Económica, La Plata. Vol. IL. Nro. 1-2, 2003

# ON MONITORING TIMING IN HIERARCHIES[1,2]

WALTER CONT [3]

## 1. Introduction

The literature on the principal-agent problem has analyzed the role that monitoring institutions play in alleviating incentive problems. Two branches of this literature have been studied separately. The first one analyzes the optimal contract when the principal hires an *ex ante* monitor (or, for lack of a better name, a supervisor, see Tirole (1986)), whereas the second one corresponds to a principal hiring an *ex post* monitor (or auditor, see Baron and Besanko (1984)).

However, not much attention has been given to what affects the choice between the two monitoring institutions.[4] Several reasons justify the importance of this problem. First, the evidence suggests that although both institutions are available to the principal, the latter chooses only one of them.[5] Hence, a formal explanation is needed to rationalize this

---

[1] JEL Classification: D82, L23.

[2] I thank Leandro Arozamena, Hongbin Cai, Pedro dal Bó, David Levine, Jean-Laurent Rosenthal, seminar participants at UCLA and at the XXXVI Annual Meeting of the *Asociación Argentina de Economía Política* and an anonymous referee for their comments.

[3] Fundación de Investigaciones Económicas Latinoamericanas and Universidad Nacional de La Plata. E-mail: walter@fiel.org.ar

[4] This paper considers situations in which the timing choice is an option for the principal. There may exist real life situations in which only one monitoring stage is feasible and hence the standard literature applies.

[5] There are also circumstances in which the principal chooses both monitors. We leave this discussion aside in this paper to focus on the principal's timing choice.

evidence. Owners of firms hire third parties to supervise employees, to monitor their effort or to audit their private information. In some circumstances, the regulatory stage is ex ante (e.g., regulatory framework, detailed contracts) while in other circumstances the control is ex post (e.g., auditing of accounting data and expenses).

Second, and more important, the monitoring timing choice is not inconsequential. Hiring an ex ante or an ex post monitor has a different effect on both agent's incentives and principal's utility. On the one hand, the ex ante monitor obtains information about the agent's productivity *before* the agent accepts the contract (and, of course, exerts effort), and his report is used to design a "flexible" contract for the agent, in which both output and the agent's compensation can be based on the monitor's report. On the other hand, the ex post monitor obtains information about the agent's effort or productivity *after* the agent exerted her effort. The principal can use the auditor's report to punish the agent (provided that a punishment scheme is available and enforceable), but she cannot make the contracting of output depend on this report. Therefore, the principal faces a *trade-off between flexibility and rigidity-punishment.*[6]

In order to find the optimal solution to this trade-off, we specify a general model (principal-monitor-agent hierarchy), which allows both supervising and auditing. The principal, who is uninformed about the agent's productivity and effort, hires the agent to produce a good or service. The standard optimal contract solves a trade-off between incentives and costly information rents. In addition, the principal may hire a monitor (whose preferences are aligned to the principal's) to elicit part of the agent's private information, which can be used to alleviate this trade-off (i.e.,

[6] The *Law and Economics* literature considered the stage of legal intervention or (benevolent) regulation of activities that generate externalities (see Shavell (1993)). This literature does not consider the trade-off analyzed in this paper.

reducing the agent's rents and improving incentives).

We find that auditing is optimal when strong and enforceable punishment schemes are available to the principal, or when punishment instruments are weak and the monitor's signal is noisy. Otherwise, supervising is optimal when punishment schemes are weak or cannot be enforced, provided that the supervisor's signal is accurate. Given a low expected punishment, the supervisor is more valuable to the principal when he learns the "right" information about the agent (which is more probable when his signal is more accurate), for the principal can reduce the agent's expected rents when she is certain about the agent's type.

The formal results in this paper fit casual observations in organizations, such as auditing of top-level managers (e.g., CEOs, who may be more exposed to punishments) or supervision of productive activities at lower levels of a firm. Typically, low-level employees have lower incomes or are protected by minimum wage regulations. Of course, company shareholders prefer to "screen" a candidate to a high-rank (i.e., managerial) position when they cannot rely on courts as an enforcement device of eventual punishments. These results also apply to regulation of "hazardous" activities. In particular, the determinants of the optimal stage of intervention such as magnitude of possible sanctions or probability of application of sanctions arise naturally under this agency-based structure. A new important determinant is the quality of the monitor's information (measured as signal accuracy of the agent's productivity), which is valuable to achieve some flexibility ex ante (i.e., prevention of damages at a low cost).

This paper connects many works dedicated to monitoring in hierarchies, which apply to either supervising or auditing. Baron and Besanko (1984) analyzed the optimal design of a regulatory contract when the government hires a benevolent regulator to audit a firm. They obtain a separation result that the pricing decision does not depend on the auditing decision (which means that the price and quantity when the firm is audited are the same as those when the firm is not audited), but the auditing decision does depend on the pricing decision (in particular, the principal sends the auditor when she infers that the firm overstated the price). Cohen (1987) characterizes the optimal enforcement for an environmental regulator to prevent oil spills (i.e., a negative externality) in an agency-based framework with moral hazard. Tirole (1986) introduces the optimal contract when a supervisor is hired, but his concern is about the effect of collusion between the supervisor and the agent on such contract (see Section 5 for a brief discussion). However, none of these works consider the optimality of supervising as compared with auditing. Finally, as we mentioned before, the Law and Economics literature (Shavell (1993) and Kolstad, Ulen y Johnson (1990)) has analyzed the optimal stage of Law enforcement of externality-generating activities in a benevolent-regulator framework. However, these papers do not consider agency problems or imperfect signals.

The paper is organized as follows. Section 2 outlines the model. Section 3 computes the optimal contract when the principal hires a supervisor or an auditor. Section 4 discuses the optimal monitoring timing, and provides some applications to organization design and regulation. Finally, Section 5 concludes and discusses some extensions.

## 2. The Model

Consider a hierarchy consisting of a principal, a monitor and an agent. The principal hires the agent to produce a good with gross value $V$ and production cost $C = \bar{\theta} - \theta e$. The payoff to the principal is $V - \bar{\theta} + \theta e$.[7] The

[7] This specification of the model nests regulation models (with cost function $C$) –where the principal is the government, the monitor is the regulator and the agent is the regulated firm– and organization models (with profit function $\pi \approx \theta e$) –where the

cost $C$, which is *observed* by the principal, is reduced by a combination of agent's productivity and effort, which are *not observed* by the principal. By exerting higher effort the agent reduces the production cost, but she derives a private effort disutility or cost $\psi(e) = e^2/2$. The agent's private productivity or type is $\theta \in \{\theta_L, \theta_H\}$, with $\theta_H > \theta_L > 0$. Let $q$ be the ex ante probability that the agent's productivity is high, i.e., $q = Pr(\theta = \theta_H)$. The parameter $\bar{\theta}$ is an upper bound on the production cost.[8] The principal reimburses the cost $C$ and pays a net transfer $t$ to the agent. The agent's reservation utility is $U_0 = 0$.

The principal also decides whether to hire a monitor to obtain an imperfect signal about the agent's productivity (this signal is also observed by the agent). The monitor obtains a signal at no cost (the results extend to a costly monitor, provided that he is hired). The signal may take the following values: With probability $1 - p$ the monitor learns nothing about the agent's type ($\sigma = 0$). Otherwise he gets an imperfect observation $\sigma \in \{L, H\}$, which is correct with probability $\alpha > 1/2$. This assumption satisfies the monotone likelihood ratio property that a correct signal is more probable. Table 1 summarizes the possible signals and their corresponding probabilities.

With the new information, the principal may reduce the agent's wage or set a fine whenever she finds that the agent misreported her type or shirked. The agent is protected by limited liability when punished: a fine $z^r$ set by the principal (depending on the monitor's report $r$) must be up to some liability bound $z$. The liability bound may be interpreted

---

principal is the owner, the monitor is either a supervisor or an auditor and the agent is the manager or worker–.

[8] We show later that the effort exerted by a type-$\theta_H$ agent is $e = \theta_H$, and hence we assume that $\bar{\theta} > \theta_H^2$ for the observed cost to be positive in all the cases analyzed in this paper.

Table 1: Monitor's Signal of Agent's Type

| Type | Observation | $\sigma$ | Probability |
|---|---|---|---|
| $\theta_H$ | 0 | 0 | $1-p$ |
| $\theta_H$ | $\theta_H$ | $H$ | $p\alpha$ |
| $\theta_H$ | $\theta_L$ | $L$ | $p(1-\alpha)$ |
| $\theta_L$ | 0 | 0 | $1-p$ |
| $\theta_L$ | $\theta_H$ | $H$ | $p(1-\alpha)$ |
| $\theta_L$ | $\theta_L$ | $L$ | $p\alpha$ |

as exogenous wealth constraints or exogenous maximum expected legal punishment.[9]

The monitor sends a report $r \in \{0, L, H\}$ to the principal, who pays him a wage $w$. He is protected by limited liability ($w \geq 0$). His reservation utility is $U_M = 0$.

We introduce the possibility for the principal to send the monitor *before* the agent accepts the contract or *after* the agent exerted effort. In the first case, the monitor obtains a signal about the agent's productivity (effort has not been exerted yet). In a large number of papers, this monitor is referred to as supervisor. In the second case, the monitor may either audit the agent's productivity or monitor the agent's effort. Given the cost structure ($C = \bar{\theta} - \theta e$), the information obtained by the principal is the same whether monitoring generates a signal $\sigma$ on produc-

[9] Alternatively, we can introduce some uncertainty on the punishment enforcement by assuming that the agent is punished with some probability $\rho \in [0, 1]$, so the expected punishment is $\rho z^r$. As it will be seen below, all that matters to the risk-neutral agent is the expected punishment. We assume that $z^r$ is expected punishment for simplicity.

tivity or effort, provided that the informativeness of signals and the cost of observing them are the same.[10] Hence we concentrate on (ex ante or ex post) productivity monitoring for convenience in the exposition. In addition, whether monitoring is ex ante or ex post, we assume the same signal distribution (same $p$ and $\alpha$) and the same cost ($c = 0$) to eliminate a possible source of timing preference derived from the quality or cost of information.

In order to make the timing decision, the principal compares costs and benefits under each alternative. If she hires a supervisor, she obtains a report that can be used to contract both output and wage. This gives some flexibility to the output choice, for the principal can create output distortions according to the probability of the events in order to reduce the expected rent to the agent. On the other hand, when the principal hires an auditor, she does not benefit from the flexibility in contracting output, but she can punish the agent (up to some point) when she finds that the agent misreported her type (or shirked). There is a trade-off: *flexibility in contracting vs. rigidity and punishment*.

The timing of the game is as follows:

1. The principal decides the monitoring timing.

---

[10] This result does not depend on the structural form of the cost function, but on the fact that, given the agent's type and effort, there is no uncertainty affecting the cost. Typical functions for cost or profit, found frequently in the principal-agent literature, are $C = \beta - e$ (where $\beta = \bar{\theta} - \theta$) and $\pi = \theta + e$, respectively. The results in this paper hold for these alternative structures. See Cont (2001).

Of course, if the informativeness of signals or the cost of obtaining them are different, the principal will choose effort monitoring or productivity auditing, whichever is more profitable. But this is a choice between two ex post monitoring alternatives, which we do not consider here.

2. Nature chooses the agent's type $\theta$. The agent learns her type, and the supervisor (if hired) observes a signal $\sigma$ of the agent's type (which is also observed by the agent).

3. At the contract stage, if the principal hires a supervisor, she offers contracts $t(C, r)$ to the agent and $w(C, r)$ to the supervisor. If the principal hires an auditor, she offers $\{t(C, r), z^r\}$ to the agent and $w(C, r)$ to the auditor. The three parties sign the contract. The agent and the supervisor (if hired) send their reports to the principal.

4. The agent chooses effort $e$, and cost $C$ is realized.

5. If an auditor is hired, he observes the signal $\sigma$, and sends a report $r$ to the principal.

6. Transfers are realized.

It is worth mentioning that we are assuming a standard information nesting used in the incentive literature: the agent knows her type and the monitor's signal, the monitor knows his signal and the distribution of agent's types, and the principal only knows the distributions of types and signals.

We also assume that all parties are risk- neutral. Since there are both moral hazard and adverse selection, a transfer of the hierarchy from the principal to the agent is not optimal. Moreover, limited liability to the monitor ensures that a transfer of the hierarchy from the principal to the monitor is not possible. The principal, agent and monitor's utilities are $U_P = V - [t + w + C]$, $U_A = t - e^2/2$ (minus an expected punishment, whenever it applies) and $U_M = w$, respectively.

When the principal observes both agent's effort and type, the problem simplifies to choosing effort and transfers in order to maximize $V - \bar{\theta} +$

$[\theta e - t]$, for $\theta \in \{\theta_L, \theta_H\}$, subject to the agent's interim participation constraint $t - e^2/2 \geq 0$. The solution to this problem is: $e_j^{FB} = \theta_j$, $t_j^{FB} = \theta_j^2/2$, for $j = L, H$. The principal's first-best utility is

$$EU_P^{FB} = V - \bar{\theta} + \frac{1}{2}\left[q\theta_H^2 + (1-q)\theta_L^2\right] \tag{1}$$

Suppose that the principal does not observe either agent's effort or type. The contract offered by the principal should be conditioned only on the observable $C$. Because of the binary nature of the problem and the fact that $C$ is deterministic for a given agent's type, we can concentrate on forcing contracts. As it is well known from revelation principle, the principal can restrict herself to direct mechanisms based on an agent's truthful report. For a report $\hat{\theta}$ there is an effort recommendation $e(\hat{\theta})$ to achieve a production cost $C(\hat{\theta}) = \bar{\theta} - \hat{\theta}e(\hat{\theta})$. When the agent's report is $\hat{\theta} = \theta_L$ $(\hat{\theta} = \theta_H)$, the principal recommends the agent to exert effort $e_L$ $(e_H)$ and pays her a transfer $t_L$ $(t_H)$ if the observed cost is $C_L$ $(C_H)$.

Let $\Delta\theta = (\theta_L/\theta_H)^2 < 1$ and $R = 1 - \Delta\theta < 1$. A feasible contract to the agent must satisfy the individual rationality (IR) and incentive compatibility (IC) constraints

$$\begin{array}{llll} IR(L): & t_L \geq e_L^2/2 & IC(L): & t_L - e_L^2/2 \geq t_H - e_H^2/2\Delta\theta \\ IR(H): & t_H \geq e_H^2/2 & IC(H): & t_H - e_H^2/2 \geq t_L - e_L^2\Delta\theta/2 \end{array}$$

A standard result is that when the constraints IR(L) and IC(H) are binding (and they are in the optimal contract), IC(L) and IR(H) are not binding (the proof is standard and hence omitted). The principal's problem with the binding constraints IR(L) and IC(H) is to choose $e_L$ and $e_H$ to maximize

$$V - \bar{\theta} + q\left[\theta_H e_H - \frac{e_H^2}{2} - R\frac{e_L^2}{2}\right] + (1-q)\left[\theta_L e_L - \frac{e_L^2}{2}\right]$$

The solution to this problem is

$$
\begin{aligned}
e_L^{NM} &= \frac{(1-q)\theta_L}{(1-q)+qR} & \qquad e_H^{NM} &= \theta_H \\
t_L^{NM} &= \frac{\left(e_L^{NM}\right)^2}{2} & \qquad t_H^{NM} &= \frac{\left(e_H^{NM}\right)^2}{2} + R\frac{\left(e_L^{NM}\right)^2}{2}
\end{aligned}
$$

and the principal achieves a utility

$$EU_P^{NM} = V - \bar{\theta} + \frac{1}{2}\left[q\theta_H^2 + \frac{(1-q)^2\theta_L^2}{(1-q)+qR}\right] \tag{2}$$

where the superscript $NM$ stands for no-monitor. Clearly $EU_P^{NM} < EU_P^{FB}$. The intuition for this result is as follows: in order to elicit high effort from the more productive agent (who has incentives to claim that she is inefficient), the principal pays her an information rent. But this rent is directly related to the type-$\theta_L$ agent's effort. Hence, the principal elicits lower effort from the less productive agent to reduce the type-$\theta_H$ agent's rent. In this way, the principal solves a trade-off between two kinds of costs: a higher cost from hiring a type-$\theta_L$ agent who exerts lower effort (than the first-best one) and a higher cost from hiring a type-$\theta_H$ agent who is paid positive rents.

To avoid paying rents, the principal may offer a contract targeted to the type-$\theta_H$ agent: $t = \theta_H^2/2$ if the observed production cost is $C_H = \bar{\theta} - \theta_H^2$, and nothing otherwise. The principal's utility is $EU_P^{NR} = q\left\{V - \bar{\theta} + \theta_H^2/2\right\}$. We assume that the principal prefers to hire both types of agents (it is sufficient to assume a high $V$).

## 3. Optimal Contract with a Monitor

This paper considers the case of a principal hiring an honest monitor to elicit –imperfectly– the agent's private information, and asks whether obtaining a monitor's signal before or after the agent exerted effort has some effect on her utility. The principal finds the monitor's signal valuable for the latter reduces the principal's information disadvantage regarding the agent's type, and hence allows her to reduce the agent's (expected) rents once she is better informed.

But the effect of this report on the agent's actions and compensation depends on the stage at which monitoring is exerted (i.e., the moment that the monitor obtains the signal and sends a report to the principal). In this section we solve for the optimal contract with a supervisor or an auditor separately and show how the monitor's report affects the agent's effort and compensation. Then we find conditions such that the principal chooses either of them (Section 4).

### 3.1. Contract with a Supervisor

The supervisor reports his signal ($r = \sigma$) to the principal at the beginning of the contract stage. The principal can use this report to contract agent's effort *and* wage. Let agent's effort be $e_{jr}$ and her compensation be $t_{jr}$ when she reports $\hat{\theta} \in \{\theta_L, \theta_H\}$ and the monitor reports $r \in \{0, L, H\}$ in a direct mechanism. The relevant constraints for a feasible contract are:[11]

---

[11] At this point it should be noticed that the problem we are interested in is that of a principal making use of the ex ante monitor's information to elicit part of the agent's private information (i.e., the monitor is functioning as a screening device, for example, through job interviews). Alternatively, we could assume that the signal is obtained after the agent signs the contract but before she exerts effort. One would expect the problem to be the same. However, $\sigma$ is revealed after the contract is signed, and the relevant agent's participation constraint becomes $\text{IR}_j$: $E_\sigma(t_{jr} - e_{jr}^2/2)$ for $j \in \{L, H\}$ and $r \in \{0, L, H\}$. The risk-neutrality assumption implies that the agent's wage could be reduced up to $-z$ in a low-probability state (compensated by a small wage increase in a high-probability stage), but this reduction could be used to alleviate incentives (by punishing ex ante). Although we consider this a technical possibility, we do not know

$$
\begin{array}{lll}
\text{IR(jr)}: & t_{jr} \geq e_{jr}^2/2 & \\
\text{IC(Lr)}: & t_{Lr} - e_{Lr}^2/2 \geq t_{Hr} - e_{Hr}^2/2\Delta\theta & \text{for } j \in \{L, H\},\ r \in \{0, L, H\} \\
\text{IC(Hr)}: & t_{Hr} - e_{Hr}^2/2 \geq t_{Lr} - e_{Lr}^2\Delta\theta/2 &
\end{array}
$$

Let $\pi_{jr}$ denote the probability of occurrence of each state, where $j \in \{L, H\}$ and $r \in \{0, L, H\}$.[12] Given that the supervisor reports to the principal honestly and that obtaining the signal is costless, the principal can set $w_{jr} = 0$. As in the case without monitor, if constraints IR(Lr) and IC(Hr) are binding (and they are in the optimal contract), the other constraints are non-binding.[13] The principal's problem with the binding constraints is to choose $\{e_{jr}\}$ to maximize

$$
V - \bar{\theta} + \sum_{r \in \{0,L,H\}} \left\{ \pi_{Hr} \left[ \theta_H e_{Hr} - \frac{e_{Hr}^2}{2} - R\frac{e_{Lr}^2}{2} \right] + \pi_{Lr} \left[ \theta_L e_{Lr} - \frac{e_{Lr}^2}{2} \right] \right\}
$$

The optimal effort and compensation are:

---

of contracts of this kind (perhaps because they are socially unacceptable), so we leave this possibility aside. Introducing risk aversion would make this option less attractive to the principal, but it would make the algebra too messy. Finally, notice also that the auditing contract would be less attractive to the principal if the agent was risk averse (Section 3.2 .

[12] There are six states with probabilities $\pi_{L0} = (1-q)(1-p)$, $\pi_{LL} = (1-q)p\alpha$, $\pi_{LH} = (1-q)p(1-\alpha)$, $\pi_{H0} = q(1-p)$, $\pi_{HL} = qp(1-\alpha)$, and $\pi_{HH} = qp\alpha$, where the first subscript corresponds to the agent's type and the second subscript corresponds to the monitor's signal.

[13] The proof is an extension of that in the No-Monitor case for a given supervisor's signal.

$$e_{H0} = e_{HL} = e_{HH} = \theta_H \qquad e_{L0} = \frac{(1-q)\theta_L}{(1-q)+qR} = e_L^{NM}$$
$$e_{LL} = \frac{(1-q)\alpha\theta_L}{(1-q)\alpha + q(1-\alpha)R} \qquad e_{LH} = \frac{(1-q)(1-\alpha)\theta_L}{(1-q)(1-\alpha)+q\alpha R} \tag{3}$$

$$t_{Lr} = e_{Lr}^2/2 \qquad t_{Hr} = \theta_H^2/2 + Re_{Lr}^2/2 \qquad j \in \{L,H\} \text{ and } r \in \{0,L,H\}$$
$$w_{jr} = 0 \tag{4}$$

$$\begin{array}{lll} \text{IR(jr)}: & t_{jr} \geq e_{jr}^2/2 & \\ \text{IC(Lr)}: & t_{Lr} - e_{Lr}^2/2 \geq t_{Hr} - e_{Hr}^2/2\Delta\theta & \text{for } j \in \{L,H\},\ r \in \{0,L,H\} \\ \text{IC(Hr)}: & t_{Hr} - e_{Hr}^2/2 \geq t_{Lr} - e_{Lr}^2\Delta\theta/2 & \end{array}$$

The next Proposition summarizes this result (see Tirole (1986)).[14]

**Proposition 1** *The optimal contract when the principal hires an honest supervisor satisfies (3)-(4). The supervisor is hired always.*

We can observe the benefits of flexibility in contracting from equations (3)-(4) and Figure 1. The principal pays only the effort cost to the type-$\theta_L$ agent for any monitor's report, but cannot eliminate the rents to the type-$\theta_H$ agent. Let the agent's rents in state Hr be $R_{Hr} = Re_{Lr}^2/2$. The optimal contract is such that $e_{LL} > e_{L0} > e_{LH}$ and $R_{HL} > R_{H0} > R_{HH}$ for any value of $\alpha$.

Consider, for example, the state in which a type-$\theta_H$ agent faces a report $r = H$ (this is a state which happens with higher probability as

[14] If the cost of supervising is positive, the contract should be corrected to internalize this cost. The results in Proposition 1 still hold, provided that the supervisor is hired.

$\alpha$ is bigger). The principal pays the agent enough rents to induce her to report her true type (otherwise the agent would exert effort $e_{LH}$ and would be compensated $t_{LH}$). But if $\theta_L$ was the agent's type and the monitor's report was $r = H$, this would be a state with lower probability for bigger $\alpha$. Hence, by obtaining the signal before the agent exerts effort, the principal finds it profitable to create higher distortions in the low-probability state LH (she reduces $e_{LH}$) when the signal is more accurate, which allows her to pay lower rents $R_{HH}$ in the high-probability state HH. Similarly, the principal reduces distortions in the high-probability state LL (she increases $e_{LL}$) when the signal is more accurate, which leads to higher rents $R_{HL}$ in the low-probability state HL. In the limiting case of perfectly informative signal ($\alpha = 1$), the high inefficiencies and rents are *ex ante* costless (states LH and HL have probability 0). Notice, in passing, that the principal designs a contract to adjust the type-$\theta_L$ agent's effort (and hence the $\theta_H$ agent's rent) to the posterior probability that the type is $\theta_L$. That is, the higher the posterior probability of $\theta = \theta_L$ given $r = L$ (i.e., $Pr(\theta_L | r = L)$), the lower the distortion and the higher the $\theta_H$ agent's rent, and viceversa, the lower the posterior probability of $\theta_L$ given $r = H$, the higher the distortion to reduce $\theta_H$ agent's rent.

Figure 2 displays the principal's expected utility as a function of the supervisor's signal accuracy (see equation (13) in the Appendix). This utility is increasing and convex, reflecting that the supervisor is more profitable to the principal when his signal is more accurate. Two cases are displayed in Figure 2. The first case corresponds to $p = 1$, and the principal's utility reaches the first best when $\alpha = 1$. This happens because the supervisor's signal is always correct and the information asymmetry between the principal and the agent is eliminated. The second case corresponds to $p < 1$. The principal's utility is always lower than the first-best utility because with some positive probability the supervisor obtains $\sigma = 0$ and the agent still keeps an informative advantage. Finally, regardless the value of $p$, the supervisor's signal is not informative at all as $\alpha \to 1/2$.

### *3.2. Contract with an Auditor*

By the time the principal sends the auditor to obtain information about the agent's type, the agent has already exerted effort and the outcome realized. Hence, the principal can use the auditor's report only to adjust compensations to the agent. From the Revelation Principle, the principal can relate a low production cost (or high output) to a type-$\theta_H$ agent. In this case there is no need to perform an audit, and the agent is paid $t_h$. When the production cost is high (or output is low), the principal cannot infer whether this is because the type-$\theta_L$ agent has exerted the right effort or the type-$\theta_H$ agent has shirked (the agent is paid $t_l$). In this case the principal sends the auditor (with probability $\delta \in [0, 1]$) and punishes the agent with a fine $z^r$ when the auditor's report does not match the agent's type report (when the agent reports her true type, this happens when the auditor reports $r \in \{0, H\}$). The punishment instrument is limited by a bound $z$, which may be understood as the agent's expected (perceived) fine, given the possibility that such punishment may not be enforced (for instance, by the courts), or as a liability bound. Using these results, the agent's participation and incentive constraints are

$$\begin{array}{ll} \mathrm{IR(L)}: & t_l - \delta[(1-p)z^0 + p(1-\alpha)z^H] \geq e_l^2/2 \\ \mathrm{IC(L)}: & t_l - \delta[(1-p)z^0 + p(1-\alpha)z^H] - e_l^2/2 \geq t_h - e_h^2/2\Delta\theta \\ \mathrm{IR(H)}: & t_h \geq e_h^2/2 \\ \mathrm{IC(H)}: & t_h - e_h^2/2 \geq t_l - \delta[(1-p)z^0 + p\alpha z^H] - e_l^2\Delta\theta/2 \end{array} \tag{5}$$

The principal pays $w^r = 0$ to the auditor for any report (since there is no incentive problem), and sends him whenever the production cost is high (i.e., she sets $\delta = 1$).[15] Let $\Omega_{ep} = \{e_h, e_l, t_h, t_l, z^0, z^H\}$ be the set of

[15] In this case, the result $\delta = 1$ depends on the auditor being costless and facing no

choice variables. The principal's problem is to choose $\Omega_{ep}$ to maximize

$$V-\bar{\theta}+q\left\{\theta_H e_h - t_h\right\}+(1-q)\left\{\theta_L e_l - t_l + \left[(1-p)z^0 + p(1-\alpha)z^H\right]\right\} \quad (6)$$

subject to constraints (5) and the limited liability constraints $z^r \leq z$ for $r \in \{0, H\}$. Let $\alpha_1^*$ denote the value of $\alpha$ such that IR(H) is non-binding for $\alpha < \alpha_1^*$, and $\alpha_2^*$ the value of $\alpha$ such that IC(H) is non-binding for $\alpha > \alpha_2^*$ (from equations (11) and (12) in the Appendix, respectively), where

$$\alpha_1^* = \frac{1}{2} + \frac{(1-q)^2\theta_L^2 R}{4pz\left[(1-q)+qR\right]^2} \qquad \alpha_2^* = \frac{1}{2} + \frac{\theta_L^2 R}{2pz}$$

The optimal effort and compensation are (this is a simplified version of Baron and Besanko (1984) , see also Kofman and Lawarrée (1993)):

$$\begin{array}{llll} & \alpha < \alpha_1^* & \alpha_1^* \leq \alpha \leq \alpha_2^* & \alpha_2^* < \alpha \\ e_l: & \dfrac{(1-q)\theta_L}{(1-q)+qR} & \sqrt{\dfrac{2p(2\alpha-1)z}{R}} & \theta_L \\ t_h: & \dfrac{\theta_H^2}{2} + \dfrac{e_l^2 R}{2} - p(2\alpha-1)z & \dfrac{\theta_H^2}{2} & \dfrac{\theta_H^2}{2} \end{array} \quad (7)$$

$$e_h = \theta_H, \qquad t_l = \frac{e_l^2}{2} + (1-p\alpha)z, \qquad z^0 = z^H = z, \qquad w^0 = w^L = w^H = 0 \quad (8)$$

The next Proposition summarizes the optimal contract.

**Proposition 2** *The optimal contract when the principal hires an honest auditor satisfies (7)-(8). The auditor is hired always.*

---

incentive problems.

Proof: See Appendix.

This contract displays some rigidity as compared to the contract with a supervisor since the agent's effort is not affected directly by the auditor's report (see Figure 3).

The auditor's report is used to punish the agent (which affects the agent's net compensation) when the principal obtains no favorable information about the agent's type (i.e., any report different from $L$). In particular, for a given informativeness $\alpha$ of the monitor's signal, the punishment threat is enough to deter the type-$\theta_H$ agent from misreporting when the liability bound is high enough, and the principal achieves first-best utility.[16] This result is summarized in the next Corollary.

**Corollary 1** *For any informativeness $\alpha$ of the auditor's signal, the principal achieves first-best utility when the expected punishment bound $z$ is sufficiently high.*

Proof: Fix a value of $\alpha \in (1/2, 1]$. The principal achieves first-best utility if $\alpha > \alpha_2^*$, which is satisfied when $z > z_B$, and $z_B = \theta_L^2 R/p(2\alpha - 1)$. *Q.E.D.*

For intermediate liability levels,

$$\frac{(1-q)^2\theta_L^2 R}{2p(2\alpha-1)[(1-q)+qR]^2} < z < \frac{\theta_L^2 R}{p(2\alpha-1)}$$

the punishment instrument is strong enough to deter the agent from misreporting her type for high signal's accuracy. In those cases, the principal

[16] However, the principal punishes the type-$\theta_L$ agent in the equilibrium of the direct mechanism. When the type-$\theta_H$ agent reports truthfully, she is not audited (see Kofman and Lawarrée (1993)).

can correct the type-$\theta_L$ agent's effort distortion. As Figure 3 shows, the effort distortion is reduced as the signal's accuracy increases. In other words, for a bad signal's quality, the principal prefers to distort effort to induce the type-$\theta_H$ agent to report her true type. As the signal's quality improves, the probability that the principal catches the type-$\theta_H$ agent with the wrong foot is higher, and hence she does not have to distort effort too much to induce the type-$\theta_H$ agent to report the true type. In the limit, as Corollary 1 states, when the signal's quality is sufficiently high (provided that $z > \theta_L^2 R/p$), the expected punishment is deterrent enough and the principal does not distort allocations.

Conversely, the agent receives some positive rent if the auditor's signal is imperfect and the expected punishment is low. For any given accuracy of the signal $\alpha$, this holds if $\alpha_1^* > 1$, i.e.,

$$(1-q)^2\theta_L^2 R > 2pz\left[(1-q)+qR\right]^2$$

In this case the agent exerts the same effort as that without monitor (Figure 3 shows this case, where $e_L$ is displayed with a dashed line).

Figure 4 shows two cases for the principal's utility. The first one corresponds to the principal achieving the first-best utility for high $\alpha$ ($\alpha_2^* < 1$, which is more probable for higher $z$). In the second case the agent obtains positive rents for any signal accuracy ($\alpha_1^* > 1$, which is more probable for lower $z$).

## 4. Optimal Timing

In this section we discuss the principal's monitoring timing decision with an honest monitor. When the principal hires an ex ante monitor (supervisor), she finds profitable to distort allocations and minimize the agent's expected rents depending on the supervisor's signal. On the other hand, the auditor's signal cannot be used to modify allocations, but can be used to punish the agent (which affects effort indirectly).

Therefore, the accuracy of the monitor's information is very useful to the principal when the latter can punish the agent strongly and the punishment can be enforced (i.e., high $z$). As we show in the Appendix (Proof of Theorem 1), there exists a minimum liability bound $\bar{z}$ such that an auditor is optimal for $z > \bar{z}$ for any degree of informativeness of his signal.[17]

However, when the agent is protected by some liability bound or, even under a high liability bound, when the punishment cannot be perfectly enforced (for instance, when there is a long delay in expedition by the courts or uncertainty about the courts' decisions), an ex ante monitor may be optimal. In particular, for values of the punishment instrument $z < \bar{z}$, the principal hires the monitor to supervise the agent when his signal is informative of the agent's type, and to audit the agent when his signal is noisy (See Figure 5). This is so because the reduction of the type-$\theta_H$ agent's rents is higher are reduced when the supervisor observes a "correct" signal (that is, $\sigma = H$, which is more probable when $\alpha$ is high). If the monitor's signal is too noisy, the benefit for the principal from distorting allocations is low, while the expected punishment still has some deterrent value. Hence an auditor is preferred.

In sum, supervising more probably dominates auditing when the expected punishment $z$ is low. The next Theorem summarizes this discussion and emphasizes the importance of monitoring timing.

**Theorem 1** *Suppose that the principal hires an honest monitor.*
• *Auditing is optimal when the principal's punishment instrument is strong and enforceable ($z$ is relatively high) for any informativeness of the signal of the agent's type, or when punishment is weak or difficult to enforce*

[17] In a paper on Law enforcement, Shavell (1993) shows that the availability of harm-based sanctions is an important determinant of the (ex post) legal intervention stage.

*(i.e., $z$ low) as long as the monitor's signal is noisy (low $\alpha$).*
• *Supervising is optimal when the punishment instrument is weak or difficult to enforce (low $z$) as long as the monitor's signal is informative (high $\alpha$).*

Proof: See Appendix.

We explained in Section 2 that effort monitoring and productivity auditing are payoff-equivalent for the principal provided that the signals' accuracy and the cost of obtaining them are the same. Suppose now that the principal is constrained to monitor the agent's type (because, say, it is too costly to monitor effort), so that the relevant problem is the timing of information gathering that is available at the outset. A Corollary of the previous Theorem is:

**Corollary 2** *Under some circumstances the principal can strategically delay the gathering of relevant information, available at the beginning of the game.*

In particular, this is so when the punishment instrument is high enough to deter the agent, so that the deterrent effect offsets the benefits of flexibility from supervising the agent.

Next, we present the principal's response to changes in the environment. Assume that the liability bound $z$ is such that $z < \bar{z}$, so that there exists a cut-off $\alpha^C \in (1/2, 1]$ of the monitor's signal informativeness for which a supervisor is optimal when $\alpha > \alpha^C$ and an auditor is optimal otherwise. Then, we have the following

**Result 1** *Consider as reference the informativeness of the monitor's signal $\alpha$.*
• *The cut-off $\alpha^C$ increases in $z$. Auditing is optimal for a broader range of the informativeness of the monitor's signal as the expected liability bound*

*increases.*

- *The cut-off $\alpha^C$ decreases in $R$. The region of optimality of a supervisor expands out as the adverse selection problem is more severe.*[18]

Proof: See Appendix.

The first result is a direct implication of Theorem 1. The second result is a consequence of the way the principal designs the contracts. The intuition is as follows: the supervisor's signal is useful to reduce the costly agent's rent in the more probable state ($R_{HH} < R_{H0}$), while the auditor's signal is only useful to punish the agent (whose rents are the rents under the No-Monitor contract net of the expected punishment, see $t_h$ in (7)). Hence, for a more disperse distribution of types, which implies a more severe adverse selection problem (captured by a higher $R$), the first contract is more suitable to control the agent's rent because of its flexibility. In words, for a given $\alpha$ and $z$, the decrease in the principal's utility with a supervisor is lower than that in the utility with an auditor, and the intersection occurs at a lower $\alpha$. Therefore, the region of optimality of a supervisor expands out.

From this discussion, we conclude that auditing is optimal when the punishment instrument is strong, and supervising is optimal when the supervisor's information is very informative about the agent's type and a punishment instrument (to be used if an auditor is hired) is weak or difficult to enforce.

These results are consistent with typical organizational structures, in which top-level managers (such as CEOs or top-level managers, who typically are able to respond to fines up to some level) are exposed to audits, while low-level workers (typically with lower incomes or protected by minimum wages) are supervised during production stage. Moreover,

[18] This result applies to the case in which the agent still earns some rent when audited (i.e., parameters are such $\alpha_1^* \geq 1$ in equation (7)), and for $q \leq 1/(1+R)$.

screening is a better instrument when the court's response to complaints is lengthy or very uncertain (and, of course, costlier).

On another line of research, the literature on incentive regulation (Laffont and Tirole (1993) and others) has studied ex ante and ex post regulation separately. The theoretical framework in this paper nests both stages of regulations, formalizing determinants of the timing choice, such as punishment instruments, enforceability and accuracy of information.

Finally, the Law and Economics literature (see Shavell (1993), Kolstad *et al.* (1990)) has studied the optimal regulatory stage of activities that generate externalities with a benevolent regulator. In this paper, we provide a conceptual agency-based framework to explain the implications of regulatory timing on the society welfare. Consider, for example, the case of a hazardous activity with "disastrous" consequences. From Theorem 1, and in accordance with the standard recommendation, the government should put all the efforts in ex ante regulation whenever the liability faced by the injurer is low (as it is the case when the bad outcome involves irremediable consequences).[19] Ex post regulation is recommended when the injurer can be strongly punished.

## 5. Conclusion and Extensions

In this paper we study the case of one-time monitoring in hierarchies. We provide insights on the optimality to the principal of using monitoring timing as a choice variable. Previous literature on the principal-agent model has analyzed both monitoring cases separately, while the literature on Law and Economics has studied the timing of legal intervention. We show that in cases where the timing choice is indeed relevant, the principal faces a trade-off in the monitoring decision (an early report provides

[19] For example, Cohen (1987) , pp. 45-46, shows estimates of very low penalties compared to the environmental damage done by oil spills.

some flexibility in contracting, while a later report can be used to punish the agent), and provide the solution to this trade-off. An auditor is optimal when the principal can expose the agent to severe fines. When the punishment is weak or difficult to enforce, the timing choice depends on the monitor's signal accuracy. In this case, the principal chooses a supervisor when his signal is informative of the agent's type, and an auditor otherwise.

The literature on collusion (Tirole (1986, 1992), Kofman and Lawarrée (1993), Baliga (1999), Faure Grimaud *et al.* (1999)) analyzes the effects of collusion on the principal's contract design, and then it is important to know how the timing choice is affected when both monitor and agent can collude. In a work in progress, we analyze the monitoring timing problem under different collusion environments, related to the degree of information manipulation, which depends on whether information is hard but non-forgeable, hard and forgeable or soft (see Cont (2003)). This issue becomes relevant here, since the incentives to collude are different depending on the monitoring timing. Under ex ante monitoring an agent has incentives to collude with the supervisor to hide or misreport his signal in order to keep the information rents. On the other hand, under ex post monitoring, the agent has an incentive to collude with the auditor to avoid being punishment. An additional trade-off arises here.

In this paper we assume that the principal has to decide between ex ante or ex post monitoring. However, when the principal has access to both monitors who are not related and do not share information, there is no reason for the principal not to choose both monitors if gathering information is costless.[20] An interesting extension is to consider that the

---

[20] There has been some progress along these lines (with a benevolent regulatory agency). For example, Kolstad *et al.* (1990) show that ex ante and ex post regulation may be complements depending on the injurer's uncertainty of his potential liability.

principal may want to hire both monitors, and allow for costly collection of signals, or information sharing and collusion between monitors (in the same lines as those mentioned in the previous paragraph). Under these circumstances, the principal may find optimal to discard one of the monitors.

Corollary 2 introduces a more general question of strategic timing. In our framework, the principal optimally chooses to delay the monitoring to later stages when she hires an auditor (in particular, when effort monitoring is not available). This result is a case of strategic timing, since under some circumstances the principal delays the gathering of relevant information that is available in the beginning of the game. Along these lines, there is a broader question that has to do with strategic contracting. Theoretical models assume that grand contracts are designed at the beginning of a general game. By constraining the set of decisions at some period of the game, the parties may get some benefit at later stages. For example, in a paper on collusion and delegation, Laffont and Martimort (1998) show that a principal finds profitable to delegate to the supervisor the direct contracting with an agent.

Finally, a new line of research is under project. The intuition in this paper extends to the stage of control (if necessary) of competing firms, where regulation (interpreted as ex ante control) and competition policy (which triggers after some incorrect behavior is detected) interact with each other. This discussion has gained room in new markets (such as the Internet) where there are many providers (multiple agents), countries (multiple principals) and regulators involved. For example, ex post control (competition policy) may be optimal when it is difficult to coordinate or organize interaction among the agents (see Laffont and Tirole (2000)).

## Appendix

**Proof of Proposition 2:** When the principal hires an honest and costless auditor, she pays $w^r = 0$ for $r \in \{0, L, H\}$. Constraint IC(L) is non-binding when the others are satisfied. The Lagrangian to problem (6) is

$$\begin{aligned}
\mathcal{L} &= V - \bar{\theta} + q\{\theta_H e_h - t_h\} + (1-q)\left\{\theta_L e_l - t_l + \left[(1-p)z^0 + p(1-\alpha)z^H\right]\right\} \\
&+ \lambda_1\left\{t_l - \left[(1-p)z^0 + p(1-\alpha)z^H\right] - \frac{e_l^2}{2}\right\} + \lambda_2\left\{t_h - \frac{e_h^2}{2}\right\} \\
&+ \lambda_3\left\{t_h - \frac{e_h^2}{2} - t_l + \left[(1-p)z^0 + p\alpha z^H\right] + \frac{e_l^2\Delta\theta}{2}\right\}
\end{aligned}$$

The Kuhn-Tucker conditions are

$$\begin{array}{lll}
\mathcal{L}_{e_h} = q\theta_H - (\lambda_2 + \lambda_3)\, e_h \leq 0, & e_h \geq 0, & \mathcal{L}_{e_h} e_h = 0 \\
\mathcal{L}_{e_l} = (1-q)\theta_L - (\lambda_1 - \lambda_3\Delta\theta)\, e_l \leq 0, & e_l \geq 0, & \mathcal{L}_{e_l} e_l = 0 \\
\mathcal{L}_{t_h} = -q + \lambda_2 + \lambda_3 \leq 0, & t_h \geq 0, & \mathcal{L}_{t_h} t_h = 0 \\
\mathcal{L}_{t_l} = -(1-q) + \lambda_1 - \lambda_3 \leq 0, & t_l \geq 0, & \mathcal{L}_{t_l} t_l = 0
\end{array}$$

$$\begin{aligned}
\mathcal{L}_{z^0} &= (1-q) - \lambda_1 + \lambda_3 = 0; \text{ if } \mathcal{L}_{z^0} < 0,\ z^0 = 0; \text{ if } \mathcal{L}_{z^0} > 0,\ z^0 = z \\
\mathcal{L}_{z^H} &= (1-q)(1-\alpha) - \lambda_1(1-\alpha) + \lambda_3\alpha = 0 \\
&\quad \text{if } \mathcal{L}_{z^H} < 0,\ z^H = 0; \text{ if } \mathcal{L}_{z^H} > 0,\ z^H = z
\end{aligned}$$

together with the participation, incentive compatibility and liability constraints. The solution involves positive $e_l$ and $e_h$. From the participation constraints, $t_l$ and $t_h$ are both positive. Then $\mathcal{L}_{e_h} = \mathcal{L}_{t_h} = 0$, which implies that the type-$\theta_H$ agent exerts first-best effort $e_h = \theta_H$. Also, $\mathcal{L}_{e_l} = \mathcal{L}_{t_l} = 0$ and

$$(1-q)\theta_L = [\lambda_1 - \lambda_3\Delta\theta]\, e_l \tag{9}$$

$$(1-q) = \lambda_1 - \lambda_3 \tag{10}$$

Using (10), $\mathcal{L}_{z^0} = 0$ and hence $z^0 = z$ without loss of generality. Also, $\mathcal{L}_{z^H} = \lambda_3(2\alpha - 1) \geq 0$ and then $z^H = z$. This is a maximum deterrence result (see Baron and Besanko (1984)). Next we consider the three possible cases for the relationship between IC(H) and IR(H) (from (5)): either of them or both of them are binding.

Case 1: $\lambda_2 = 0$ and $\lambda_3 = q$. IR(H) is non-binding and IC(H) is binding. Using equations (9) and (10), $e_l = (1-q)\theta_L / [(1-q) + qR]$. Using the IR and IC constraints, $t_l = e_l^2/2 + (1-p\alpha)z$, $t_h = \theta_H^2/2 + e_l^2 R/2 - p(2\alpha - 1)z$. This is the solution if IR(H) is non-binding (i.e., $e_l^2 R/2 > p(2\alpha - 1)z$), which is satisfied for $\alpha < \alpha_1^*$, where

$$\alpha_1^* = \frac{1}{2} + \frac{(1-q)^2 \theta_L^2 R}{4pz\left[(1-q) + qR\right]^2} \tag{11}$$

Case 2: $\lambda_2 = q$ and $\lambda_3 = 0$. IR(H) is binding and IC(H) is non-binding. Using (10) in (9) we have $e_l = \theta_L$. From the IC and IR constraints, $t_l = \theta_L^2/2 + (1-p\alpha)z$, $t_h = \theta_H^2/2$, and IC(H) must hold as inequality (i.e., $\theta_L^2 R/2 < p(2\alpha - 1)z$), which is satisfied for $\alpha > \alpha_2^*$, where

$$\alpha_2^* = \frac{1}{2} + \frac{\theta_L^2 R}{2pz} \tag{12}$$

Case 3: Both $\lambda_2$ and $\lambda_3$ are non-negative (and both less than or equal to $q$). Using IR and IC constraints, the type-$\theta_H$ agent's rent must be zero, i.e., $e_l^2 R/2 = p(2\alpha - 1)z$. Hence, $e_l = \sqrt{2p(2\alpha - 1)z/R}$. Using equations (9) and (10), $\lambda_3 = (1-q)\left(\theta_L/e_l - 1\right)/R$. The agent compensation is $t_l = e_l^2/2 + (1-p\alpha)z$ and $t_h = \theta_H^2/2$. The Lagrangian is concave since it is linear in compensations and punishments, there are no cross terms among them and efforts, and the second derivative with respect to effort is negative. The summary of effort and compensations is presented in equations (7)-(8). *Q.E.D.*

**Proof of Theorem 1:** Define $U_H = V - \bar{\theta} + q\frac{\theta_H^2}{2}$. On the one hand, the principal's utility with a supervisor is

$$EU_P^S = U_H + \frac{(1-q)^2\theta_L^2}{2}\left\{\frac{p\alpha^2}{(1-q)\alpha + q(1-\alpha)R} + \frac{p(1-\alpha)^2}{(1-q)(1-\alpha) + q\alpha R} + \frac{1-p}{(1-q)+qR}\right\} \tag{13}$$

where $EU_P^S$ is increasing and convex in $\alpha$. On the other hand, the principal's utility with an auditor is

$$EU_P^A = \begin{cases} U_H + \frac{(1-q)^2\theta_L^2}{2[(1-q)+qR]} + qp(2\alpha-1)z = EU_P^{NM} + qp(2\alpha-1)z & \text{if } \alpha < \alpha_1^* \\ U_H + (1-q)\left\{\sqrt{\frac{2p(2\alpha-1)z}{R}}\theta_L - \frac{p(2\alpha-1)z}{R}\right\} & \text{if } \alpha_1^* \leq \alpha \leq \alpha_2^* \\ EU_P^{FB} & \text{if } \alpha > \alpha_2^* \end{cases} \tag{14}$$

where $\alpha_1^*$ is from (11), $\alpha_2^*$ is from (12), $EU_P^{NM}$ is equation (2) and $EU_P^{FB}$ is equation (1). The first part of this utility function is increasing and linear in $\alpha$, the second part is concave and the last part is constant. By construction, the principal's utility is continuous in all parameters.

If $z = 0$, $EU_P^A = EU_P^{NM}$, while $EU_P^S > EU_P^{NM}$. On the other hand, when $z$ is very high, $EU_P^A > EU_P^S$ for all $\alpha$. So, for intermediate values of $z$, $EU_P^S$ intersects $EU_P^A$ at some cut-off value $\alpha^C$ such that $EU_P^A > EU_P^S$ for $\alpha < \alpha^C$, and $EU_P^A < EU_P^S$ for $\alpha > \alpha^C$. This intersection must happen when effort is distorted or the agent receives some positive rent when audited (that is, cases 1 or 2 in the Proof of Proposition 2).

When the intersection exists, if $z$ increases given a value of $\alpha$, $EU_P^A$ increases (until it reaches $EU_P^{FB}$) while $EU_P^S$ remains the same. Therefore the cut-off $\alpha^C$ increases in $z$ (implying that auditing is optimal for a broader range of the signal informativeness $(1/2, \alpha^C]$).[21] In particular, there exists a critical liability bound $\bar{z}$ (which corresponds to the punish-

[21] We prove this result assuming that the intersection exists. The more general

ment $z$ for which $\alpha^C = 1$) such that $EU_P^A > EU_P^S$ for all $\alpha$ when $z > \bar{z}$. *Q.E.D.*

**Proof of Result 1:** The region of optimality of a supervisor expands out as adverse selection is more severe (i.e., $R$ increases, caused by an increase in $\theta_H$). The proof is done assuming that parameters are such that $\alpha_1^* \geq 1$ (from equation (11)) and $q \leq 1/(1+R)$. We show that $EU_P^S$ decreases less than $EU_P^A$ does as $R$ increases. When this is the case, the new cut-off $\alpha^C$ corresponds to a lower value of $\alpha$ (keeping all the other parameters fixed). From equations (13) and (14), eliminate the common parts $U_H$ and $(1-q)^2\theta_L^2/2$ to get that

$$\frac{\partial}{\partial R}\left(\frac{p\alpha^2}{(1-q)\alpha+q(1-\alpha)R}+\frac{p(1-\alpha)^2}{(1-q)(1-\alpha)+q\alpha R}+\frac{1-p}{(1-q)+qR}\right) > \\ > \frac{\partial}{\partial R}\left(\frac{1}{[(1-q)+qR]}\right) \tag{15}$$

which, after several steps (omitted for convenience), simplifies to

$$\alpha(1-\alpha)\left\{\left[2\left(\alpha^2+(1-\alpha)^2\right)-4\alpha(1-\alpha)\right](1-q)^3+\left[\alpha(1-\alpha)-\alpha^3-(1-\alpha)^3\right]q^3R^3\right\} \\ +\left[\alpha^4+(1-\alpha)^4+2\alpha^2(1-\alpha)^2-\alpha(1-\alpha)\right](1-q)^2qR > 0$$

The first term within $\{.\}$ is non-negative and the second term is non-positive. The last term is non-positive. But note that $(1-q)^2qR \geq q^3R^3$ and $(1-q)^3 \geq q^3R^3$ when $q \leq 1/(1+R)$, and that the sum of all brackets simplifies to $(1-2\alpha)^2 > 0$, for $\alpha > 1/2$. Hence, inequality (15) is satisfied. *Q.E.D.*

---

result is that the cut-off $\alpha^C$ is non-decreasing in $z$, which corresponds to including the case of no intersection. This generalization holds throughout the proof.

## References

Baliga, Sandeep, 1999, Monitoring and Collusion with "Soft" Information, *Journal of Law, Economics and Organization*, vol. 15, No. 2, pp. 434-440.

Baron, David and David Besanko, 1984, Regulation, Asymmetric Information and Auditing, *RAND Journal of Economics*, vol. 15, No. 4, pp. 447-470.

Cohen, Mark, 1987, Optimal Enforcement Strategy to Prevent Oil Spills: An Application of a Principal-Agent Model with "Moral Hazard", *Journal of Law and Economics*, vol. 30, No. 1, pp. 23-51.

Cont, Walter, 2001, *Essays on Contract Design*, Ph.D. Dissertation, University of California, Los Angeles.

Cont, Walter, 2003, Optimal Monitoring Timing under Collusion, mimeo.

Faure-Grimaud Antonie, Jean-Jacques Laffont and David Martimort, 1999, The Endogenous Transaction Costs of Delegated Auditing, *European Economic Review*, vol. 43, No. 4-6, pp. 1039-1048.

Kofman, Fred and Jackes Lawarrée, 1993, Collusion in Hierarchical Agency, *Econometrica*, vol. 61, No. 3, pp. 629-656.

Kolstad, Charles, Thomas Ulen and Gary Johnson, 1990, Ex Post Liability vs. Ex Ante Safety Regulation: Substitutes or Complements?, *American Economic Review*, vol. 80, No. 4, pp. 888-901.

Laffont, Jean-Jacques and David Martimort, 1998, Collusion and Delegation, *RAND Journal of Economics*, vol. 29, No. 2, pp. 280-305.

Laffont, Jean-Jacques and Jean Tirole, 1993, *A Theory of Incentives in Procurement and Regulation*, The MIT Press.

Laffont, Jean-Jacques and Jean Tirole, 2000, *Competition in Telecommunications*, The MIT Press.

Shavell, Steven, 1993, The Optimal Structure of Law Enforcement, *Journal of Law and Economics*, vol. 36, No. 1-2, pp. 255-287.

Tirole, Jean, 1986, Hierarchies and Bureaucracies: On the Role of Collusion in Organizations, *Journal of Law, Economics and Organization*, vol. 2, No. 2, pp. 181-214.
Tirole, Jean, 1992, Collusion and the Theory of Organizations. In J.J. Laffont (ed.), *Advances in Economic Theory*, vol. 2, Cambridge University Press, pp. 151-205.

Figure 1: Effort as a Function of Supervisor's Signal Accuracy ($\alpha$).

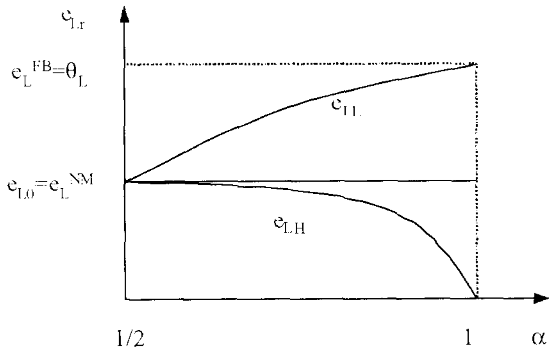



Figure 2: Principal's Expected Utility as a Function of Supervisor's Signal Accuracy ($\alpha$).

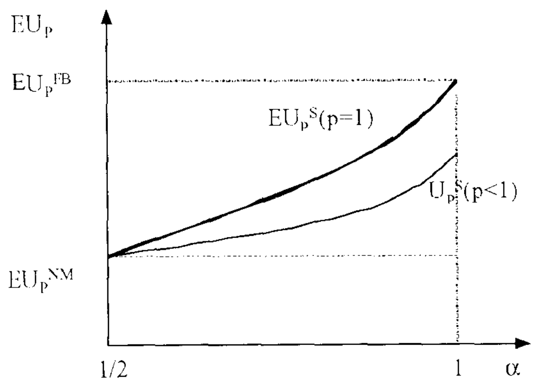


Figure 3: Effort as a Function of Auditor's Signal Accuracy ($\alpha$).

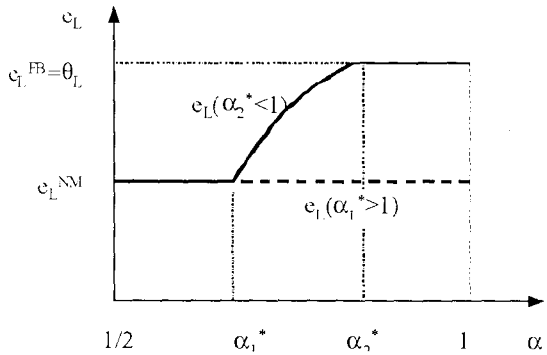



Figure 4: Principal's Expected Utility as a Function of Auditor's Signal Accuracy ($\alpha$).

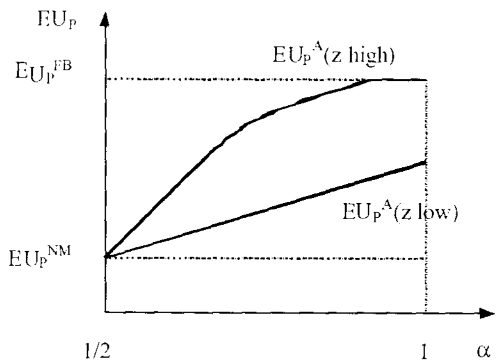
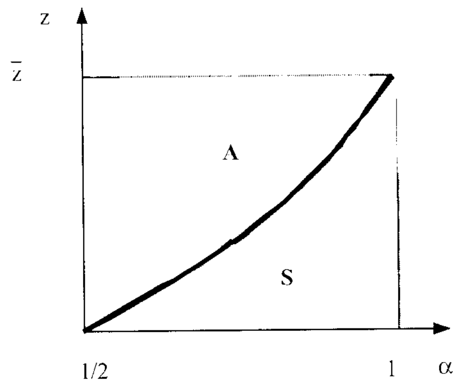


Figure 5: Optimal Monitoring Timing.



## ON MONITORING TIMING IN HIERARCHIES
WALTER CONT


## RESUMEN

Clasificación JEL: D82, L23.

Este paper muestra que en un modelo principal-monitor-agente el principal enfrenta un *trade-off* al elegir el momento en el cual monitorear al agente. Una señal obtenida por un monitor *ex ante* (supervisor) provee flexibilidad en la contratación con el agente (dado que se puede contratar el producto y las remuneraciones basados en esta señal), mientras que una señal obtenida por un monitor *ex post* (auditor) puede ser utilizada para penalizar al agente. Este trabajo muestra que auditar al agente es óptimo para el principal si (i) éste puede implementar y hacer cumplir esquemas de penalización o si (ii) los instrumentos de penalización no son suficientemente disuasores siempre y cuando la señal del monitor sea poco informativa de la característica del agente. De otra manera, resulta óptimo supervisar al agente.


## ON MONITORING TIMING IN HIERARCHIES
WALTER CONT


## SUMMARY

JEL Classification: D82, L23.

In a principal-monitor-agent model we show that the principal's choice of the timing to monitor the agent presents a trade-off. On the one hand, a signal from an *ex ante* monitor (supervisor) provides flexibility in contracting (since both output and wages can be contracted on this signal). On the other hand, a signal from an *ex post* monitor (auditor) can be used to punish the agent. Auditing is optimal when (i) strong punishment schemes can be implemented and enforced by courts or (ii) when punishment instruments are not expected to be strong and the monitor's signal is noisy. Supervising is optimal otherwise.