion in Organizations and Management of Conflicts through Job Design and Authority Delegation", *Journal of Economic Research* 12. pp. 203-241
12. Suzuki, Y., (2008) "Mechanism Design with Collusive Supervision: A Three-tier Agency Model with a Continuum of Types," *Economics Bulletin*, Vol. 4 no. 12 pp. 1-10
13. Tirole, J. (1986) "Hierarchies and Bureaucracies: On the role of Collusion in Organizations". *Journal of Law, Economics and Organization*. 2. 181-214.
14. Tirole, J. (1992) "Collusion and the Theory of Organizations" in *Advances in Economic Theory: The Sixth World Congress* edited by J.J.Laffont. Cambridge: Cambridge University Press.
15. Topkis, D (1978) "Minimizing a submodular function on a lattice", *Operations Research*, 26(2), 255-321.

APPENDICES

Appendix1 Proof of Lemma2

Proof: The "$\Rightarrow$" part was established above. It remains to show that local IC and monotonicity imply that $U\left(\hat{\theta},\theta\right) \leqslant U(\theta)$ for all $\hat{\theta},\theta$. For $\hat{\theta} > \theta$, we can write

$$\begin{aligned}
U\left(\hat{\theta},\theta\right) - U(\theta) =& W(\hat{\theta}) - C\left(X\left(\hat{\theta}\right),\theta\right) - U(\theta) \\
=& U\left(\hat{\theta}\right) + C\left(X\left(\hat{\theta}\right),\hat{\theta}\right) - C\left(X\left(\hat{\theta}\right),\theta\right) - U(\theta) \\
=& \left[C\left(X\left(\hat{\theta}\right),\hat{\theta}\right) - C\left(X\left(\hat{\theta}\right),\theta\right)\right] + \left[U\left(\hat{\theta}\right) - U(\theta)\right] \\
=& \int_{\theta}^{\hat{\theta}} \frac{\partial C\left(X\left(\hat{\theta}\right),\tau\right)}{\partial \tau} d\tau + \int_{\theta}^{\hat{\theta}} \left[-\frac{\partial C(X(\tau),\tau)}{\partial \tau}\right] d\tau \\
=& \int_{\theta}^{\hat{\theta}} \left[\frac{\partial C\left(X\left(\hat{\theta}\right),\tau\right)}{\partial \tau} - \frac{\partial C(X(\tau),\tau)}{\partial \tau}\right] d\tau \leqslant 0
\end{aligned}$$

Here the last equality obtains by **(ICFOC)**[7], and the inequality obtains by SCP and the fact that $X\left(\hat{\theta}\right) \geqslant X(\tau)$ by **(M)**. The proof for $\theta > \hat{\theta}$ is similar. ■

Appendix2 Proof of Lemma3

Proof: We transform the *expected information rents* by exploiting "Integration by Parts".

Now, remember that

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[U(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} \frac{\partial C(X(\tau),\tau)}{\partial \tau} d\tau\right] f(\theta)\, d\theta = \int_{\underline{\theta}}^{\overline{\theta}} U(\theta) f(\theta)\, d\theta$$

Because $[U(\theta) F(\theta)]' = U(\theta) f(\theta) + \underbrace{\frac{dU(\theta)}{d\theta}} F(\theta) = U(\theta) f(\theta) \underbrace{- \frac{\partial C(X(\theta),\theta)}{\partial \theta}}_{\text{(Due to Envelope Theorem)}} F(\theta)$, and so $U(\theta) f(\theta) = [U(\theta) F(\theta)]' + \frac{\partial C(X(\theta),\theta)}{\partial \theta} F(\theta)$, we have

$$\begin{aligned}
\int_{\underline{\theta}}^{\overline{\theta}} U(\theta) f(\theta)\, d\theta =& [U(\theta) F(\theta)]_{\underline{\theta}}^{\overline{\theta}} + \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta} F(\theta) d\theta \\
=& U\left(\overline{\theta}\right) + \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta} F(\theta) d\theta \\
=& U(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta} d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta} F(\theta) d\theta \\
& \left(\because U\left(\overline{\theta}\right) = U(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta} d\theta\right) \\
=& U(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta} (1 - F(\theta))\, d\theta \\
=& U(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta} \frac{1 - F(\theta)}{f(\theta)} f(\theta)\, d\theta \quad ■
\end{aligned}$$

[7] $U\left(\hat{\theta}\right) - U(\theta) = \int_{\theta}^{\hat{\theta}} \frac{dU}{d\tau}(\tau)\, d\tau \underbrace{=}_{\because \text{Envelope Theorem}} \int_{\theta}^{\hat{\theta}} -\frac{\partial C(X(\tau),\tau)}{\partial \tau} d\tau$