法政大学学術機関リポジトリ
HOSEI UNIVERSITY REPOSITORY

PDF issue: 2024-07-27

A Three-Tier Agency Model with Collusive Auditing : Two-Type Case

SUZUKI, Yutaka

(出版者 / Publisher)
法政大学比較経済研究所 / Institute of Comparative Economic Studies, Hosei University

(雑誌名 / Journal or Publication Title)
Journal of International Economic Studies

(巻 / Volume)
25

(開始ページ / Start Page)
29

(終了ページ / End Page)
48

(発行年 / Year)
2011-03

(URL)
https://doi.org/10.15002/00007859

Journal of International Economic Studies (2011), No.25, 29–48
©2011 The Institute of Comparative Economic Studies, Hosei University

A Three-Tier Agency Model with Collusive Auditing: Two-Type Case

Yutaka Suzuki
Faculty of Economics, Hosei University

ABSTRACT

We construct a simple three-tier agency model, which is an extension of the familiar screening (self selection) models and can be placed in line with the collusion literature à la Tirole (1986, 1992), including Kofman and Lawarree (1993)'s auditing application. The basic trade-off consists of the discrete reduction in information rent vs. the improvement of marginal incentives (outputs), and thus we have "Efficiency at the top" and "Downward distortion at the bottom" at the optimum. Then, we theoretically compare the "collusion-proof" regime and the "no-commitment/renegotiation" regime. We extensively utilize a graphical explanation, which provides us with clear intuition and logic of the optimal solutions and their comparative statics, as well as robust implications for corporate governance reform. Last, as an extension, we show that when the private information of the two agents is *perfectly correlated*, the principal can implement the full information first best optimum *at no incentive cost*.

Key words: Mechanism Design, Adverse Selection, Collusion, Supervision, Renegotiation, Yardstick Mechanism, Corporate Governance

JEL Classification: D82, D86

1. Introduction

Recently, auditing has been rapidly growing 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 effective corporate governance. In those cases, collusive supervision/auditing and revelation of false information were common occurrences. 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 Japanese firms' top management organization, a shareholders' meeting elects a director (or the board of directors) and an auditor who audits the execution of the work of the management and makes a report to the shareholders. With this auditing system, which has been legally amended several times, it is often said that the auditor has 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 auditing information in order to increase the interests of the shareholders through an arrangement in which the auditor and manager do not collude. Many Japanese firms, such as Toyota and Canon, do preserve and try to improve the traditional Japanese auditing system. Our paper can be viewed as an analysis of such 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]

In this paper, we construct a simple three-tier agency model in line with the above collusion models. Since it is an extension of the familiar screening (self-selection) models à la Baron and Myerson (1982) and Maskin and Riley (1984), the basic trade-off in the model consists of the discrete reduction in information rent vs. the improvement of marginal incentives (outputs), and thus we preserve "efficiency at the top" and "downward distortion at the bottom" at the optimum.

Then, the optimal collusion-proof contract in the principal-supervisor-agent three-tier regime built by adding the auditor (supervisor) into the hierarchy has the property whereby (1) *efficiency at the top* (for the good-type agent) and (2) *downward distortion at the bottom* (for the bad-type agent), and the downward distortion is *mitigated* at the optimum, in comparison with the principal-agent two-tier regime. The optimal solution allows 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—i.e., 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).

Though we basically consider a situation where the principal can *commit* to a collusion-proof contract, i.e., a situation of "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 and 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. Hence, the principal can strictly improve his payoff ex-post, but must bear the ex-ante incentive cost. In this situation, it would

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

be interesting to see whether the principal can do better in the equilibrium without her partial commitment than if she can fully commit.

Throughout the paper, we extensively utilize a graphical explanation, which is indeed beneficial in that we can provide clear intuition and logic of the optimal solutions of the model and their comparative statics, as well as robust implications for reform of corporate governance.

Incidentally, the principal cannot attain the first-best under the principal-supervisor-one agent hidden information setting with collusion, and downward distortion at the bottom always occurs. Introduction of the supervisor can mitigate, but cannot solve the problem perfectly, and indeed in some case the distortion may become bigger (as in the no-commitment/renegotiation regime). So, finally, we show that in the case where the private information of the two agents is *perfectly correlated*, the principal can implement the full information first best optimum *at no incentive cost*. Intuitively, the principal places the two agents in a *prisoner's dilemma* game and both truth-telling and the first-best optimal incentives can be induced in the unique dominant strategy equilibrium, without giving any information rent.

2. Principal-Agent Hidden Information 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. Gross profits are $X = \theta + e$, where θ is the manager's ability to run the firm and e is the effort he supplies. θ a priori belongs to $\{\underline{\theta}, \overline{\theta}\}$ and the prior beliefs are $\Pr\left(\theta = \overline{\theta}\right) = h$, where $0 < h < 1$. Expending effort e costs the manager $C(e)$ in disutility, which satisfies $C(e) > 0, C'(e) > 0, C'' > 0, \forall e \in \mathrm{R}_{+}$. W is the wage payment the agent receives, and then his utility is $W - C(e)$. 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 form of the contract is $\{X, W\}$. X is the level of gross profits[2] the manager is required to obtain and W the wage he will be paid if he generates the required level. If he produces less than X, he receives no pay. If he generates more than X, he will still receive only W. If he rejects the offer, the game ends. If he accepts the offer, a contract is signed and the principal is fully committed. This is a standard *screening problem*.

Now, we examine the optimal solution. Let $X\left(\overline{\theta}\right)$ and $X\left(\underline{\theta}\right)$ be the profits specified for the good-type agent ($\theta = \overline{\theta}$) and the bad-type agent ($\theta = \underline{\theta}$), respectively. We write X_H and X_L for $X\left(\overline{\theta}\right)$ and $X\left(\underline{\theta}\right)$, respectively. Defining $W\left(\overline{\theta}\right)$ and $W\left(\underline{\theta}\right)$ similarly, we write W_H and W_L for $W\left(\overline{\theta}\right)$ and $W\left(\underline{\theta}\right)$, respectively. These are the wages specified by the contracts.

The benchmark first best solution maximizes the expected profits, subject to the IR (Individual Rationality) constraints, which require that the manager be willing to sign a contract whatever her type. That is,

$$\max_{\{X_H, W_H\},\{X_L, W_L\}} h[X_H - W_H] + (1-h)[X_L - W_L]$$
$$\text{s.t.} \quad W_H - C(X_H - \bar{\theta}) \geq 0$$
$$W_L - C(X_L - \underline{\theta}) \geq 0$$

Substituting $W_H = C(X_H - \bar{\theta})$ and $W_L = C(X_L - \underline{\theta})$ into the objective function results in the expected total surplus maximization:

$$\max_{\{X_H, X_L\}} h\left[X_H - C(X_H - \bar{\theta})\right] + (1-h)\left[X_L - C(X_L - \underline{\theta})\right]$$

The first order conditions for the optimum are:

$$\left.\begin{aligned} 1 - \frac{\partial C(X_H - \bar{\theta})}{\partial X_H} = 0 \\ 1 - \frac{\partial C(X_L - \underline{\theta})}{\partial X_L} = 0 \end{aligned}\right\} \Leftrightarrow 1 = \frac{\partial C(X_H - \bar{\theta})}{\partial X_H} = \frac{\partial C(X_L - \underline{\theta})}{\partial X_L}$$

In the first best optimum, the marginal benefit of output 1 is equal to the marginal cost of output for both types $\bar{\theta}, \underline{\theta}$. Hence, we have $X_H - \bar{\theta} = X_L - \underline{\theta}$ and $e_H^{FB} = e_L^{FB} = e^{FB}$. This means that first best efforts are equal for both types $\bar{\theta}, \underline{\theta}$. We also see that $W_H^{FB} = W_L^{FB} = C(e^{FB})$

Next, under the assumption of asymmetric information on θ, we seek the separating contracts, which induce the two types to behave differently. For this, the contracts must be incentive compatible.

IC (Incentive Compatibility) requires:

$$W_H - C(X_H - \bar{\theta}) \geq W_L - C(X_L - \bar{\theta}) \tag{1a}$$

$$W_L - C(X_L - \underline{\theta}) \geq W_H - C(X_H - \underline{\theta}) \tag{1b}$$

(1a) states that the good-type agent $(\theta = \bar{\theta})$ prefers to select the contract intended for him rather than the contract intended for the bad-type agent $(\theta = \underline{\theta})$, i.e., the good-type agent's IC constraint. (1b) states that the bad-type agent $(\theta = \underline{\theta})$ prefers to select the contract intended for him rather than the contract intended for the good-type agent $(\theta = \bar{\theta})$, i.e., the bad-type agent's IC constraint.

The IR (Individual Rationality) constraints require:

$$W_H - C(X_H - \bar{\theta}) \geq 0 \tag{2a}$$

$$W_L - C(X_L - \underline{\theta}) \geq 0 \tag{2b}$$

The first best solutions $\{X_H^{FB}, X_L^{FB}\} = \{\bar{\theta} + e^{FB}, \underline{\theta} + e^{FB}\}, W_H^{FB} = W_L^{FB} = C(e^{FB})$ are *not* incen-

[2] We will later rephrase (gross) profit *X* as output *X* in order to make the terminology clearer in the theoretical analysis.

tive compatible for the good-type agent $\overline{\theta}$, since he has an incentive to tell a lie (mimic/pretend that type $\theta = \underline{\theta}$). Indeed, we can check the incentive of the good type $\overline{\theta}$.
If he tells the truth "$\theta = \overline{\theta}$", he obtains $W_H^{FB} - C\left(X_H^{FB} - \overline{\theta}\right) = 0$.
If he says "$\theta = \underline{\theta}$" (i.e., he lies), he obtains

$$W_L^{FB} - C\left(X_L^{FB} - \overline{\theta}\right) = C\left(e^{FB}\right) - C\left(e^{FB} - \left(\overline{\theta} - \underline{\theta}\right)\right) > 0.$$

Hence, he has an incentive to tell a lie (mimic/pretend), i.e., *not incentive compatible*.

As is typical in such problems, only the good type's IC (1a) and the bad type's IR (2b) bind at the optimum. From (2b), $W_L = C\left(X_L - \underline{\theta}\right)$. Substituting it into (1a) with equality, we have

$$W_H - C\left(X_H - \overline{\theta}\right) = W_L - C\left(X_L - \overline{\theta}\right) = C\left(X_L - \underline{\theta}\right) - C\left(X_L - \overline{\theta}\right) \tag{3}$$

This is the *information rent* for the good-type agent $\overline{\theta}$. Hence, the optimization problem can be written as follows

$$\max_{\{X_H, W_H\},\{X_L, W_L\}} h\left[X_H - W_H\right] + (1-h)\left[X_L - W_L\right]$$
$$\text{s.t.} \quad W_H - C\left(X_H - \overline{\theta}\right) = C\left(X_L - \underline{\theta}\right) - C\left(X_L - \overline{\theta}\right)$$
$$W_L = C\left(X_L - \underline{\theta}\right)$$

Substituting $W_L = C\left(X_L - \underline{\theta}\right)$ and $W_H = C\left(X_H - \overline{\theta}\right) + \left[C\left(X_L - \underline{\theta}\right) - C\left(X_L - \overline{\theta}\right)\right]$ into the objective function yields

$$\max_{X_H, X_L} \underbrace{h\left[X_H - C\left(X_H - \overline{\theta}\right)\right] + (1-h)\left[X_L - C\left(X_L - \underline{\theta}\right)\right]}_{\text{Expected Total Surplus}} - \underbrace{h\left[C\left(X_L - \underline{\theta}\right) - C\left(X_L - \overline{\theta}\right)\right]}_{\substack{\text{"Information Rent"} \\ \text{for the good type}}}$$

The first order conditions for the optimum are:

$$1 - \frac{\partial C\left(X_H - \overline{\theta}\right)}{\partial X_H} = 0 \Leftrightarrow X_H^* = X_H^{FB}$$

$$\underbrace{(1-h)\left[1 - \frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L}\right]}_{\substack{\text{Marginal Surplus} \\ \text{for the bad type}}} - \underbrace{h\left[\frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L} - \frac{\partial C\left(X_L - \overline{\theta}\right)}{\partial X_L}\right]}_{\substack{\text{Marginal Information Rent} \\ \text{for the good type}}} = 0$$

$$\Leftrightarrow 1 - \frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L} - \frac{h}{1-h} \cdot \left[\frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L} - \frac{\partial C\left(X_L - \overline{\theta}\right)}{\partial X_L}\right] = 0$$

From these conditions, we have the following proposition, which is a familiar result in the literature (e.g. Baron and Myerson (1982), Maskin and Riley (1984), and Bolton and Whinston (2005))

Proposition 0

In the principal-agent regime with no supervisor, the second-best solution has the properties of
(1)*Efficiency at the to*p (for the good-type agent) $X_H^* = X_H^{FB}$
(2)*Downward distortion at the bottom* (for the bad-type agent) $X_L^* < X_L^{FB}$

Proof: As for X_H, the first order condition is the same as the first best case, so $X_H^* = X_H^{FB}$.
As for X_L, evaluating the first order condition at $X = X_L^{FB}$, we have

$-\frac{h}{1-h}\left[\frac{\partial C\left(X_L^{FB}-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L^{FB}-\overline{\theta}\right)}{\partial X_L}\right]<0$. This means that the principal can raise his virtual payoff by decreasing X_L from the first best level X_L^{FB}. Hence, we have $X_L^* < X_L^{FB}$.■

Graphical Explanation

Let us explain the argument so far in a graphical manner. First, the payoff function of the type θ agent is $U(W,X;\theta)=W-C(X-\theta)$. In order to depict the indifference curve of the type θ agent in the (X,W) diagram, we totally differentiate both sides of $U_\theta = W - C(X-\theta)$, and obtain $dW - \frac{\partial C(X-\theta)}{\partial X}dX = 0$. Then, putting it in order, we have the marginal rate of substitution $MRS_{XW}^{\theta} = \left.\frac{dW}{dX}\right|_{U=\text{const}} = \frac{\partial C(X-\theta)}{\partial X}$. We easily see that the marginal cost of output $\frac{\partial C(X-\theta)}{\partial X}$ is decreasing in type θ , i.e., the good type $\overline{\theta}$ has a gentler indifference curve (a smaller MRS_{XW}) for any point (X,W). Remember that the first order conditions for the first best optimum are $1=\frac{\partial C\left(X_H-\overline{\theta}\right)}{\partial X_H}=\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}$. We have $X_H^{FB}-\overline{\theta} = X_L^{FB}-\underline{\theta}$, which means that $e_H^{FB} = e_L^{FB} = e^{FB}$. That is, at the first best solution, $1=C'\left(X_L^{FB}-\underline{\theta}\right)=C'\left(X_H^{FB}-\overline{\theta}\right)=C'\left(e^{FB}\right)$ and $W_H^{FB}=W_L^{FB}=C\left(e^{FB}\right)$. From these facts, we can depict the indifference curves of both types and the first best contracts G and B in the (X,W) diagram.

However, the first best solution $G:\left\{X_H^{FB},W_H^{FB}\right\}=\left\{\overline{\theta}+e^{FB},C\left(e^{FB}\right)\right\}$ is *not* incentive compatible for the good type $\overline{\theta}$ under asymmetric information, since he has an incentive to tell a lie (mimic type $\underline{\theta}$) and select $B:\left\{X_L^{FB},W_L^{FB}\right\}=\left\{\underline{\theta}+e^{FB},C\left(e^{FB}\right)\right\}$. Indeed, the good-type agent $\overline{\theta}$ can obtain $W_L^{FB}-C\left(X_L^{FB}-\overline{\theta}\right)=C\left(e^{FB}\right)-C\left(e^{FB}-\left(\overline{\theta}-\underline{\theta}\right)\right)>0$ by telling a lie, instead of $W_H^{FB}-C\left(X_H^{FB}-\overline{\theta}\right)=0$ by telling the truth. Below Figure 2 depicts this fact.

So, the principal takes the optimal balance between the expected total surplus and the information rent for the good type. As a result, we have the results of (1) *Efficiency at the top* (for the

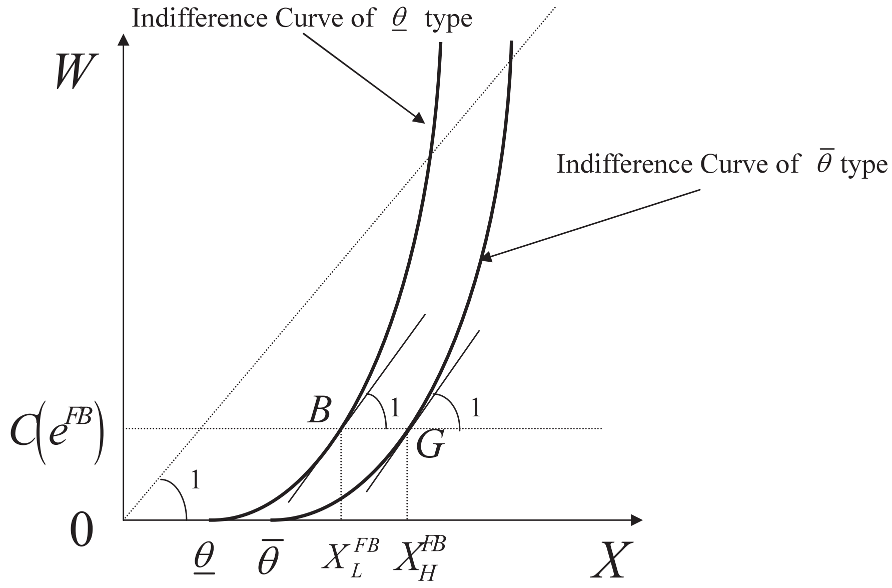

Figure 1

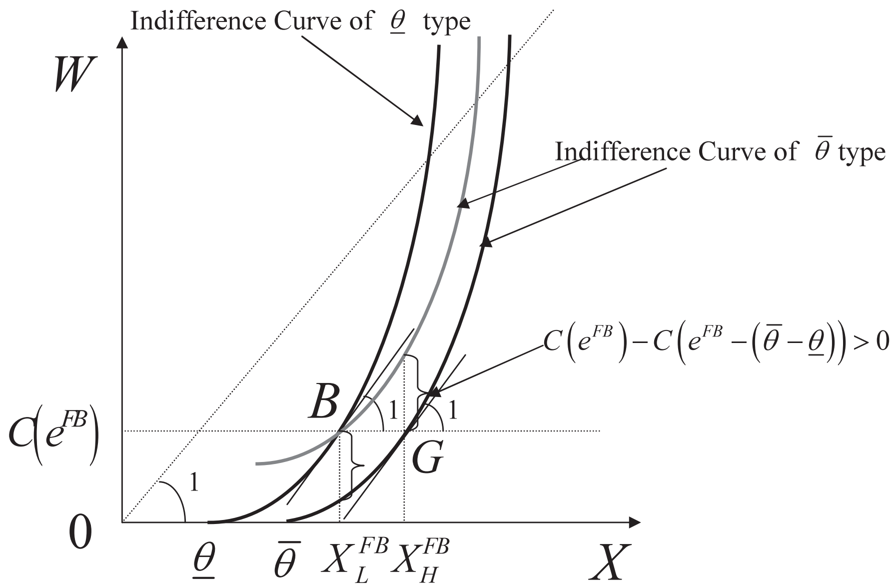

Figure 2

good-type agent $\overline{\theta}$) $X_H^* = X_H^{FB}$ and (2) *Downward distortion at the bottom* (for the bad-type agent $\underline{\theta}$) $X_L^* < X_L^{FB}$. The intuition is that a small reduction in X_L from the first best X_L^{FB} results in a *second-order (marginal) reduction* in total surplus for the bad type $\underline{\theta}$, but generates a *first-order (discrete) reduction* in the good type $\overline{\theta}$'s information rent through relaxing the IC for the good type $\overline{\theta}$ and allowing the principal to reduce *W* discretely. The optimal wage payments are

$W_L^* = C\left(X_L^* - \underline{\theta}\right) = C\left(e_L^*\right)$ for the bad type $\underline{\theta}$, and

$W_H^* = \underbrace{C\left(X_H^{FB} - \overline{\theta}\right)}_{\text{effort cost}} + \underbrace{C\left(X_L^* - \underline{\theta}\right) - C\left(X_L^* - \overline{\theta}\right)}_{\text{information rent}}$ for the good type $\overline{\theta}$.

The Figure3 shows the result.

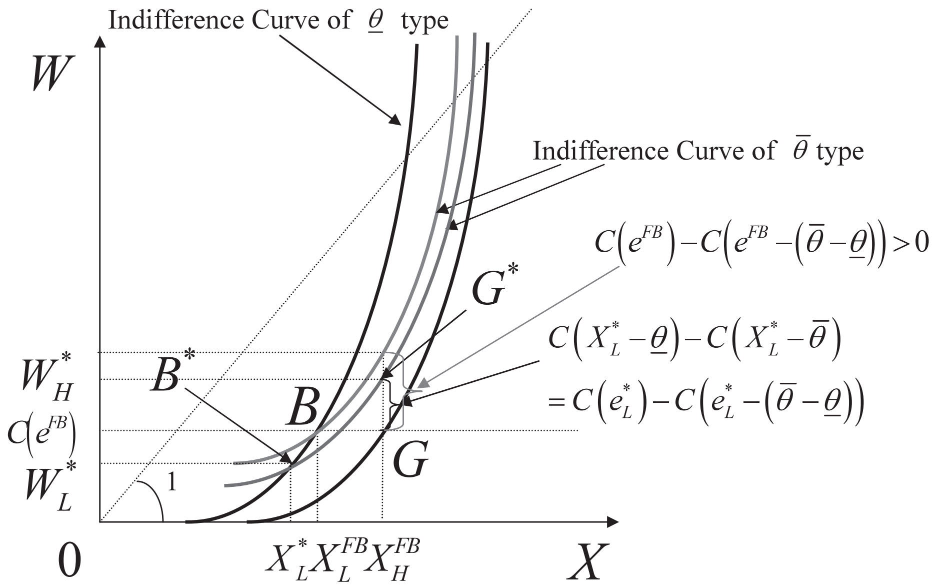

Figure 3

The result of $X_L^* < X_L^{FB}$ can be understood by looking at Figure 4, which shows that the optimal solution X_L^* is determined such that the marginal benefit 1 equals the marginal *virtual cost* (the marginal cost $\frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L}$ plus the marginal information rent $\frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L} - \frac{\partial C\left(X_L - \overline{\theta}\right)}{\partial X_L}$).

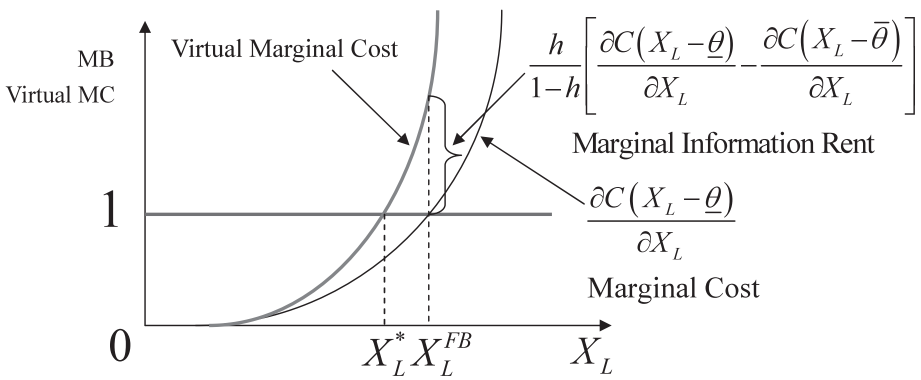

Figure 4

3. Collusion and Supervision

Now, we introduce a third player, the supervisor, into the model. The principal has access, at a cost z, to the supervisor who is an internal auditor and can, for each θ, provide proof of the fact (θ) with probability p, and with $1-p$, is unable to obtain any information.[3] We assume that proofs of θ cannot be falsified, and thus the agent is protected against false claims that his type θ is higher/lower than what it really is. On the other hand, the agent can potentially benefit from a failure by the supervisor to truthfully report that his type is $\theta=\bar{\theta}$, when the supervisor observes the signal θ. A self-interested supervisor will collude with the agent only if he benefits from such behavior. Specifically, let us 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, being prevented by the principal, may be hard to organize and are subject to resource losses. We follow the literature in assuming that side-contracts of this sort are possible (see, e.g., Tirole 1992).

To avoid collusion, the principal will have to offer the supervisor a reward W_S for providing the information $\theta=\bar{\theta}$, such that the following *coalition incentive compatibility constraint* is satisfied.

$$W_s \geq kU_H = k\left[C\left(X_L-\underline{\theta}\right)-C\left(X_L-\bar{\theta}\right)\right]$$

Indeed, once the information $\theta=\bar{\theta}$ is obtained, the principal will drop agent $\bar{\theta}$'s payment W_H to $C\left(X_H-\bar{\theta}\right)$, and not pay the information rent to agent $\bar{\theta}$. The agent is thus ready to pay the supervisor an amount of $U_H=C\left(X_L-\underline{\theta}\right)-C\left(X_L-\bar{\theta}\right)$, and the value of this side payment to the supervisor is kU_H, where $k\in[0,1]$. Therefore, hiring a supervisor and eliciting his information requires the principal to pay kU_H to the supervisor if the (hard) information of $\theta=\bar{\theta}$ is provided. Substituting $W_S = kU_H$ into the principal's objective function, the <u>virtual surplus</u> for the bad type $\theta=\underline{\theta}$ in the principal-supervisor-agent regime is,

$$\underbrace{(1-h)\left[X_L-C\left(X_L-\underline{\theta}\right)\right]}_{\text{Total Surplus for the bad type}}-h\cdot\left[\underbrace{(1-p)U_H}_{\substack{\text{Information Rent}\\ \text{for the good type}}}+\underbrace{pkU_H}_{\substack{\text{Information Rent}\\ \text{for the supervisor}}}\right]$$

$$=\underbrace{(1-h)\left[X_L-C\left(X_L-\underline{\theta}\right)\right]}_{\text{Total Surplus for the bad type}}-h\left[(1-p)+pk\right]\underbrace{\left[C\left(X_L-\underline{\theta}\right)-C\left(X_L-\bar{\theta}\right)\right]}_{\text{Information Rent}}$$

Hence, the expected total virtual surplus is:

$$\underbrace{h\left[X_H-C\left(X_H-\bar{\theta}\right)\right]+(1-h)\left[X_L-C\left(X_L-\underline{\theta}\right)\right]}_{\text{Expected Total Surplus}}-h\left[(1-p)+pk\right]\underbrace{\left[C\left(X_L-\underline{\theta}\right)-C\left(X_L-\bar{\theta}\right)\right]}_{\text{Information Rent}}$$

The first order conditions for the optimum are:

---

[3] The supervisor's signal received from the agent may be informative $s=\theta$ with probability p, or non-informative $s=\phi$ with probability $1-p$.

$$1-\frac{\partial C\left(X_H-\bar{\theta}\right)}{\partial X_H}=0 \Leftrightarrow X_H^S=X_H^{FB}$$

$$\underbrace{1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}}_{\text{Marginal Total Surplus}}-\frac{h}{1-h}\underbrace{\left[\left(1-p\right)+pk\right]}_{\le 1}\underbrace{\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\bar{\theta}\right)}{\partial X_L}\right]}_{\text{Marginal Information Rent}}=0$$

Proposition 1:

In the principal-supervisor-agent regime, the optimal collusion-proof solution has the following properties:

(1) *Efficiency at the top* (for the good-type agent) $X_H^S=X_H^{FB}$

(2) *Downward distortion at the bottom* (for the bad-type agent) is mitigated, i.e.,

$$X_L^* \underbrace{\le}_{\substack{\text{Equality}\\ \text{holds at } k=1}} X_L^S < X_L^{FB}.$$

Result (2) comes from the reduction in the virtual cost, i.e., the total and marginal information rents by the introduction of a supervisor with $k \le 1$. The point is the reduction in the virtual marginal cost due to $\left(1-p\right)+pk \le 1$, compared with the standard no-supervisor case. The following figure clearly shows this point.

$$\frac{h}{1-h}\underbrace{\left[\left(1-p\right)+pk\right]}_{\le 1}\underbrace{\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\bar{\theta}\right)}{\partial X_L}\right]}_{\text{Marginal Information Rent}}$$

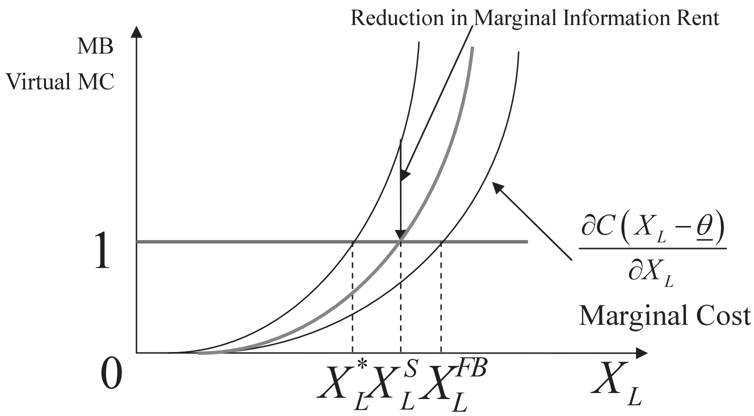

Figure 5

We can readily perform a comparative statics, and the rationale of the results is clear and intuitive.

Proposition2: Comparative statics on X_L^S

The optimal collusion-proof solution X_L^S is non-decreasing in the parameter *p*, and non-increasing in the parameter *k*.

Proof:

The coefficient of the marginal information rent $(1-p)+pk=1-p(1-k)$ decreases as the parameter *p* increases. Hence, the marginal information rent (and so the marginal virtual cost) $\frac{h}{1-h}\left[(1-p)+pk\right]\left[\frac{\partial C(X_L-\underline{\theta})}{\partial X_L}-\frac{\partial C(X_L-\overline{\theta})}{\partial X_L}\right]$ decreases as *p* increases. This brings about the increase in the optimal output $X_L^S\uparrow$. Similarly, the coefficient of the marginal information rent $(1-p)+pk=1-p(1-k)$ increases as the parameter *k* increases. Hence, the marginal information rent (and so the marginal virtual cost) increases as *k* increases. This brings out the decrease in the optimal output $X_L^S\downarrow$. ■

4. No-Commitment /Renegotiation Regime

So far, we have considered a situation where the principal can *commit* to the collusion-proof contract. Now, we examine more explicitly the timing of the game. The principal has access to the supervisor, who chooses a message $\mathrm{m}\in\{\varnothing,\theta\}$, where $\varnothing$ 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 *renegotiate* the original contract. The principal can raise her payoff by *eliminating the downward distortion in the bad type* $\underline{\theta}$.[4] Namely, instead of the contract at the no-information phase $\{X_L^{NC},W_L^{NC}\}$, the principal will offer the efficient (first best) contract $\{X_L^{FB},W_L^{FB}\}$, and exploit the information rent from the good type $U(\overline{\theta})=C(X_L^{NC}-\underline{\theta})-C(X_L^{NC}-\overline{\theta})$. If the good-type agent $\overline{\theta}$ anticipates this modification, since he can benefit from a failure by the supervisor to report his type $\overline{\theta}$ truthfully, he will offer the supervisor the side payment $U(\overline{\theta})$, for which the supervisor benefits up to *kt*, where $k\in[0,1]$. Thus, the principal must pay $W_s(\theta)=kU(\overline{\theta})=k\left[C(X_L^{NC}-\underline{\theta})-C(X_L^{NC}-\overline{\theta})\right]$ to the supervisor in opposition to the collusive offer by the good type, in order to elicit the true information. In summary, the principal strictly improves his payoff by changing X_L^{NC} into X_L^{FB} ex-post, but the ex-ante incentive cost $-kU(\overline{\theta})$ arises. This is the trade-off for the principal when the supervisor obtains the proof of true information $\theta=\overline{\theta}$, 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 and the same trade-off between the total surplus and the information rent emerges.

Formally, the expected virtual surplus in the principal-supervisor-agent regime is written as

[4] This idea is similar to the renegotiation problem from lack of commitment to the long-term contract, which was first considered by Dewatripont (1988)

$$\underbrace{(1-p)}_{\theta \text{ is not revealed}}\left\{\underbrace{h\left[X_H-C\left(X_H-\overline{\theta}\right)\right]+(1-h)\left[X_L-C\left(X_L-\underline{\theta}\right)\right]}_{\text{Expected Total Surplus}}-\underbrace{hU\left(\overline{\theta}\right)}_{\substack{\text{information rent}\\ \text{for the good type}}}\right\}$$

$$+\underbrace{p}_{\theta \text{ is revealed}}\left\{\underbrace{h\left[X_H^{FB}-C\left(X_H^{FB}-\overline{\theta}\right)\right]+(1-h)\left[X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right]}_{\text{(Ex post) First Best Allocative Efficiency}}-\underbrace{khU\left(\overline{\theta}\right)}_{\substack{\text{information rent}\\ \text{for the supervisor}}}\right\}$$

When the principal determines the output targets $\{X_L, X_H\}$ for the no-information phase $\varnothing$, she must consider the information rent for the supervisor $pkU\left(\overline{\theta}\right)$ as well as the information rent for the good agent $(1-p)U\left(\overline{\theta}\right)$. The principal will optimize the bad agent $\underline{\theta}$'s output X_L, in order to mitigate the collusive pressure by the good agent when the supervisor observed the signal $\theta=\overline{\theta}$, and to deal with the standard trade-off between the total surplus generated by X_L and the information rent for the good agent $\overline{\theta}$ when the supervisor could not obtained any information for θ. Thus, in this regime, the principal maximizes the following modified *virtual surplus*.

$$\max_{\{X_H, X_L\}}(1-p)\left\{\underbrace{h\left[X_H-C\left(X_H-\overline{\theta}\right)\right]+(1-h)\left[X_L-C\left(X_L-\underline{\theta}\right)\right]}_{\text{Expected Total Surplus}}-\underbrace{hU\left(\overline{\theta}\right)}_{\substack{\text{information rent}\\ \text{for the good type}}}\right\}-p\underbrace{khU\left(\overline{\theta}\right)}_{\substack{\text{information rent}\\ \text{for the supervisor}}}$$

The first order conditions for the optimum are,

$$1-\frac{\partial C\left(X_H-\overline{\theta}\right)}{\partial X_H}=0 \Leftrightarrow X_H^{NC}=X_H^{FB}$$

$$\underbrace{1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}}_{\text{Marginal Total Surplus}}-\frac{h}{1-h}\underbrace{\left[1+\frac{pk}{1-p}\right]}_{\geq 1}\underbrace{\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\overline{\theta}\right)}{\partial X_L}\right]}_{\text{Marginal Information Rent}}=0$$

We now have the following proposition on the comparison of equilibrium incentives.

Proposition3:

Let X_H^{NC} and X_L^{NC} be the outputs (in the no-information phase) of the good type $\overline{\theta}$ and the bad type $\underline{\theta}$, respectively. Then, we have:

(1) Efficiency at the top (for the good-type agent) $X_H^{NC}=X_H^{FB}$

(2) Downward distortion at the bottom (for the bad-type agent) is aggravated:

$$X_L^{NC}\leq X_L^{*}\leq X_L^{S}\leq X_L^{FB}$$

$X_L^{NC}\leq X_L^{*}$ in the above proposition comes from the increase in the virtual cost, i.e., the total and marginal information rents in this regime. Virtual marginal cost increases by $pk/(1-p)$, compared with the standard no-supervisor case. Figure 6 clearly shows this point.

$$\frac{h}{1-h}\underbrace{\left[1+\frac{pk}{1-p}\right]}_{\geq 1}\underbrace{\left[\frac{\partial C(X_L-\underline{\theta})}{\partial X_L}-\frac{\partial C(X_L-\overline{\theta})}{\partial X_L}\right]}_{\text{Marginal Information Rent}}$$

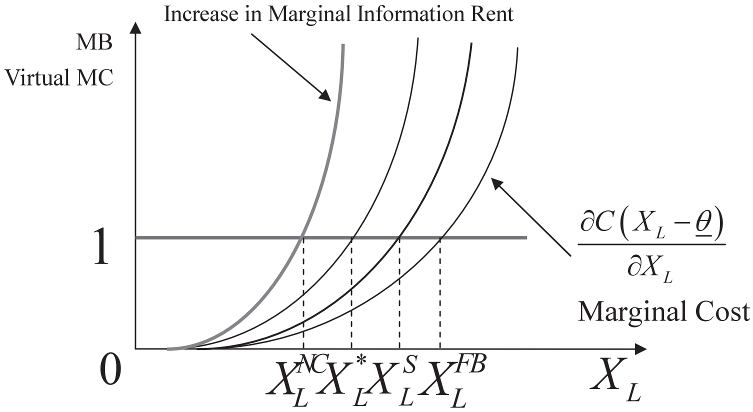

Figure 6

Now, we can perform a comparative statics on the optimal solution X_L^{NC}.

<u>Proposition4</u>: Comparative statics on X_L^{NC}

The optimal output X_L^{NC} in this no commitment/renegotiation regime is non-increasing in the parameter *p*, and non-increasing in the parameter *k*.

Proof:

The coefficient of the marginal information rent $1+\frac{pk}{1-p}$ increases as the parameter *p* increases. Hence, the marginal information rent (and so the marginal virtual cost) $\frac{h}{1-h}\left[1+\frac{pk}{1-p}\right]\left[\frac{\partial C(X_L-\underline{\theta})}{\partial X_L}-\frac{\partial C(X_L-\overline{\theta})}{\partial X_L}\right]$ increases as *p* increases. This brings about the decrease in the optimal output $X_L^{NC}\downarrow$. Similarly, the coefficient of the marginal information rent $1+\frac{pk}{1-p}$ increases as the parameter *k* increases. Hence, the marginal information rent (and so the marginal virtual cost) increases as *k* increases. This brings about the decrease in the optimal output $X_L^{NC}\downarrow$. ■

5. The Trick of the Model

Indeed, we have so far considered the following parameterized objective function:

$$\max_{X_H, X_L} \underbrace{h\left[X_H - C\left(X_H - \bar{\theta}\right)\right] + (1-h)\left[X_L - C\left(X_L - \underline{\theta}\right)\right]}_{\text{Expected Total Surplus}} - z \cdot \underbrace{h\left[C\left(X_L - \underline{\theta}\right) - C\left(X_L - \bar{\theta}\right)\right]}_{\text{Information Rent}}$$

where z is a parameter, and $z = 0$ corresponds to total surplus maximization and $z = 1$ corresponds to the standard second best problem of principal and one agent. Similarly, $z = (1-p) + pk$ corresponds to the principal-supervisor-agent regime with full commitment (see Section 3) and $z = 1 + \frac{pk}{1-p}$ corresponds to the no-commitment, renegotiation regime (see Section 4).

We see that the objective function is additively separable in X_L and X_H. Hence the program can be broken into two:

$$\max_{X_H} h\left[X_H - C\left(X_H - \bar{\theta}\right)\right]$$

$$\max_{X_L} (1-h)\left[X_L - C\left(X_L - \underline{\theta}\right)\right] - z \cdot h\left[C\left(X_L - \underline{\theta}\right) - C\left(X_L - \bar{\theta}\right)\right]$$

From the first line we see that X_H maximizes the total surplus for the good type, and obtain the "efficiency at the top" result.

As for X_L, the formula for the optimal solution of X_L in each regime is

$$\underbrace{1}_{\text{Marginal Benefit}} = \underbrace{\underbrace{\frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L}}_{\text{Marginal Cost}} + z \cdot \underbrace{\frac{h}{1-h}\left[\frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L} - \frac{\partial C\left(X_L - \bar{\theta}\right)}{\partial X_L}\right]}_{\text{Marginal Information Rent}}}_{\text{Marginal Virtual Cost}} \qquad (*)$$

Since the square bracket of the second term of right-hand side is positive, we see that as the parameter z becomes bigger (smaller), the marginal virtual cost becomes bigger (smaller). Hence, the optimal solution for X_L becomes smaller (bigger). As a result, we have Proposition 4 on the comparison of the optimal solutions.

6. Yardstick Mechanism in Perfectly Correlated Environments

As shown so far, the principal cannot attain the first best under one-agent hidden information setting. That is, "downward distortion at the bottom $X_L^* \leq X_L^{FB}$" occurs. The introduction of the supervisor can mitigate, but cannot solve the problem perfectly, and indeed in some case the distortion may become bigger (as in the no commitment /renegotiation regime).

In this section, we show that in the case where the private information of the two agents is *perfectly correlated*, the principal can implement the full information first best optimum *at no incentive cost*. That is, we show that the principal can attain the points **G** and **B** in Figure 1 in the unique equilibrium.

Remember that $\left\{X_H^{FB}, X_L^{FB}\right\} = \left\{\bar{\theta} + e^{FB}, \underline{\theta} + e^{FB}\right\}, W_H^{FB} = W_L^{FB} = C\left(e^{FB}\right)$. When the agent of type $\overline{\theta}$ chooses the point **B** $\{W_L^{FB}, X_L^{FB}\} = \{C(e^{FB}), \underline{\theta} + e^{FB}\}$ in *one- agent* setting, he can obtain the information rent $C(e^{FB}) - C(e^{FB} - (\overline{\theta} - \underline{\theta}))$, as already examined above.

Now, we assume that there are two agents, whose private information is *perfectly correlated*, and do not introduce the supervisor into the setting (for simplicity).

Formally, we focus on the case where the private information (θ_i, θ_j) of the agents i, j is *perfectly correlated*, i.e.,

$$(\theta_i, \theta_j) = \begin{cases} (\overline{\theta}, \overline{\theta}) & \text{with prob. } \quad h \\ (\underline{\theta}, \underline{\theta}) & \text{with prob. } \ 1-h \end{cases}$$

The principal proposes the following contract $W_i(X_i, X_j)$ to the agent i = 1, 2, $i \neq j$.

$$W_i(X_i, X_j) = \begin{cases} \underbrace{C(e^{FB})}_{=W_H^{FB}} & \text{if } (X_i, X_j) = (X_H^{FB}, X_H^{FB}) \\ \underbrace{C(e^{FB})}_{=W_L^{FB}} & \text{if } (X_i, X_j) = (X_L^{FB}, X_L^{FB}) \\ \underbrace{C(e^{FB})}_{=W_H^{FB}} + \underbrace{\left[C(e^{FB}) - C(e^{FB} - (\overline{\theta} - \underline{\theta})) + \Delta\right]}_{\text{Information Rent}+\Delta} & \text{if } (X_i, X_j) = (X_H^{FB}, X_L^{FB}) \\ \underbrace{C(e^{FB})}_{=W_L^{FB}} - \underbrace{\left[C(e^{FB}) - C(e^{FB} - (\overline{\theta} - \underline{\theta})) + \Delta\right]}_{\text{Information Rent}+\Delta} & \text{if } (X_i, X_j) = (X_L^{FB}, X_H^{FB}) \end{cases}$$

First, we explore the incentive of the agent of type $\theta_i = \overline{\theta}$ in the state $(\overline{\theta}, \overline{\theta})$ under the above scheme. His payoff function is written as $W_i(X_i, X_j) - C(X_i - \overline{\theta})$, which is the payoff when the agent $\theta_i = \overline{\theta}$ chooses the output X_i given that the other agent chooses X_j.

Suppose that the other agent $\theta_j = \overline{\theta}$ chooses the output X_L^{FB}.

Then, if the agent $\theta_i = \overline{\theta}$ chooses the output X_L^{FB}, he will obtain the payoff

$W_i(X_L^{FB}, X_L^{FB}) - C(X_L^{FB} - \overline{\theta}) = C(e^{FB}) - C(e^{FB} - (\overline{\theta} - \underline{\theta}))$.

If the agent $\theta_i = \overline{\theta}$ chooses the output X_H^{FB}, he will obtain the payoff

$$W_i(X_H^{FB}, X_L^{FB}) - C(X_H^{FB} - \overline{\theta}) = C(e^{FB}) + \underbrace{\left[C(e^{FB}) - C(e^{FB} - (\overline{\theta} - \underline{\theta})) + \Delta\right]}_{\text{Information Rent}+\Delta} - C(e^{FB}) =$$

$$C(e^{FB}) - C(e^{FB} - (\overline{\theta} - \underline{\theta})) + \Delta.$$

Hence, the agent $\theta_i = \overline{\theta}$ has an incentive to choose the output X_H^{FB}.

Next, suppose that the other agent $\theta_j = \overline{\theta}$ chooses the output X_H^{FB}.

Then, if the agent $\theta_i = \overline{\theta}$ chooses the output X_L^{FB}, he will obtain the payoff

$$W_i\left(X_L^{FB}, X_H^{FB}\right) - C\left(X_L^{FB} - \bar{\theta}\right) = C\left(e^{FB}\right) - \underbrace{\left[C\left(e^{FB}\right) - C\left(e^{FB} - \left(\bar{\theta} - \underline{\theta}\right)\right) + \Delta\right]}_{\text{Information Rent}+\Delta} - C\left(e^{FB} - \left(\bar{\theta} - \underline{\theta}\right)\right) = -\Delta$$

If the agent $\theta_i = \bar{\theta}$ chooses the output X_H^{FB}, he will obtain the payoff

$W_i\left(X_H^{FB}, X_H^{FB}\right) - C\left(X_H^{FB} - \bar{\theta}\right) = C\left(e^{FB}\right) - C\left(e^{FB}\right) = 0$.

Hence, the agent $\theta_i = \bar{\theta}$ has an incentive to choose the output X_H^{FB}. That is, regardless of the other player $\theta_j = \bar{\theta}$'s choices, the agent $\theta_i = \bar{\theta}$ has an incentive to choose the output X_H^{FB}.

The incentive structure of the agent $\theta_j = \bar{\theta}$ is also the same. Regardless of the other player $\theta_i = \bar{\theta}$'s choices, the agent $\theta_j = \bar{\theta}$ has an incentive to choose the output X_H^{FB}. The choice of X_H^{FB} is the dominant strategy for the agent $\theta_j = \bar{\theta}$.

The payoff matrix is as follows.

$\theta_i = \bar{\theta}$ \ $\theta_j = \bar{\theta}$	X_L^{FB}	X_H^{FB}
X_L^{FB}	$C\left(e^{FB}\right) - C\left(e^{FB} - \left(\bar{\theta} - \underline{\theta}\right)\right)$ $C\left(e^{FB}\right) - C\left(e^{FB} - \left(\bar{\theta} - \underline{\theta}\right)\right)$	$C\left(e^{FB}\right) - C\left(e^{FB} - \left(\bar{\theta} - \underline{\theta}\right)\right) + \Delta$ $-\Delta$
X_H^{FB}	$-\Delta$ $C\left(e^{FB}\right) - C\left(e^{FB} - \left(\bar{\theta} - \underline{\theta}\right)\right) + \Delta$	0 0

Perfect correlation of the private information $\left(\theta_i, \theta_j\right) = \left(\bar{\theta}, \bar{\theta}\right)$ is crucial. As the payoff matrix shows, the principal places the two agents in a *prisoner's dilemma* game. By exploiting this structure, the principal can implement the full information first best optimum in the unique dominant strategy equilibrium.[5]

Next, we explore the incentive of the agent of type $\theta_i = \underline{\theta}$ in the state $\left(\underline{\theta}, \underline{\theta}\right)$. His payoff function is written as $W_i\left(X_i, X_j\right) - C\left(X_i - \underline{\theta}\right)$ when the agent $\theta_i = \underline{\theta}$ chooses the output X_i given that the other agent chooses X_j.

Suppose that the other agent $\theta_j = \underline{\theta}$ chooses the output X_L^{FB}.

Then, if the agent $\theta_i = \underline{\theta}$ chooses the output X_L^{FB}, he will obtain the payoff

$W_i\left(X_L^{FB}, X_L^{FB}\right) - C\left(X_L^{FB} - \underline{\theta}\right) = C\left(e^{FB}\right) - C\left(e^{FB}\right) = 0$.

If the agent $\theta_i = \underline{\theta}$ chooses the output X_H^{FB}, he will obtain the payoff

[5] A key problem in the design of optimal contracts in correlated environments is the possibility of *multiple equilibria* in the subgame played by the parties whose private information is correlated. As noted by Demski and Sappington (1984), multiple equilibria do not pose a problem when the private information is *perfectly correlated*.

$$W_i\left(X_H^{FB}, X_L^{FB}\right) - C\left(X_H^{FB} - \underline{\theta}\right)$$

$$= C\left(e^{FB}\right) + \left[C\left(e^{FB}\right) - C\left(e^{FB} - \left(\overline{\theta} - \underline{\theta}\right)\right) + \Delta\right] - C\left(e^{FB} + \left(\overline{\theta} - \underline{\theta}\right)\right)$$

$$= 2\left[C\left(e^{FB}\right) - \frac{C\left(e^{FB} - \left(\overline{\theta} - \underline{\theta}\right)\right) + C\left(e^{FB} + \left(\overline{\theta} - \underline{\theta}\right)\right)}{2}\right] + \Delta < 0$$

The negative sign is due to the convexity of the cost function $C' > 0, C'' > 0$.
Hence, the agent $\theta_i = \underline{\theta}$ has an incentive to choose the output X_L^{FB}.

Next, suppose that the other agent $\theta_j = \underline{\theta}$ chooses the output X_H^{FB}.
Then, if the agent $\theta_i = \underline{\theta}$ chooses the output X_L^{FB}, he will obtain the payoff

$$W_i\left(X_L^{FB}, X_H^{FB}\right) - C\left(X_L^{FB} - \underline{\theta}\right) = C\left(e^{FB}\right) - \left[C\left(e^{FB}\right) - C\left(e^{FB} - \left(\overline{\theta} - \underline{\theta}\right)\right) + \Delta\right] - C\left(e^{FB}\right)$$

$$= -\left[C\left(e^{FB}\right) - C\left(e^{FB} - \left(\overline{\theta} - \underline{\theta}\right)\right) + \Delta\right] < 0$$

If the agent $\theta_i = \underline{\theta}$ chooses the output X_H^{FB}, he will obtain the payoff
$W_i\left(X_H^{FB}, X_H^{FB}\right) - C\left(X_H^{FB} - \underline{\theta}\right) = C\left(e^{FB}\right) - C\left(e^{FB} + \left(\overline{\theta} - \underline{\theta}\right)\right) < 0$.
Taking the difference of the payoffs, we have

$$C\left(e^{FB}\right) - C\left(e^{FB} + \left(\overline{\theta} - \underline{\theta}\right)\right) + \left[C\left(e^{FB}\right) - C\left(e^{FB} - \left(\overline{\theta} - \underline{\theta}\right)\right) + \Delta\right]$$

$$= 2\left[C\left(e^{FB}\right) - \frac{C\left(e^{FB} - \left(\overline{\theta} - \underline{\theta}\right)\right) + C\left(e^{FB} + \left(\overline{\theta} - \underline{\theta}\right)\right)}{2}\right] + \Delta < 0$$

Hence, the agent $\theta_i = \underline{\theta}$ has a strict incentive to choose the output X_L^{FB}, or no incentive to deviate from X_L^{FB} to X_H^{FB}.

Thus, regardless of the other agent $\theta_j = \underline{\theta}$'s choices, the agent $\theta_i = \underline{\theta}$ has an incentive to choose the output X_L^{FB}. The choice of X_L^{FB} is the dominant strategy for the bad agent $\theta_i = \underline{\theta}$.

The incentive structure of the agent $\theta_j = \underline{\theta}$ is the same. Regardless of the other player $\theta_i = \underline{\theta}$'s choices, the agent $\theta_j = \underline{\theta}$ has a strict incentive to choose the output X_L^{FB}. The choice of X_L^{FB} is the dominant strategy for the bad agent $\theta_j = \underline{\theta}$.
The payoff matrix is as follows.

$\theta_j = \underline{\theta}$ \ $\theta_i = \underline{\theta}$	X_L^{FB}	X_H^{FB}
X_L^{FB}	0 0	$2\left[C\left(e^{FB}\right)-\frac{C\left(e^{FB}-\left(\bar{\theta}-\underline{\theta}\right)\right)+C\left(e^{FB}+\left(\bar{\theta}-\underline{\theta}\right)\right)}{2}\right]+\Delta<0$ $-\left[C\left(e^{FB}\right)-C\left(e^{FB}-\left(\bar{\theta}-\underline{\theta}\right)\right)+\Delta\right]<0$
X_H^{FB}	$-\left[C\left(e^{FB}\right)-C\left(e^{FB}-\left(\bar{\theta}-\underline{\theta}\right)\right)+\Delta\right]<0$ $2\left[C\left(e^{FB}\right)-\frac{C\left(e^{FB}-\left(\bar{\theta}-\underline{\theta}\right)\right)+C\left(e^{FB}+\left(\bar{\theta}-\underline{\theta}\right)\right)}{2}\right]+\Delta<0$	$C\left(e^{FB}\right)-C\left(e^{FB}+\left(\bar{\theta}-\underline{\theta}\right)\right)<0$ $C\left(e^{FB}\right)-C\left(e^{FB}+\left(\bar{\theta}-\underline{\theta}\right)\right)<0$

Perfect correlation of the private information $\left(\theta_i,\theta_j\right) = \left(\underline{\theta},\underline{\theta}\right)$ is also crucial here. As the payoff matrix shows, the principal places the two agents in a *prisoner's dilemma* game, thereby implementing the full information first best optimum in the unique dominant strategy equilibrium at no incentive cost.

Summarizing the arguments so far, we have:

<u>Proposition5</u>:

Under perfect correlation of the private information of the two agents, the principal can implement the full information first best optimum in the unique dominant strategy equilibrium[6], without giving any information rent. The equilibrium contracts are Pareto efficient in both states $\left(\bar{\theta},\bar{\theta}\right)$ and $\left(\underline{\theta},\underline{\theta}\right)$.

6. Conclusion

Recently, auditing to meet the needs of corporate governance has been rapidly growing 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 constructed a simple three-tier agency model, which is a natural extension of the familiar screening (self selection) models and can be placed in line of the collusion literature à la Tirole (1986, 1992), including Kofman and Lawarree (1993)'s auditing application. 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 margin-

[6] Of course, in this dominant strategy equilibrium, the Nash incentive compatibility constraints are also satisfied, in the sense that it is optimal for the agent i to behave truthfully *given* that the agent j behaves truthfully, and vice versa.

al 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, i.e., "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.[7]

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 and 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. 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" regime. The principal can do better in the equilibrium without some of her commitments than if she can fully commit.

In any case, the principal cannot attain the first best under the principal-supervisor-one agent hidden information setting with collusion, and "downward distortion at the bottom" always occurs. Introduction of the supervisor can mitigate, but cannot solve the problem perfectly, and indeed in some case the distortion may become bigger (as in the no-commitment /renegotiation regime). So, finally, we show that in the case where the private information of the two agents is *perfectly correlated*, the principal can implement the full information first best optimum *at no incentive cost*. Intuitively, the principal places the two agents in a *prisoner's dilemma* game and both the truth-telling and the first best optimal incentives can be induced in the unique dominant strategy equilibrium, without giving any information rent.

Acknowledgements

This research was supported by Grant-in-Aid for Scientific Research by Japan Society for the Promotion of Science(C) No. 20530162 (2008-2010)

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

REFERENCES

Baron,D. and R, Myerson (1982) “Regulating a Monopolist with Unknown Cost,” *Econometrica* 50. 911-930

Bolton, P and M. Dewatripont (2005) *Contract Theory* MIT Press

Demski,J.S. and D. Sappington (1984) “Optimal Incentive Contracts with Multiple Agents,” *Journal of Economic Theory* 33(1), June 152-171

Dewatripont, M (1988) “Commitment and Information Revelation over Time: The case of Optimal Labor Contracts.” *Quarterly Journal of Economics*, 104, 589-619.

Kofman, F, and J. Lawarree (1993) “Collusion in Hierarchical Agency”, *Econometrica*, Vo1.61, No3, May, 629-656.

Laffont, J-J and D. Martimort (1997) “Collusion under Asymmetric Information” *Econometrica*, Vol.65, No4, 875-911.

Laffont, J-J, and J. Tirole (1988) “The Dynamics of Incentive Contracts,” *Econometrica*, Vol. 56 No.5, 1153-75, September

Laffont, J-J and J. Tirole (1991) “The Politics of Government Decision-Making: A Theory of Regulatory Capture," *Quarterly Journal of Economics*.106, 1089-1127.

Ma, C., J. Moore and S . Turnbull (1988) "Stopping Agents from Cheating," *Journal of Economic Theory*, Vol. 46, 355-372

Maskin,E and J,Riley (1984) "Monopoly with Incomplete Information," *RAND Journal of Economics*, Vol.15. No.2. 171-196.

Suzuki, Y. (2007) “Collusion in Organizations and Management of Conflicts through Job Design and Authority Delegation”, *Journal of Economic Research* 12. 203-241

Suzuki, Y. (2007) “Mechanism Design with Collusive Auditing : A Three-Tier Agency Model with "Monotone Comparative Statics" and an Implication for Corporate Governance”, Institute of Comparative Economic Studies, Hosei University, Working Paper No.128.

Suzuki, Y. (2008)"Mechanism Design with Collusive Supervision: A Three-tier Agency Model with a Continuum of Types,” *Economics Bulletin*, Vol. 4 No. 12. 1-10

Tirole, J. (1986) “Hierarchies and Bureaucracies: On the role of Collusion in Organizations". *Journal of Law, Economics and Organization*. 2. 181-214.

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.

Shleifer,A.(1985) “A Theory of Yardstick Competition,” *RAND Journal of Economics*, Vol. 16, No. 3. 319-327.