E C O N O M I C S B U L L E T I N

Mechanism design with collusive supervision: a three-tier agency model with a continuum of types

Yutaka Suzuki
Faculty of Economics, Hosei University

Abstract

We apply the "Monotone Comparative Statics" method ala 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 ala Tirole (1986, 1992), thereby provide 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. In addition to such a technical contribution, the paper derives some clear and robust implication applicable to corporate governance reform (Propositions 1 (2) and 3).

The author is grateful to an associate editor for his helpful comments and suggestions. This research was financially supported by the Research Project at Hosei Institute on Aging, Hosei University supported by the Ministry of Education and the Grant from the Zengin Foundation for Studies on Economics and Finance in Japan.
Citation: Suzuki, Yutaka, (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
Submitted: May 16, 2008. **Accepted:** May 16, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume4/EB-08D80022A.pdf

1 Introduction

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 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.

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.

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

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 $\max_{X \in \mathrm{X}} U(X,\theta) := W(X) - C(X,\theta)$ 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 $\theta'' > \theta'$, $X' \in \arg\max_{X \in \mathrm{X}} U(X,\theta')$, and $X'' \in \arg\max_{X \in \mathrm{X}} U(X,\theta'')$. 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: $\max_{X \in \mathrm{X}} X - C(X,\theta)$ 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 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] Edlin and Shannon (1998) introduced this SCP under the name of "increasing marginal returns".

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:

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

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

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. 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 $k \in [0, 1]$. 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.

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

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 $X < X^S(\overline{\theta}) = X^{FB}(\overline{\theta})$. This result is a generalization of the two-type case. 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:

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

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 $k \in [0,1]$. 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. ■

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 Conclusion

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. Hence, we can say that the overall contribution of our note is to apply the monotone comparative statics method to the three-tier agency model with hidden information and collusion, to provide a framework that can address the issues treated in the existing literature *in a much simpler fashion*, and to derive some clear and robust implication applicable to corporate governance reform.

References

1. Bolton, P and M. Dewatripont (2005) *Contract Theory* MIT Press
2. Edlin, A and Shannon, C (1998) "Strict Monotonicity in Comparative Statics." *Journal of Economic Theory*, 81, July, 201-219.
3. Kofman, F, and Lawarree, J, (1993) "Collusion in Hierarchical Agency", *Econometrica*, Vo1.61, No3, May, 629-656.
4. Laffont, J and D. Martimort (1997) "Collusion under Asymmetric Information" *Econometrica*, Vol.65, No4, 875-911.
5. Laffont, J. and J. Tirole (1991) "The Politics of Government Decision-Making: A Theory of Regulatory Capture," *Quarterly Journal of Economics*.106, 1089-1127.
6. Milgrom, P, and Segal, I (2002) "Envelope Theorems for Arbitrary Choice Sets," *Econometrica* 70(2), March, 583-601
7. Mirrlees, J.A. (1971) "An Exploration in the Theory of Optimum Income Taxation" *Review of Economic Studies*, 38, 175-208.
8. Suzuki, Y., (1999) Incomplete Contracts, Collusion and Authority Delegation in Hierarchical Agency. Chapter 1 in his Ph.D. Thesis at The University of Tokyo. February, 1999.
9. Tirole, J. (1986) "Hierarchies and Bureaucracies: On the role of Collusion in Organizations". *Journal of Law, Economics and Organization*. 2. 181-214.
10. 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.
11. Topkis, D (1978) "Minimizing a submodular function on a lattice", *Operations Research*, 26(2), 255-321.

APPENDICES

Appendix1 Proof of Lemma2

Proof: The "⇒" 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

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

Here the last equality obtains by **(ICFOC)**[3], 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

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

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

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

---

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