Optimal Auditing in Hierarchical Relationships

Anke S. Kessler*

Revised version: October 2002

Abstract

This paper studies an agency framework in which a principal hires an auditor to conduct audits on a productivity parameter that is private information of a manager. Auditing technologies are distinguished according to the quantity (frequency) and the quality (accuracy) of the information they deliver. We show that the frequency of audits is irrelevant if the auditor is either honest or colludes with the manger but auditing evidence can only be concealed and not forged. In either case, the first-best can be achieved if the audit is sufficiently precise even though unbounded punishments are not feasible. Only if auditing evidence can be falsified, the principal benefits both from the frequency and the accuracy of the auditor's observations. The findings therefore indicate that firms should opt for accurate rather than frequent audits under a wide range of circumstances.(JEL: D82, L23)

1 Introduction

In the standard agency model with adverse selection, the principal designs a contract for an agent who is privately informed about a parameter relevant for the transaction. It is by now well understood that the outcome of this relationship will be inefficient due to the prevailing informational asymmetry and the principal's desire to extract informational rents. In order to reduce the prevailing incentive problem, therefore,

*University of Bonn and CEPR. I wish to thank Chong-en Bai, Matthias Kräkel, Christoph Lülfesmann, Rüdiger Schils, Patrick Schmitz, Urs Schweizer, Dirk Sliwka, and various participants of the EARIE 2001 conference for their suggestions and discussions. I am also indebted to two anonymous referees whose comments have helped to improve the paper. Remaining errors are my own.

the principal has an interest to conduct audits. Examples of models that incorporate this possibility include ?, ?, ?, and ?, among others. Yet, the value of audits is diminished if the principal has to employ a third party for this task: the potential collusion between auditor and agent generally reduces the welfare of the principal as has been shown by ??, ??, and ?.[1] An assumption common to these contributions is that they take the information technology the principal has access to as exogenously given.[2] In practice, however, the principal may have various information technologies with different characteristics at her disposal. A company's headquarter, for example, can decide both on how often auditors are send to its subsidiaries and on how intense those audits should be. Similarly, by choosing a particular monitoring procedure an employer may either improve the quality of an evaluation on the employee's performance or, alternatively, increase the frequency of inspection rounds.

The purpose of this paper is to study the virtues of two different aspects of an information technology that have already been illustrated by the above examples. The first characteristic is the frequency of making an observation, i.e., how often evaluations or audits are carried out. The second property is the accuracy of the observation made, i.e., how precisely the outcome of an evaluation or audit reflects the true value of the variable which is to be observed. The analysis is conducted in a simple agency framework where a principal employs an auditor to monitor a manager who carries out a productive task. The manager is privately informed about a technology parameter and the effort he exerts to produce output. The outcome of the audits is used as a control instrument by the principal to improve the manager's incentives.[3] Unbounded punishments are ruled out by assuming that the manager is wealth constrained and can therefore be held liable only in a limited way.[4] I first consider the case where the auditor cannot collude with the manager, which is equivalent in the model to a situation in which the principal undertakes the audit herself. Next, manager and auditor are allowed to collude through writing unobservable side contracts. Here, I will distin-

[1] A different view of hierarchies is taken in ? and ? who show that employing the supervisor can enhance the principal's commitment to the contract, thereby increasing her expected return from the relationship.

[2] An exception is the paper by ?, which is briefly discussed below.

[3] Hence, audits or performance evaluations do not serve to obtain information on the agent' talents that would make him (better) suited for a particular job (rather than another).

[4] If arbitrarily high penalties are allowed for, even weakly correlated ex post signals are sufficient to eliminate the agent's rent and ensure the first-best outcome, provided the parties are risk neutral (see, e.g., ? or ?).

guish between the case where the auditor/manager coalition can only conceal (hard) evidence and a situation where the auditor's report may contain fake (soft) evidence. For each of these three possibilities, I investigate how – ignoring relative costs – the principal's return from the relationship varies with the frequency (quantity) and the accuracy (quality) of the information collected by the auditor. The analysis indicates that only the quality of information matters under a wide range of circumstances: quantitative aspects are shown to be irrelevant if either the auditor is honest or his report is hard information. The principal benefits from frequent observations only if the auditor is collusive and his signal is non-verifiable. These results imply that there is often no trade-off involved in choosing between accurate rather than frequent audits or performance evaluations, contrary to what is perhaps at first glance suggested.

Up to now, there have been few attempts to draw conclusions on how the principal values different information systems in an agency framework with adverse selection.[5] **?** analyze a competitive labor market framework in which firms can test for a worker's ability. Testing is costly and both the accuracy of the test and the percentage of workers tested are choice variables in the firms' decision problem. It is shown that workers are tested with strictly positive probability in equilibrium. Only if the testing technology is very accurate and unbounded punishments can be used, the full information equilibrium can be approximated. **?** investigate a principal-agent model with a continuum of types and limited liability. The authors demonstrate that the principal's welfare and efficiency increase in the responsiveness of an additional ex post signal to the agent's type, which corresponds to our finding that more precise audits are preferred by the principal.[6] Contrary to the present paper, however, they neither consider collusion nor the possibility that no informative signal is received, i.e., the probability (frequency) of making a valuable observation is equal to one. **?** study the choice between input and output monitoring: the principal prefers to monitor the inputs if she herself is the residual claimant whereas output monitoring is superior if the agent is the residual claimant. Finally, **?** analyzes a principal–agent relationship in which a manager searches for investment opportunities. He compares incentive systems that differ in

[5] In contrast, the ranking of information structures in principal-agent models with moral hazard has received more attention in the literature, starting with **?**'s seminal result on sufficient statistics. See also **?** for a comparison based on **??**'s theory, **?** who develops a mean preserving spread criterion that is applicable to a broader class of information systems, and **?** who consider a risk-neutral agent protected by limited liability.

[6] **?** analyzes a cost function for the accuracy of an audit and shows that the optimal accuracy depends on maximum punishments.

the (accounting) information they use and shows that a performance measure based on residual income economizes on agency costs relative to performance measure based on realized cash flows.

The remainder of the paper is organized as follows. Section 2.1. introduces the basic framework. Section 2.2 derives the optimal contract and its properties under collusion-free and collusive monitoring with hard and soft information, respectively. A final section concludes.

2 The Model

2.1 The Basic Framework

Consider the following simple agency model with three risk neutral parties:[7] a principal (P) hires a manager (M) to carry out a productive task and employs an auditor (A) to monitor production. The manager is privately informed about a random productivity parameter θ and the effort e he exerts to produces output x, which is given by

$$x = \theta + e.$$

Productivity can either be high or low, $\theta_h > \theta_l$, with q as the commonly known ex ante probability that it is high, $\theta = \theta_h$. Exerting effort is costly to the manager. His disutility of effort is represented by an increasing and strictly convex function $\psi(e)$ which satisfies $\psi(0) = \psi'(0) = 0$. To ensure that the principal's problem of designing an optimal contract is globally concave and has a unique solution, we assume in what follows that $q \leq \frac{1}{2}$ and $\psi'''(\cdot) \geq 0$ (see also, e.g., **?**). Denoting by t the transfer that the principal pays to the manager for his work, his utility is $u_M = t - \psi(e)$.

The auditor can conduct audits and has access to the following information technology: with probability $p \in (0, 1]$ an informative signal $s \in \{s_h, s_l\}$ is received which is imperfectly correlated with the manager's productivity θ. With probability $1 - p$, the signal $s = \emptyset$ is received which contains no information, i.e., the auditor observes 'nothing'. Conditional upon receiving an informative signal, the probability that the signal is correct is denoted by $\alpha \in (\frac{1}{2}, 1]$.[8] In the remainder, I will refer to p as the *frequency*

[7] The model is a slightly modified version of **?** who focuses on the cost of collusion and does not consider the value of different auditing technologies.

[8] Restricting the probability α to the interval $(\frac{1}{2}, 1]$ is by convention only and ensures that the

and to α as the precision or *accuracy* of the information generated through audits.[9] The main objective of this paper is to analyze the effect of these two dimensions of the information system, (α, p), on the optimal contract and the principal's utility. The auditor receives a wage w for his services and his utility is $u_A = w$. For simplicity only, the respective reservation utilities of the auditor and the manager are normalized to zero. Throughout the analysis, M and A are assumed to be wealth constrained so that transfers must be non-negative and unbounded punishments are unfeasible. The principal owns the production and designs contracts for the manager and the auditor which can be conditioned on commonly observable variables such as output x and the auditor's report r.[10] Her utility is $u_P = \theta + e - t - w$.

The timing is as follows. Before contracting takes place, nature chooses the state of productivity θ and the manager learns his type. Next, the principal proposes contracts to the manager and the auditor which both can accept or reject. If contracts have been accepted, the manager exerts e and the auditor observes $s \in \{\emptyset, s_h, s_l\}$. The outcome of the audit is revealed to the manager. Then, output is realized and the auditor (if called for) reports the result of his investigation to the principal, possibly after first conferring with the manager. Finally, the principal pays the transfers in accordance with the contracts.

signal is correct on average. For parameter values α in the range $[0, \frac{1}{2})$, the signal would be incorrect on average but still valuable. Clearly, what matters for its informational content is $|\alpha - \frac{1}{2}|$.

[9]Although the auditor's activity generates a signal on θ and thus better reflects a situation where he conducts an 'audit', one could equally well assume that he is instead a supervisor in charge of a 'performance evaluation' and observes a noisy signal on e. Since x is a linear function of e and θ, it is straightforward to show that both possibilities are equivalent provided everything else (in particular, the informational content of an observation) is held equal. The assumption that $\alpha = Prob\{s = s_l|\theta_l\} = Prob\{s = s_h|\theta_h\}$ is equally innocuous [see also **?**]. If those probabilities differed in magnitude, the relevant measure of precision is given by the ratio of $Prob\{s = s_l|\theta_l\}$ to $Prob\{s = s_h|\theta_l\}$.

[10]The assumption on r being commonly observable is imposed by most of the literature on monitoring. As we will see below, it may be of interest for the principal to claim that A has observed 'nothing' rather than the true value of s. It is therefore necessary to assume that the principal cannot hide monitoring evidence. See also footnote 19

Benchmark Contracts

First I examine two benchmark contracts starting with the first-best solution and following with the standard problem under asymmetric information. To analyze the first case, suppose θ is publicly observable. The principal maximizes her utility, $u_p = \theta_i + e_i - t_i$, subject to the manager's participation constraint $t_i - \psi(e_i) \geq 0$, $i \in \{l, h\}$. Thus, the first-best contract is characterized by $\psi'(e^{FB}) = 1$ and $t^{FB} = \psi(e^{FB})$, independent of the manager's productivity θ.

Next, suppose θ is private information of the manager but $p \equiv 0$. Since there is no role for the auditor, the principal's problem is now a standard adverse selection problem (see, e.g., **?**). Supposing that the principal can commit herself not to renegotiate the contract, the Revelation Principle states that the analysis can be confined to a direct mechanism which guarantees truthful revelation and which prescribes an output $x(\hat{\theta})$ and a transfer $t(\hat{\theta})$ contingent upon the manager's announcement of his type, $\hat{\theta}$. For notational simplicity, let (t_i, x_i), $i = h, l$ be the contract for a manager who claims to be of type θ_i, $i \in \{l, h\}$. Using $e_i = x_i - \theta_i$, the manager's participation are

$$(\text{PC}_i) \qquad t_i - \psi(e_i) \geq 0, \qquad i \in \{l, h\}$$

and incentive-compatibility requires[11]

$$(\text{IC}_l) \qquad t_l - \psi(e_l) \geq t_j - \psi(e_h + \Delta\theta)$$

$$(\text{IC}_h) \qquad t_h - \psi(e_h) \geq t_l - \psi(e_l - \Delta\theta),$$

where $\Delta\theta = \theta_h - \theta_l > 0$. Maximizing the principal's expected revenue, $u_P = q(\theta_h + e_h - t_h) + (1-q)(\theta_l + e_l - t_l)$, subject to the (PC_l) and (IC_h) constraints (the other two constraints do not bind at the optimum and can be ignored) yields

$$\begin{aligned} &1 = \psi'(e_h^{SB}) \quad \Rightarrow \quad e_h^{SB} = e^{FB} \quad \text{and} \\ &1 = \psi'(e_l^{SB}) + \frac{q}{1-q}\left[\psi'(e_l^{SB}) - \psi'(e_l^{SB} - \Delta\theta)\right] \quad \Rightarrow \quad e_l^{SB} < e^{FB}. \end{aligned}$$

as the first-order conditions for the second best effort levels. The corresponding transfers are $t_l^{SB} = \psi(e_l^{SB})$ and $t_h^{SB} = \psi(e^{FB}) + \left[\psi(e_l^{SB}) - \psi(e_l^{SB} - \Delta\theta)\right]$. Under the optimal contract, the low-productivity manager exerts a level of effort which is distorted downward and obtains his reservation utility. In contrast, the effort of the high-productivity type is efficient and he earns an informational rent equal to $\Phi(e_l^{SB}) = \psi(e_l^{SB}) - \psi(e_l^{SB} - \Delta\theta) > 0$.

[11]It is assumed here that the principal always wants to employ the manager. Also, the (IC_h) constraint implicitly requires $e_l \geq \Delta\theta$.

2.2 *Optimal Contracts and Auditing*

Going back to the initial model, recall from the previous section that p is the frequency of making a 'valuable' observation, i.e., the probability that a signal $s \in \{s_h, s_l\}$ is received. There are two interpretations of p that are consistent with the present analysis. On the one hand, p may simply be the frequency of audits (which then generate an informative signal with certainty). On the other hand, p could be the probability of A receiving an informative signal conditional upon the audit having been conducted (with certainty). As long as the information technology is such that the costs are the same in both cases, I demonstrate in the Appendix that the two *a priori* distinct possibilities are formally equivalent. For ease of presentation, the exposition is in what follows confined to the second interpretation. Specifically, I will assume without loss of generality that audits are conducted with probability one.[12] The parameter p thus becomes the probability of A observing s_h or s_l, given that the audit has taken place with certainty. The reader should keep in mind, however, that all results are equally valid for the alternative interpretation where p is the actual probability (frequency) of audits.

Collusion Free Auditing:

We begin by studying the case where the auditor always reports his signal truthfully. This problem is formally similar to a situation in which the principal undertakes the audit herself. Before embarking on the formal analysis, some preliminary considerations are useful. First, note from the previous section that a low-productivity manager has no incentive to claim that he is of the high-productivity type under the no-auditor second best contract (and under the first-best contract). Hence, anticipating that the respective incentive constraint will not be binding, there is no need to relax the constraint through audits. If a the manager has (truthfully) reported that he is of the high-productivity type, I can therefore without loss of generality assume that no report is requested, although an audit may have been conducted. If output is low, though, a high-productivity manager may have misrepresented his true type (he may have shirked) and the principal will always request a report. Let t_l^r and w^r, respectively, denote the compensations for the θ_l-type manager and the auditor, conditional on the report $r \in \{0, l, h, \}$. Under the information technology specified above, the manager's

[12] Recall that it is not costly to employ the auditor and note that the principal can always choose to ignore A's report.

participation constraints are

$$(PC_l) \qquad p[\alpha t_l^l + (1-\alpha)t_l^h] + (1-p)t_l^0 - \psi(e_l) \geq 0,$$
$$(PC_h) \qquad t_h - \psi(e_h) \geq 0.$$

and the incentive compatibility constraints read

$$(\mathrm{IC}_l) \qquad p[\alpha t_l^l + (1-\alpha)t_l^h] + (1-p)t_l^0 - \psi(e_l) \geq t_h - \psi(e_h + \Delta\theta)$$
$$(\mathrm{IC}_h) \qquad t_h - \psi(e_h) \geq p[(1-\alpha)t_l^l + \alpha t_l^h] + (1-p)t_l^0 - \psi(e_l - \Delta\theta).$$

Next, note that because the signal is informative on average a high-productivity manager who shirks faces a higher probability of a report $r = h$ than a manager with low productivity. To satisfy the corresponding incentive constraint (IC_h) – and thus ensure equilibrium truth-telling – at the lowest cost to herself, the principal therefore optimally sets the transfer to the agent for this report, t_l^h, as low as possible (Maximum Deterrence Principle, see **?**). Recalling that limited liability requires transfers to be non-negative, we thus must have $t_l^h = 0$ under the optimal contract. Finally, because the auditor is honest, there is no need for his wage to vary with reports so that $w_l^r \equiv 0$ without loss of generality. The principal's problem is now to

$$(1) \qquad \text{maximize} \quad q\{\theta_h + e_h - t_h\} + (1-q)\{\theta_l + e_l - p\alpha t_l^l - (1-p)t_l^0\}$$
$$\text{subject to} \qquad (\mathrm{PC}_l),(\mathrm{PC}_h),(\mathrm{IC}_h) \quad \text{and} \quad t_h, t_l^l, t_l^0 \geq 0,$$

where I have again omitted the (IC_l) constraint which will be satisfied by the solution to (1). The results of the formal analysis in the Appendix are gathered in the following proposition.

Proposition 1 Consider collusion free auditing and define $\alpha^* \equiv \psi(e^{FB})/[\psi(e^{FB}) + \psi(e^{FB} - \Delta\theta)]$. *The optimal contract is characterized by a first-best effort of the high-productivity type,* $e_h^* = e^{FB}$, *and an effort of the low-productivity type that strictly exceeds his second-best effort,* $e_l^* > e_l^{SB}$.
Moreover, for any $p \in (0, 1]$ *the utility of the principal is independent of the frequency* p *of the signal. It is strictly increasing in the accuracy* α *of the signal for values* $\alpha < \alpha^*$ *and the first best is achieved for values* $\alpha \geq \alpha^*$.

The first part of the proposition simply states that additional information through audits increases the power of the incentive scheme for the low-productivity manager.

While this conclusions is standard in the literature,[13] the second part is more surprising. Under the optimal contract, the principal's utility is not affected by the probability that the auditor actually observes a signal on θ. It solely depends on the accuracy of A's information, even though the *ex ante* probability that A detects shirking by M increases with p. Besides, the principal can implement a first-best solution if α is above some critical level, regardless of p. The result implies that the principal strictly prefers an information system with accurate rather than frequent observations, although both dimensions would be valuable for a decision maker according to, e.g., **??**'s notion of informativeness.[14]

The intuition for this finding can most easily be obtained by taking a closer look at the informational rent of a high-productivity manager, which I can by substituting for t_l^0 from (PC_l) into (IC_h) write as

$$\begin{aligned} \Phi(e_l) &= \psi(e_l) - \psi(e_l - \Delta\theta) - p(2\alpha - 1)(t_l^l - t_l^h) \\ &= \Phi^{NA}(e_l) - p(2\alpha - 1)(t_l^l - t_l^h), \end{aligned} \tag{2}$$

where $\Phi^{NA}(e_l)$ is the corresponding rent in the no-auditor second best case ($p = 0$). The last term on the right-hand side of (2) is the expected penalty if the manager is not truthful and $p > 0$. It reflects the fact that a dishonest high-productivity manager faces a different probability distribution of the signal than a honest manager with low productivity. In particular, the auditor will detect shirking of the high-productivity manager with probability $p\alpha$ while a low-productivity manager is punished only with probability $p(1 - \alpha)$. As one would expect, the term therefore strictly decreases in α *and* p. Importantly, however, the penalty also increases in $t_l^l - t_l^h$, which is the difference between the transfer t_l^l that the manager receives if the signal (falsely) indicates that he was honest and the transfer when the signal (correctly) indicates non-compliance, t_l^h. The principal thus optimally sets the latter payment as low as possible, i.e., $t_l^h = 0$. At the same time, t_l^l should be as high as possible: the penalty of not paying the agreed transfer t_l^l if $s = s_h$ is more severe, the larger t_l^l. As can be seen from (PC_l), this in turn implies that t_l^0 should be as small as possible, i.e., $t_l^0 = 0$. In other words, under the optimal contract the manager is rewarded only if an audit results in confirmatory

[13]Similar results have been obtained in, e.g., **?** and **?**.

[14]To see this consider an information structure where the signal $s \in \{s_l, s_h\}$ is received with certainty. Now suppose s 'garbled' in the following way: with probability $p < 1$, the signal observed is $s' = s$. With probability $1 - p$, $s' = 0$. The distribution of s' depends on θ only via s which is therefore more informative than s' in the sense of Blackwell.

evidence and is punished otherwise.[15]

Setting $t_l^h = t_l^0 = 0$, the manager's participation requires $p\alpha t_l^l = \psi(e_l)$. As a consequence, if the principal has to raise t_l^l in response to a drop in p in order to keep the expected compensation constant [(PC$_l$) satisfied], the expected punishment does not vary [see (2)] and the manager's rent is thus independent of p. Formally, the parameter p cancels out once we solve for t_l^l from (PC$_l$) and substitute the value into (IC$_h$). Moreover, the expected payoff of a manager who shirked [the right-hand side of the transformed (IC$_h$)] eventually becomes negative for if α is sufficiently high. Once this is the case at $e_l = e_l^{FB}$, the principal can implement the first-best effort level at no additional cost of inducing truthful revelation. The parameter range for which this happens is $\alpha \geq \alpha^*$, again irrespective of p.

It should be emphasized that this finding does not depend on the normalization of the manager's wealth and reservation utility to zero. In particular, his limited liability could more generally be represented by a minimum transfer $\underline{t}$ which may be positive (if M must be paid a minimum wage $\underline{t} > 0$) or negative (if M has initial wealth $-\underline{t} > 0$). As is easily verified, the above reasoning remains unaffected and so does the conclusion that the frequency of monitoring is irrelevant. Neither does the result depend on the fact that the auditor's reservation utility is zero, i.e., on the assumption of costless auditing (see also, e.g., **??** and **??**). Since our finding solely refers to the benefit side of raising p or α, it would remain unaffected once audits are conducted with positive probability in equilibrium. Indeed, incorporating the cost side would be a straightforward matter even if variations in the two parameters p and α were not equally costly as I have assumed. For instance, if the marginal cost of raising p is strictly positive over some range, the principal's utility will simply be strictly decreasing for these parameter values.[16] Thus, extending the model in this direction would not add additional insights.

At the same time however, the result is sensitive to the representation of limited liability

[15] Although such incentive mechanisms may appear unusual at first sight, they are not uncommon. In the internal labor markets of organizations, for instance, employees with a negative or mixed performance are often not promoted. Good examples are the 'up-or-out' schemes used in the military or in academia (where frequent evaluations are notably rare).

[16] There is one technical problem that arises in this case, though. As we have seen, P's payoff is discontinuous at $p = 0$: the value an increase in this parameter is strictly positive at $p = 0$ and then drops to zero provided we confine attention to the benefit side, ignoring cost differentials. Hence, if the marginal cost of raising p is strictly positive everywhere, the problem of choosing this dimension of the information system optimally will either have a corner solution at $p = 0$ or no solution at all.

as a (common) lower bound on transfer payments. First, this assumption implies that penalties are transfer dependent: the higher the manager's remuneration, the more he can be held liable (see also, e.g., **?**, **?**, and **?**). In contrast, **?** and **?** interpret limited liability as an exogenous upper bound on penalties, say $\bar{t} > 0$, that does not vary with how much money M has or receives. To see why the result no longer holds, observe that P will again inflict the maximal punishment for an unfavorable report $r = h$. Hence, we must have $t_l^h = t_l^l - \bar{t}$ under the optimal contract. Inserting this value into (2), we see that the informational rent is now given by $\psi(e_l) - \psi(e_l - \Delta\theta) - p(2\alpha - 1)\bar{t}$, which – since $\bar{t}$ is exogenously fixed – strictly decreases in p. The reason is simply that the principal is no longer in a position to (costlessly) adjust the punishment upward as a response to a drop in p.[17] Second, the assumption that the maximal penalty is independent of the report is also critical.[18] Suppose for instance that t_l^0 could not be lowered as much as t_l^h, perhaps because courts are not willing to enforce large penalties without evidence. Then, the result would not go through essentially for the same reason as above. If p were to fall, the principal could not keep the (PC) constraint of the low-productivity manager satisfied without raising the expected rent of the high-productivity manager. As a point in case, consider a situation where the manager cannot be punished at all if the auditor observed nothing, i.e., where $t_l^0 \equiv t_l^l$.[19] This assumption is implicitly made in **?** who consider a monitoring model in which the supervisor receives signals on the agent's effort that are either correct or not informative (mistakes do not occur). The authors conclude that a first best cannot be achieved even though $\alpha = 1$ and it is straightforward to show that the frequency of valuable signals matters in their model.

Collusive Monitoring:

In line with the literature, collusion is modelled as a fully enforceable side-contract between manager auditor that the two parties sign after the auditor has received the signal but before he sends his report r.[20] This contract is unobservable to the prin-

[17] Of course, this possibility also implicitly requires that the principal is not wealth constrained.

[18] I wish to thank the referees for making me aware of this point.

[19] Observe that the principal ex post has an incentive to manipulate the auditor's report in order to retain the manager's transfer. The present analysis does not consider the possibility of manipulating r, e.g, through collusion with the auditor (see **?** for an analysis of the principal's incentives to misreport the monitoring outcome and how employing an auditor as a third party can constitute a commitment device). But while falsifying evidence without the managers's consent may be quite difficult, it is less plausible that the principal cannot bribe the auditor to conceal the auditing outcome. If the principal can hide evidence, the presumption that she can only partially condition the manager's compensation on r may be appropriate.

[20] The assumption that side contracts are perfectly enforceable is standard in the literature on

cipal and specifies payments from the manager to the auditor (or vice versa) which may be contingent on realized output and the auditor's report. Since side-contracting takes place under symmetric information (recall that M observes the outcome of A's investigation) and e is already sunk at this stage, the optimal report from the auditor/manager coalition's point of view maximizes the total wage bill $t(x, r) + w(x, r)$.

Concerning the auditor's information, I will distinguish two cases. Following **??** and others, suppose first the auditor's information is 'hard' in the sense that the outcome of his investigation could be concealed but not forged. Accordingly, if the auditor has observed a signal $s = s_h$, he can either present the true evidence ($r = h$) or else claim that he has observed nothing ($r = 0$). Since this is the only case where collusion can improve the welfare of the coalition, the following coalition incentive constraint must be satisfied:

$$\text{(CICH)} \qquad t_l^h + w^h \geq t_l^0 + w^0.$$

But it is evident from the preceding discussion that the optimal contract as characterized by Proposition 1 already satisfies the (CICH) constraint: the auditor always obtains his reservation utility and the manager receives the same payment for $r = h$ and for $r = 0$.[21] We thus have

collusion in hierarchies. In practice, of course, it may be impossible for the parties to rely on a court for enforcement, not least because side transfers are often illegal. An alternative mechanism to judicial enforcement could be that the parties are concerned with their reputation of keeping promises. For an extensive discussion of this point see **?**.

[21] See also **?** who shows in addition that this result is robust as to how the manager's limited liability is interpreted.

Proposition 2 If the auditor's information is hard, the optimal contracts under collusion free monitoring and collusive monitoring coincide. Hence, the principal's return from the relationship is again independent of the frequency p and strictly increasing in the accuracy α of the signal for $\alpha < \alpha^*$. The first best is achieved for $\alpha \geq \alpha^*$.

Finally, let us turn to a situation where the manager and the auditor can falsify evidence, i.e., where the auditor's signal is non-verifiable. This is the case of 'soft' information which has been investigated for example by **?**. The manager can now bribe the auditor to report that he did not shirk, which may be optimal if A has observed that he shirked *and* if A has observed nothing. Since both parties can agree in those states to jointly claim that the auditor has observed the manager worked correctly, we must have

$$\text{(CICS)} \qquad t_l^0 + w^0 \geq t_l^l + w^l \quad \text{and} \quad t_l^h + w^h \geq t_l^l + w^l.$$

It is straightforward to check that these two constraints bind at the optimum and that other possible coalition incentive constraint as well as the incentive constraint for the low-productivity manager do not bind and can be ignored. Also, there is also no gain for the principal from rewarding the auditor if his report was favorable for the manager. We can therefore set $w^l = 0$ without loss of generality. The optimal contract for the principal taking into account the (CIC) constraints then solves

$$\text{(3)} \qquad \begin{aligned} \max_{e,t,w} \quad & q\{\theta_h + e_h - t_h\} \\ & + (1-q)\left\{\theta_l + e_l - p[\alpha t_l^l + (1-\alpha)(t_l^h + w^h)] - (1-p)(t_l^0 + w^0)\right\} \\ \text{subject to} \quad & (\text{PC}_l),(\text{PC}_h),(\text{IC}_h),(\text{CICS}) \text{ and } t_h, t_l^l, t_l^0, t_l^h, w^h, w^0 \geq 0, \end{aligned}$$

Note that I continue to interpret p as the probability of an informative signal being observed during an audit, rather than as the probability of conducting an audit in the first place. Indeed, the above program (without loss of generality) presumes that audits are conducted with probability one, although the results presented below are equally valid for the alternative interpretation of p as the actual audit probability (see the Appendix). As before, contractual terms will be such that no side transfers are made in equilibrium. Yet, in contrast to the previous analysis the principal is now strictly worse off as compared to a situation where side contracts between M and A are not feasible. As is formally shown in the Appendix, the possibility of collusion now prompts the principal not to rely on the auditor's report if the signal is too noisy. More specifically, if α is below some critical level $\hat{\alpha}$, the manager's compensation does not depend on r and the optimal contract is equal to the no-auditor second-best scheme as defined above. If the signal is sufficiently accurate the effort of the low-productivity manager is higher than e_l^{SB}. In this case, the principal again optimally punishes the manager if the auditor has observed that he shirked. If A has observed nothing, however, the manager will be paid a positive amount, i.e. it is no longer optimal to retain the transfer in this case. Furthermore, the principal now benefits from both a higher frequency and a higher precision of the signal and a first-best solution can only be attained if $\alpha = 1$ and p is sufficiently large. Proposition 3 summarizes these findings.

Proposition 3 Suppose the auditor's information is soft and define $\hat{\alpha} = \frac{1}{1+q}$. *The optimal contract is characterized as follows:*

a) if $\alpha \leq \hat{\alpha}$, *the principal does not request the auditor's report and the optimal contract is equal to the no-auditor second-best scheme with* $e_h = e^{FB}$ *and* $e_l = e_l^{SB}$. *The principal's expected payoff is independent of* p *and* α.

b) if $\alpha > \hat{\alpha}$, *the report is requested with probability one. The manager is punished only when detected shirking and obtains the same positive transfer if the investigation was unfruitful and favorable. The effort levels satisfy* $e_h = e^{FB}$ *and* $e_l \in (e_l^{SB}, e_l^*)$. *The principal's expected return from the relationship is increasing in* p *and* α.

c) the first best is achieved if and only if $p \geq \hat{p} = [\psi(e^{FB}) - \psi(e^{FB} - \Delta)]/\psi(e_l^{FB})$ *and* $\alpha = 1$.

The intuition for part a) of the proposition is as follows. Since the optimal contract is collusion-proof, the principal must pay the auditor an amount equal to the punishment of the manager if the former has detected shirking. Using the auditor's report is therefore relatively costly if mistakes are likely to occur. To see why $\hat{\alpha}$ is inversely related to q, recall that A is only used if output is low which occurs with probability $1 - q$ because in equilibrium the manager never shirks. If this probability is high (q low) and the auditor's information rather imprecise, the expected wage of the auditor exceeds the principal's benefit from monitoring.[22]

[22]For the case of an exogenously given punishment, this result has already been obtained by **?**.

Part b) states that if α is sufficiently large, the principal uses the auditor's report with probability one. In this case, it is still optimal for her to punish the manager if the report was unfavorable ($r = h$). The rationale behind this result is that the principal has to pay at least t_l^l, either to the manager or to the auditor (this is directly implied by the CICS constraints). From the point of view of the principal, this transfer serves two purposes. First, it can be retained in order to give incentives to a high-productivity manager not to shirk. Second, it is used to compensate a low-productivity manager for his efforts. If it is efficient to monitor, the former effect dominates the latter since incentives can be given at relatively low cost.[23] Concerning the manager's compensation in case the audit revealed nothing ($r = 0$), however, the latter effect dominates the former. Incentives cannot be given through t_l^0 because both types of managers face the same probability that the auditor does not receive a signal. At the same time, raising t_l^0 puts the principal into a position to lower t_l^l which is beneficial (recall that she has to pay t_l^l in any case). This line of reasoning also explains why, contrary to the two previous cases, the principal's payoff now increases with the frequency p of the signal. Because $t_l^0 = t_l^l > t_l^h$ under the optimal contract, a rise in p will require the transfer to a θ_l-type manager to go up (unless $r = h$) which is costly for P. But this effect is overcompensated by the reduction in the informational rent of the θ_h-type manager that is caused by more frequent observations. More generally, P is no longer in a position to adjust t_l^l as p varies without affecting the expected compensation of a truthful and a non-truthful manager differently. Figure 1 depicts the principal's indifference curve in the p/α-space.

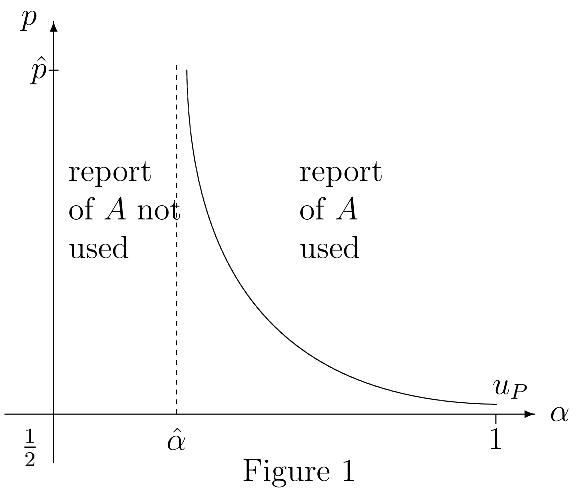

Figure 1

[23] P has to compensates A only if he observed that M shirked which occurs with a small probability given that the signal is very informative.

Finally, to see why a first best can only be achieved if $\alpha = 1$ and p sufficiently large consider first a situation where the auditor's signal is perfectly accurate ($\alpha = 1$) but $p < \hat{p}$. If the auditor's information is soft, he can claim that M did not shirk even though he has observed nothing. This prevents the principal from imposing the punishment $t_l^0 = 0$ that is needed to relax the incentive constraint of the high-productivity manager (see Proposition 1). Due to $t_l^0 > 0$, (IC_h) continues to bind if the probability of this transfer being paid is large enough, i.e. p sufficiently low. Next, consider $p \geq \hat{p}$ but $\alpha < 1$. Clearly, if A has observed that M shirked he must be given a wage equal to the punishment of M in order to tell the truth. Hence, only if the principal knows that the signal is perfectly accurate ($\alpha = 1$), she can costlessly use a compensation scheme in which A is paid a wage equal to the punishment of M in case he detected shirking. Since shirking never occurs in equilibrium, the probability of this wage being paid is zero and the first best can be implemented at no cost for the principal.

3 Concluding Remarks

This paper has reconsidered the issue of auditing and collusive behavior in hierarchies. The analysis has focused on how different information technologies affect the contractual terms and the principal's expected return from the relationship. To this end, I have distinguished between quantitative and qualitative aspects of the information technology: first, the probability that a signal is received (the frequency of the observation) and second, the probability that this signal is correct (the accuracy of the observation). It was demonstrated that if the auditor and the manager cannot collude or, alternatively, if the auditor's signal is hard information, the principal's utility under the optimal contract is independent of the frequency with which the additional information is actually observed. In those cases, the principal's sole interest lies in improving the precision of audits. A first-best allocation from the principal's point of view can be implemented as long as the auditing outcome is sufficiently accurate. If the signal of a collusive auditor is non-verifiable, however, the possibility of collusion imposes a cost on the principal, which induces her not to rely on the auditor's report provided that mistakes are likely to occur. Again, this decision is independent of the signal's reception probability. If the signal is relatively accurate, the auditor will be used in spite of collusion. This is the only case in which the frequency of audits has a positive impact on the principal's return from the relationship.

An additional insight may be gained if one compares these findings to those in **??** and **?**, where the auditor observes the manager's type *ex ante*, i.e. before the manager makes his announcement. These papers presume that the auditor either observes the true value of the parameter ($\alpha = 1$) or else observes 'nothing' and that the signal is hard information. A general result of this line of research is that the first best cannot be achieved unless $p = 1$, which seemingly contradicts Propositions 1 and 2 where I demonstrate that the first best is possible independently of p.[24]

The crucial difference between these models and this paper is the timing of the observation of the auditor. If A monitors ex ante and the manager simultaneously observes the signal, the manager's participation constraint must hold for all possible signals separately. Punishments are therefore infeasible. In contrast, the present model follows another branch of the literature in that the auditor receives his signal ex post (as in, e.g., **?**). This implies that the manager is at the time of contracting still uncertain about the outcome of subsequent audits, which facilitates punishments and, hence, greatly improves his incentives.[25] Therefore, the principal should prefer to audit ex post rather than ex ante, because in the former case the manager is definitely uncertain about the outcome of the auditor's investigation at the time of contracting and at the time he has to take the prescribed action. The principal benefits from the manager's uncertainty since it relaxes his participation and incentive constraints and punishments can be invoked. As already argued in **?**, one possibility for the principal to prevent ex ante observations of the auditor[26] is to monitor the manager's effort (rather than productivity), simply because it is impossible to observe a signal on e before its value has been realized. This conclusion suggests for example that firms should focus more on evaluating a worker's performance on the job instead of relying on test to determine his ability for the task.

The present framework could be extended in several directions. First, the assumption of the manager being risk neutral (above a certain threshold) may be dropped as this is likely to play a role in the determination of the optimal information system. Second,

[24] Also, collusion reduces the value of the relationship in this literature, which is not the case here. Interestingly, **?** shows that Tirole's findings remain unaffected if the auditor's information is 'soft' rather 'hard'.

[25] This point can also be made by comparing the results presented here with those of **?** who consider competitive firms conducting pre-employment tests on workers' abilities. In their paper, the probability of testing matters and the first best can only be achieved if unbounded punishments are feasible.

[26] Note that the auditor-manager coalition gains from early observations.

one could allow for coalitions other than the auditor/manager coalition. It is clear from the analysis that the principal frequently has an incentive to falsify or at least to hide monitoring evidence. The fact that she can collude with the auditor thus alters the optimal contract in a non-trivial way, which may give rise to additional considerations concerning her preferred auditing technology.

Appendix

Proof of Proposition 1

We first show that there is no need to formally distinguish between the probability of an audit and the probability of receiving a signal in an audit. To this end, let $\gamma \in [0,1]$ be the probability that an audit takes place and $\pi \in [0,1]$ be the probability that the auditor receives a valuable signal, contingent on an audit having been conducted. The corresponding transfers are denoted by t_i^{NA} if no audit takes place, t_i^{NS} if an audit has not generated a valuable signal, and t_i^r if the audit yields a (truthful) report, $i, r \in \{l, h\}$. For example, the expected transfer to a manager who truthfully claims to be of type i is then equal to $(1-\gamma)t_i^{NA} + \gamma\big\{(1-\pi)t_i^{NS} + \pi\big[\alpha t_i^i + (1-\alpha)t_i^r\big]\big\}$.

Lemma 1 Without loss of generality, we can restrict attention to contracts where $t_i^{NA} = t_i^{NS} \equiv t_i^0$.

Proof:. Instead of a contract specifying possibly different transfers t_i^{NA} and t_i^{NS}, consider an alternative contract where both transfers are replaced by

$$t_i^0 \equiv t_i^{NA} - \frac{\gamma(1-\pi)}{1-\gamma\pi}(t_i^{NA} - t_i^{NS}) = t_i^{NS} + \frac{1-\gamma}{1-\gamma\pi}(t_i^{NA} - t_i^{NS}). \tag{4}$$

Leaving the remaining payments t_i^r unchanged, the expected transfers to all types of managers remain the same under this new contract, irrespective of whether they announce their type truthfully or not. The principal's payoff is unchanged as well, provided auditing and actually receiving a signal are both equally costly in the information technology (as is assume here). Finally, the (IC) and (PC) constraints are satisfied and, due to $t_i^0 \geq \min\{t_i^{NA}, t_i^{NS}\}$, so are the limited liability constraints. Q.E.D.

Since the expected transfer to a manager who truthfully (respectively, falsely) claims to be of type $i \neq j$ now reduces to $(1-\gamma\pi)t_i^0 + \gamma\pi\big[\alpha t_i^i + (1-\alpha)t_i^j\big]$ (respectively,

$(1-\gamma\pi)t_i^0 + \gamma\pi\big[\alpha t_j^i + (1-\alpha)t_i^i\big]$), a direct consequence of Lemma 1 is that only the probability of making an observation, $p \equiv \gamma\pi$, is relevant for the analysis.

Proof of Proposition 1. From the arguments made in main text, the Lagrangian of the principal's problem (1) reads

$$
\begin{aligned}
\mathcal{L} = {} & q\{\theta_h + e_h - t_h\} + (1-q)\{\theta_l + e_l - p\alpha t_l^l + (1-p)t_l^0\} \\
& + \lambda_1\{p\alpha t_l^l + (1-p)t_l^0 - \psi(e_l)\} + \lambda_2\{t_h - \psi(e_h)\} \\
& + \lambda_3\{t_h - \psi(e_h) - p(1-\alpha)t_l^l - (1-p)t_l^0 + \psi(e_l - \Delta\theta)\},
\end{aligned}
$$

plus the non-negativity constraints on e_i, t_h, t_l^l and t_l^0. Taking derivatives, we obtain

$$
\begin{aligned}
&(5) \quad \frac{\partial L}{\partial e_h} = q - (\lambda_2 + \lambda_3)\psi'(e_h) \le 0, \quad e_h \ge 0 \quad \text{and} \quad e_h \frac{\partial L}{\partial e_h} = 0 \\
&(6) \quad \frac{\partial L}{\partial t_h} = -q + (\lambda_2 + \lambda_3) \le 0, \quad t_h \ge 0 \quad \text{and} \quad t_h \frac{\partial L}{\partial t_h} = 0 \\
&(7) \quad \frac{\partial L}{\partial e_l} = 1 - q - \lambda_1\psi'(e_l) + \lambda_3\psi'(e_l - \Delta\theta) \le 0, \quad e_l \ge 0 \quad \text{and} \quad e_l \frac{\partial L}{\partial e_l} = 0 \\
&(8) \quad \frac{\partial L}{\partial t_l^l} = -(1-q)p\alpha + \lambda_1 p\alpha - \lambda_3 p(1-\alpha) \le 0, \; t_l^l \ge 0 \quad \text{and} \quad t_l^l \frac{\partial L}{\partial t_l^l} = 0 \\
&(9) \quad \frac{\partial L}{\partial t_l^0} = -(1-q)(1-p) + \lambda_1(1-p) - \lambda_3(1-p) \le 0, \; t_l^0 \ge 0 \quad \text{and} \quad t_l^0 \frac{\partial L}{\partial t_l^0} = 0,
\end{aligned}
$$

and the constraints and their complementary slackness conditions. (5) and $\psi'(0) = 0$ imply $e_h > 0$ so that $t_h > 0$ by (PC$_h$). (5) and (6) then yield

$$
(10) \qquad q = \lambda_2 + \lambda_3,
$$

and $\psi'(e_h) = 1$ or $e_h = e^{FB}$. By a similar argument, we also have $e_l > 0$ [(7) holds with equality, see also footnote 11] so that either t_l^l or t_l^0 must be strictly positive from (PC$_l$). Now suppose $t_l^l = 0$, implying $t_l^0 > 0$. As (9) then holds with equality, we must have $\lambda_1 = (1-q) + \lambda_3 > 0$. But since inserting this expression in (8) gives $\lambda_3 \le 0$, $t_l^l = 0$ requires $\lambda_3 = 0$ which in turn implies $e_l = e^{FB}$ from (7) and contradicts (IC$_h$) together with (PC$_l$).

Hence, $t_l^l > 0$ which, using (8) implies $\lambda_1 = (1-q) + \frac{1-\alpha}{\alpha}\lambda_3 > 0$. From (7),

$$
(11) \qquad 1 = \psi'(e_l) + \frac{\lambda_3}{1-q}\left[\frac{1-\alpha}{\alpha}\psi'(e_l) - \psi'(e_l - \Delta\theta)\right].
$$

If $\lambda_3 = 0$, $e_l = e^{FB}$ and $\lambda_2 = q > 0$ from (10), so (PC$_h$) binds. (IC$_h$) is then satisfied only if

$$\frac{1-\alpha}{\alpha}\psi(e^{FB}) - \psi(e^{FB} - \Delta\theta) \leq 0$$

or $\alpha \geq \alpha^*$ as defined in the proposition. Obviously, α^* is independent of p.

If $\alpha < \alpha^*$, in contrast, we must have $\lambda_3 > 0$, which implies that (IC$_h$) binds. Inserting the expression for λ_1 into (9) gives $\lambda_3(1-2\alpha) \leq 0$ and, consequently, $t_l^0 = 0$ (more generally, t_l^0 should be equal to the maximal penalty/the agent's wealth, see **?**). If the high-types participation constraint is slack ($\lambda_2 = 0$), we have $\lambda_3 = q$ and e_l is implicitly given by

$$1 = \psi'(e_l) + \frac{q}{1-q}\left[\frac{1-\alpha}{\alpha}\psi'(e_l) - \psi'(e_l - \Delta\theta)\right];$$

from (11). If (PC$_h$) and (IC$_h$) are binding at the optimum, e_l is determined by

$$\frac{1-\alpha}{\alpha}\psi(e_l) - \psi(e_l - \Delta\theta) = 0.$$

In either case, e_l does not depend on p.[27] Since $t_l^0 = 0$, $t_l^l = [\psi(e_l)]/(\alpha p)$ from (PC$_l$). Therefore, the informational rent of the high-productivity type is independent of p [if $\lambda_3 = 0$ and (IC$_h$) is not binding, the values of t_l^l and t_l^0 can be chosen arbitrarily so as to satisfy (PC$_l$)]. Applying the envelope theorem, we have after substituting for t_l^l and λ_1,

$$\frac{\partial L}{\partial \alpha} = \frac{\lambda_3}{\alpha^2}\psi(e_l) > 0 \quad \Leftrightarrow \quad \lambda_3 > 0,$$

which completes the proof. Q.E.D.

Proof of Proposition 3

Similar to the case of collusion-free audits, I first show that the probability of an audit γ and the probability of observing a valuable signal π (contingent on an audit having been conducted) can be subsumed in one parameter $p \equiv \gamma\pi$. Using the previous notation, let the corresponding transfers be t_i^{NA} (no audit), t_i^{NS} (audit but no signal reported), and t_i^r (audit and signal r reported), $i, r \in \{l, h\}$.

Lemma 2 Without loss of generality, we can restrict attention to contracts where $t_i^{NA} = t_i^{NS} \equiv t_i^0$ *and* $w_i^{NA} = w_i^{NS} \equiv w_i^0$.

[27] Note that depending on the value of α, e_l may actually exceed e^{FB}. Nevertheless, it is straightforward to show that the incentive constraint of low-productivity never binds under the optimal solution (see **?**).

Proof:. I first argue that $t_i^{NA} + w^{NA} = t_i^{NS} + w_i^{NS}$ without loss of generality. To see this, note that because A's report maximizes the total wage bill, we must have $t_i^{NS} + w^{NS} = t_i^r + w^r$, $r \in \{l, h\}$ under any contract that satisfies the $(CICS)$ constraints (recall that collusion can emerge in any state other than $r = NA$). Now suppose $t_i^{NA} + w^{NA} < t_i^r + w^r$, $r \in \{NS, l, h\}$ and consider the alternative contract with $t_i^{NA'} + w^{NA'} = t_i^{NA} + w^{NA} + \epsilon$ and $t_i^{r'} + w^{r'} = t_i^r + w^r - \frac{1-\gamma}{\gamma}\epsilon$, $\epsilon > 0$. By construction, all coalition incentive constraints still hold. As to limited liability, change only the payments of that party whose respective constraints constraints are slack (which must be the case for at least one party, either M or A), leaving the transfers to the other party unchanged. Because expected payoffs are unaffected, the remaining incentive and participation constraints continue to hold. The case $t_i^{NA} + w^{NA} > t_i^r + w^r$ is analogous, which shows that no restriction is imposed by setting $t_i^{NA} + w^{NA} = t_i^{NS} + w_i^{NS}$. The remainder of the proof is essentially the same as in Lemma 1. In particular, since the only difference in P's problem are the new coalition proofness constraints, I only need to show that we can still set $t_i^{NA} = t_i^{NS} = t_i^0$ and $w_i^{NA} = w_i^{NS} = w_i^0$ even with $(CICS)$ binding. From the proof of Lemma 1, consider again an alternative contract that leaves the payments t_i^r and w^r, $i, r \in \{h, l\}$ unchanged and replaces t_i^{NA} and t_i^{NS} by t_i^0 as specified in (4) and w_i^{NA} and w_i^{NS} by

$$w_i^0 \equiv w_i^{NA} - \frac{\gamma(1-\pi)}{1-\gamma\pi}(w_i^{NA} - w_i^{NS}) = w_i^{NS} + \frac{1-\gamma}{1-\gamma\pi}(w_i^{NA} - w_i^{NS}). \tag{12}$$

Since there is no collusion in the no-audit state, payments for $r = NA$ can be disregarded. Furthermore, because t_l^r and w^r, $r = l, h$ are unaltered, we only need to check whether the constraint $t_l^0 + w^0 = t_l^l + w^l$ holds under this new agreement. Substituting for $t_l^0 + w^0$ using (4) and (12), and employing $t_i^{NA} + w^{NA} = t_i^{NS} + w_i^{NS}$ from above, we see that this constraint holds if and only if $t_l^{NS} + w^{NS} = t_l^l + w^l$, i.e., if and only if $(CICS)$ was satisfied under the initial contract. Q.E.D.

After we substitute for t_i^0 and w_i^0 in the expected transfer to M and A, payoffs only depend on the probability of making an observation, $p = \gamma\pi$. Again, I will for simplicity interpret p as the probability of receiving a signal, contingent on the audit having been conducted. Some further observations are in order: first, we can w.l.o.g. restrict attention to contracts where A's report is requested with probability one if M claims to be of type θ_l (note that P can always choose not rely on A's report by setting $t_i^l = t_i^h = t_i^0$). Second, because (IC_l) constraint is slack under the optimal contract, there is no need for P to request a report if M claims to be of type θ_h. Finally, recall that the binding coalition constraints are $t_l^0 + w^0 = t_l^l + w^l$ and $t_l^h + w^h = t_l^l + w^l$. But because it can never be optimal for P to set $w^l > 0$, it must be that $w^l = 0$ implying $w^0 = t_l^l - t_l^0$ and $w^h = t_l^l - t_l^h$. A's limited liability (respectively, participation) constraints can therefore be written as $w^0 = t_l^l - t_l^0 \geq 0$ and $w^h = t_l^l - t_l^h \geq 0$.

Proof of Proposition 3. The Lagrangian of the principal's problem (3) is now

$$\begin{aligned}L &= q\{\theta_h + e_h - t_h\} + (1-q)\{\theta_l + e_l - t_l^l\}\\ &+ \lambda_1\{p\alpha t_l^l + (1-p)t_l^0 + p(1-\alpha)t_l^h - \psi(e_l)\} + \lambda_2\{t_h - \psi(e_h)\}\\ &+ \lambda_3\{t_h - \psi(e_h) - p(1-\alpha)t_l^l - (1-p)t_l^0 - p\alpha t_l^h + \psi(e_l - \Delta\theta)\}\\ &+ \lambda_4\{t_l^l - t_l^0\} + \lambda_5\{t_l^l - t_l^h\},\end{aligned}$$

with the non-negativity constraints. The Kuhn-Tucker conditions are

$$\frac{\partial L}{\partial e_h} = q - (\lambda_2 + \lambda_3)\psi'(e_h) \leq 0, \quad e_h \geq 0 \quad \text{and} \quad e_h\frac{\partial L}{\partial e_h} = 0 \tag{13}$$

$$\frac{\partial L}{\partial t_h} = -q + (\lambda_2 + \lambda_3) \leq 0, \quad t_h \geq 0 \quad \text{and} \quad t_h\frac{\partial L}{\partial t_h} = 0 \tag{14}$$

$$\frac{\partial L}{\partial e_l} = 1 - q - \lambda_1\psi'(e_l) + \lambda_3\psi'(e_l - \Delta\theta) \leq 0, \quad e_l \geq 0 \quad \text{and} \quad e_l\frac{\partial L}{\partial e_l} = 0 \tag{15}$$

$$\frac{\partial L}{\partial t_l^l} = -(1-q) - \lambda_1 p\alpha - \lambda_3 p(1-\alpha) + \lambda_4 + \lambda_5 \leq 0,\ t_l^l \geq 0, t_l^l\frac{\partial L}{\partial t_l^l} = 0 \tag{16}$$

$$\frac{\partial L}{\partial t_l^0} = \lambda_1(1-p) - \lambda_3(1-p) - \lambda_4 \leq 0, \quad t_l^0 \geq 0 \quad \text{and} \quad t_l^0\frac{\partial L}{\partial t_l^0} = 0 \tag{17}$$

$$\frac{\partial L}{\partial t_l^h} = \lambda_1 p(1-\alpha) - \lambda_3 p\alpha - \lambda_5 \leq 0, \quad t_l^h \geq 0 \quad \text{and} \quad t_l^h\frac{\partial L}{\partial t_l^h} = 0. \tag{18}$$

As in the proof of Proposition 1, we have $e_h, t_h, e_l, t_l^l > 0$ with $\psi'(e_h) = 1$,

$$q = \lambda_2 + \lambda_3 \tag{19}$$

$$(1-q) = \lambda_1\psi'(e_l) - \lambda_3\psi'(e_l - \Delta\theta) \tag{20}$$

$$(1-q) = \lambda_1 p\alpha - \lambda_3 p(1-\alpha) + \lambda_4 + \lambda_5. \tag{21}$$

Suppose first $\lambda_4, \lambda_5 > 0$ so that $t_l^l = t_l^0 = t_l^h > 0$. From (17) and (18), $\lambda_4 = (1-p)(\lambda_1 - \lambda_3)$ and $\lambda_5 = p(1-\alpha)\lambda_1 - p\alpha\lambda_3$. Substituting these equations into (21) and using (20), we obtain

$$1 = \psi'(e_l) - \frac{\lambda_3}{1-q}[\psi'(e_l) - \psi'(e_l - \Delta\theta)]. \tag{22}$$

Note that since the manager's compensation does not depend on the auditor's report in this case, the solution must be identical to the no-auditor second-best scheme. From (22), we have $e_l = e_l^{SB}$ if $\lambda_3 = q$ ($\lambda_2 = 0$). But $\lambda_3 = q$ implies $\lambda_1 = 1$ by (20) and (22) and we can have $\lambda_5 > 0$ only if $(1-\alpha) - \alpha q > 0$ or $\alpha < \hat{\alpha} = \frac{1}{1+q}$. Conversely, $\alpha \geq \hat{\alpha}$ contradicts the assumption that $\lambda_4, \lambda_5 > 0$ and we must have either $\lambda_4 = 0$ or $\lambda_5 = 0$ or both. $\lambda_4 = 0, \lambda_5 > 0$ immediately yields a contradiction from (17), (18) and (21). Likewise, $\lambda_4 = \lambda_5 = 0$ again contradicts (17), (18) and (21) under the assumption that $q < \frac{1}{2}$. Hence, we must have $\lambda_4 > 0$ and $\lambda_5 = 0$. Then, $\lambda_1(1-\alpha) - \lambda_3\alpha \leq 0$ from (18) as long as $\lambda_3 \leq q$ with strict inequality if $\alpha > \hat{\alpha}$. Hence, $t_l^h = 0$. There are three possibilities. First, $\lambda_2 = 0$ implies $\lambda_3 = q$. Substituting for λ_4 in (21), we get $\lambda_1 = [(1-q) + q(1-p\alpha)]/(1-p+p\alpha) > 0$. Note that $\lambda_1 < 1$ as long as $\alpha > \hat{\alpha}$. Thus, $e_l > e_l^{SB}$. Using the expression for λ_1, the optimal level of e_l can be recovered from (20) which, after some rearrangements yields

$$1 = \frac{1}{1-p+p\alpha}\psi'(e_l) + \frac{q}{1-q}\left\{\frac{1-p\alpha}{1-p+p\alpha}\psi'(e_l) - \psi'(e_l - \Delta\theta)\right\}.$$

Since $\alpha > \frac{1}{2}$, e_l is strictly lower than the optimal e_l under collusion-free monitoring. Second, $\lambda_2 > 0$ implies that both (PC_h) and (IC_h) are binding at the optimum. e_l can then be obtained from

$$\frac{1-p\alpha}{1-p+p\alpha}\psi(e_l) - \psi(e_l - \Delta\theta) = 0. \tag{23}$$

Finally, if $\lambda_2 > 0$ and $\lambda_3 = 0$, we have $e_l = e^{FB}$ and (20), (21) and (23) hold only if $\alpha = 1$ and $p \geq \hat{p} = [\psi(e_l^{FB}) - \psi(e_l^{FB} - \Delta)]/\psi(e_l^{FB})$. Again, I can apply the envelope theorem which, after substituting $t_l^h = 0$ and $t_l^0 = t_l^l$ yields

$$\frac{\partial L}{\partial p} = \lambda_3\alpha - \lambda_1(1-\alpha) \geq 0,$$

by (18), with strict inequality if $\alpha > \hat{\alpha} = \frac{1}{1+q}$ and $\lambda_3 > 0 \Leftrightarrow p < \hat{p}$ and $\alpha < 1$ which completes the proof. Q.E.D.

Anke S. Kessler
Simon Fraser University
Department of Economics
8888 University Drive
Burnaby, B.C., V5A 1S6

Canada

email: akessler@sfu.ca