Partial Monitoring, Managerial Compensation, and the Internal Efficiency of the Firm

by

Christoph Lülfesmann*

University of Bonn

July 1998

SFB-Discussion Paper A-532

Summary

The paper investigates an adverse selection model with monitoring of managerial effort. In contrast to the literature, we assume that the manager can be punshished only if his effort is below a certain level that is monitored by the principal. Surprisingly, the optimal labor contract may induce an equilibrium effort which is *lower* than in the standard model without monitoring. This result holds for any discrete distribution of managerial types. Moreover, we show in the continuous type case that the optimal contracts for high-quality (low-quality) managers are purely output-dependent (effort-dependent).

Keywords: Monitoring, Productive Efficiency, Adverse Selection.
JEL-Classification: H57, L51.

*Address: Department of Economics, Adenauerallee 24-42, University of Bonn, D-53113 Bonn, Germany. e-mail: clmann@united.econ.uni-bonn.de. phone: +49-228-739288. fax: +49-228-739239. I would like to thank Ingolf Dittmann, Gabor Gyarfas, Nico Hansen, David Wettstein and, in particular, Anke Kessler for valuable comments and suggestions. All remaining errors are my own. Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged.

1 Introduction

The determinants of the internal (or productive) efficiency of the firm rank high on the research agenda of economic theory. Information theoretical approaches have succeeded in opening the black box of production in the neoclassical paradigm. While one branch of the literature has investigated the moral-hazard problem and its consequences for productive efficiency and optimal compensation schemes, a different line of research has concentrated on principal-agent models with adverse selection: in these models, the agent knows more about a parameter that is relevant for production (e.g., his intrinsic ability or the firm's production cost) than the principal. As one of the key insights, it was found that informational asymmetries between owner (principal) and manager (agent) diminish the productive efficiency of the firm relative to a first best situation.[1] Moreover, the agent's optimal compensation scheme is an incentive contract which allows the principal to elicit the agent's private information and discriminates among agents of different 'types'.

The present article studies the implications of monitoring in such a situation. We follow Laffont and Tirole (1992, 1993) and Kofmann and Lawarrée (1993) who investigate the effect of audits on managerial effort. To this end, we consider an adverse selection model where the agent is privately informed on his ability to perform a particular task. In addition, he can exert (unobservable) effort that improves the productive efficiency of the firm.[2] Kofman and Lawarrée (1993) consider an auditing technology where a (possibly colluding) supervisor receives an imperfect signal on the agent's intrinsic ability rather than his effort level. The authors derive comparative static results with respect to he maximum punishment that can be invoked if the supervisor's report on the agent's ability differs from his own report. If the maximum penalty is low, the possibility to monitor leaves managerial equilibrium effort unaffected but reduces informational rents. Only if the penalty can be set sufficiently large so as to ensure zero rents, monitoring has allocative consequences and managerial effort increases.[3] Laffont and Tirole (1992,93), in contrast, focus on random audits of managerial effort. In their framework, the principal detects shirking with an exogenously given probability, and the maximum penalty is imposed

[1]See, e.g., Baron and Myerson (1982) or Laffont and Tirole (1986).

[2]Among others, Polinsky and Shavell (1979), Baron and Besanko (1984), Nalebuff and Scharfstein (1987), and Melumad and Mookherjee (1989) study monitoring in adverse selection models without an effort component.

[3]See also Dittmann (1996) who endogenizes the auditor's reliability in this framework.

when shirking is detected. Since the agent is assumed to be protected by limited liability, this maximum punishment coincides with the agent's contracted remuneration (his wage). In equilibrium, monitoring reduces the informational rent of an agent with high-productivity: as a random inspection may prove that he shirked, mimicking low-ability agents becomes less attractive. The authors conclude that '... monitoring of effort enables [the principal] to reduce the informational rent and consequently leads to a *smaller* distortion of effort for the inefficient type'.[4]

To summarize, existing models suggest that monitoring in general increases the agent's effort and therefore has a positive impact on the productive efficiency of the firm. The present paper demonstrates that this conclusion, though intuitive, is very sensitive with respect to the considered monitoring technology. Specifically, it crucially depends on the assumption that random inspections can generate enough evidence for a court to enforce maximum punishments. The monitoring technology proposed in our model, in contrast, strongly limits the power of a single observation.[5]

As the previous approaches, our model is motivated by the observation that even though labor or procurement contracts frequently specify input targets as the agent's working hours (or the attention he has to devote to an assigned task) the principal will most likely be unable or unwilling to spend the resources needed to audit an agent's entire activities. In practice, she may therefore often engage in different forms of partial monitoring. Among these forms are random inspections as well as continuous auditing processes where the agent is fully controlled over a certain time interval. Those different types of monitoring may also be sequentially used. For example, the principal may start with random inspections. Provided that these point audits bring evidence on the agent's shirking, she will frequently intensify her monitoring attempts and switch to a continuous monitoring process. A continuous gathering of evidence may be needed if courts are reluctant to penalize an agent heavily unless the employer proves shirking over a longer period.[6] As stated above, standard models analyzing monitoring in adverse selection situations assume that a single observation (a random inspection) suffices to impose the highest possible punishment on an agent. In contrast to that view, we consider a situation

[4]Quotation from Laffont-Tirole (1993), p.529; emphasis added.

[5]We also allow for monitoring with probabilty one (random inspections would not alter our qualitative results).

[6]This may not be true when the agent's shirking takes a criminal form (e.g., he is catched stealing), where a single observation may be sufficient to impose maximum penalties.

where a continuous auditing process is needed to detect the agent shirking and impose judicial penalties against him. The idea is captured by the simple monitoring technology developed in this paper: monitoring allows the principal to choose a particular effort level and observe whether the agent worked less than this amount.

To give an example what we have in mind here, suppose the principal is the owner of a small factory. Surely, the owner can control an employee's working hours and his performance at times she is physically present. She may not be able to control the worker, however, at times she is absent. Hence, as long as the worker meets a specified output target, the owner will have no means to penalize him for any shirking that occured at times where no audits are conducted.[7] In this scenario, the principal faces a natural trade off. She can be present over the entire week (thereby precluding the agent from shirking) which may be very costly if her opportunity costs of doing so are high. Alternatively, she may pursue a part-time monitoring strategy that enables her to enforce a certain minimum effort level which corresponds to the monitoring technology described above. We show that - as in the existing literature - monitoring reduces the informational rents of more productive workers in this case. We also find, however, that monitoring can lead the principal to induce an equilibrium effort that is *lower* than in the standard model without monitoring. Hence, the availability of audits may induce the principal to reduce working loads and, hence, sacrifice productive efficiency. This stands in sharp contrast to common beliefs as well as previous findings. Yet, the intuition behind this outcome is relatively simple: while monitoring allows the principal to reduce informational rents, the marginal rents granted to highly productive agents expand for any given effort level. Moreover, we find that the optimal contract may have to specify an effort target although parties rationally forsee that the agent will subsequently provide an effort in excess of what has been contractually specified.

[7] A different interpretation of the monitoring technology arises in - public or private - procurement contracts for building projects. In such a situation, it is likely that the initial contract can be made contingent on the service of technical equipment such as concrete mixers or slewing cranes whereas the quality of the employed workers or the used concrete cannot be measured, at least not at reasonable costs. Accordingly, it may be not be possible to institute legal proceedings against the contractor as long as inputs are above some *verifiable* level. For instance, it may be impossible to prove that the contractor shifts employees and equipment between different activities ('cost-shifting'), which has been seen as a major impediment against input-based cost-plus contracts in the literature on defense procurement [see Kovacic (1991)]. Such evidence will be even harder to provide if the agent at the same time met the contractually specified output target.

While this rather counterintuitive result on productive efficiency applies for any finite type space, it does not carry over to a continuous distribution of managerial abilities. In this limit case, monitoring leaves the equilibrium effort of high ability agents unaffected. Less productive types, however, work strictly more than under a standard contract. Interestingly, there exists a dichotomy between the optimal contracts proposed to different types of agents in this case. Whereas productive agents sign incentive contracts that are solely contingent on output, less efficient agents are pooled and the optimal contract specifies only an effort requirement. We believe that both outcomes mirror labor arrangements found in practice. While output dependent bonuses and stock options are commonly granted to managers on high hierarchy levels, the pay of low-tier employees sually depend on input measures such as their working hours.

An outline of the paper is as follows: Section 2 sets up a discrete two-type model and solves for the second-best standard screening contract in Subsection 2.1. The optimal effort-contingent contract is derived in Subsection 2.2. Subsection 2.3 compares the principal's profits under both types of contractual arrangements and endogenizes her monitoring decision. Section 3 considers the limit case where the distribution of agent's types becomes continuous. The final Section 4 concludes with a discussion of the results.

2 A Discrete Model

2.1 Setup

Consider a situation in which a principal (P) hires an agent or manager (A) to carry out a productive task. Both parties are risk-neutral. There are two possible types of agents $\theta \in \{\bar{\theta}, \underline{\theta}\}$ characterized by different abilities $\bar{\theta} > \underline{\theta}$. Let $p \in [0, 1]$ be the principal's perceived ex-ante probability that the manager is of the high-productivity type. The production function

$$y(e, \theta) = e + \theta \tag{1}$$

is additive in managerial ability and effort (a nonnegative scalar).[8] The manager's idiosyncratic effort costs $\psi(e)$ are assumed to be increasing and strictly convex. In addition, we impose the Inada conditions $\psi(0) = 0$, $\lim_{e\to 0}\psi'(e) = 0$ and $\lim_{e\to\infty}\psi'(e) = +\infty$. The manager's utility is $U = t - \psi(e)$, where t denotes the transfer (wage) payment he recieves in compensation for his work.

The principal is the owner of the production technology or, alternatively, purchases the good from the agent. While the agent's intrinsic productivity θ is his private information, it is possible for the principal to (partially) monitor the effort exerted by A. In particular, we assume that she can gather verifiable evidence on whether the manager's true effort is lower than some boundary level E. We will call E the *enforceable* effort level.[9] Therefore, the principal can - provided that the bilateral contract specifies an effort requirement $E^R \leq E$ - successfully sue the agent only if his true effort falls short of E. In other words, the principal can require the agent to exert any effort level $E^R \leq E$ by punishing him sufficiently if his effort falls below E^R. To ensure that the punishment is high enough for A to comply with the contractual terms, we assume that the principal can retain the transfer t in this case.[10] If the manager has provided this minimum level, in contrast, it is useless to appeal to the court and punishments cannot be invoked: any required effort level larger than E cannot be contractually enforced. Throughout the paper, we impose the following assumption:

Assumption (Free Disposal): The manager can unobservably and costlessly waste or hide output y.

Note that in a standard adverse-selection model, it can never be optimal for an agent to waste output because his compensation strictly increases in that variable. As will become clear below, the high-productivity manager gains from the option to hide output in our framework and this assumption is essential. Although we find that output is never wasted in equilibrium, the manager's option to do so positively affects

[8]We adopt an additive separable function [as, e.g., in Laffont and Tirole (1993) and Kofman and Lawarrée (1993)] for analytical convenience only. All qualitative results do not depend on this specification.

[9]It is not important whether or not the principal actually *observes* the agent's true effort through monitoring as long as she cannot prove shirking to the court when the true effort exceeded E.

[10]See also Laffont and Tirole (1993). While their results crucially depend on this assumption, the exact size of the maximum punishment in the present paper is not significant. In particular, our results apply for any maximum punishment level from the interval $[m, \infty]$ where the value m is characterized by some positive $m < t$.

his equilibrium utility.[11] In what follows, we will refer to $\tilde{y} \leq y$ as the manager's *realized* production level which is assumed to be publicly observable and verifiable. The principal's utility (her profit) is accordingly given by $P = \tilde{y} - t = \tilde{y} - \psi(e) - U$. Note that the principal's valuation of managerial utility is negative, since a higher utility implies higher wage payments.

As a point of reference, let us briefly state the first-best solution when θ is publicly observable. Maximizing the principal's utility subject to the constraint that A receives his reservation payoff (normalized to zero) yields

$$\psi'(e^{FB}) = 1, \quad \theta \in \{\bar{\theta}, \underline{\theta}\} \quad \text{or} \quad e^{FB}(\theta) = \psi'^{-1}(1) \tag{2}$$

and $U(\theta) = 0 \Leftrightarrow t^{FB} = \psi(e^{FB})$. Hence, in a world without informational asymmetries, the manager receives only his reservation payoff, and the principal induces an effort level that equates marginal benefits and marginal costs and is independent of θ.

Now suppose θ is private information of the agent. Invoking the revelation principle, we can restrict attention to contracts which ensure that the agent truthfully announces his productivity type in equilibrium. We will call a contract $\mathcal{C}_S = \{t(\hat{\theta}), \tilde{y}(\hat{\theta})\}$ that prescribes a level of realized output $\tilde{y}(\hat{\theta})$ and a remuneration $t(\hat{\theta})$ conditional upon the agent's announcement $\hat{\theta}$ an output-dependent *standard* screening contract. When a positive monitoring level E can be enforced, the principal can alternatively offer an *effort-contingent* contract $\mathcal{C}_M = \{t(\hat{\theta}), \tilde{y}(\hat{\theta}), E^R(\hat{\theta})\}$ which in addition specifies a (minimal) effort requirement $E^R \leq E$.[12]

Standard Contract under Incomplete Information

For later reference, we calculate the second-best solution when the principal does not dispose of a positive enforceable effort level or, equivalently, offers a standard screening contract $\mathcal{C}_S$ to the manager. Let $e(\theta, \hat{\theta}) = y(\theta, \hat{\theta}) - \theta$ represent the effort level of an agent of type θ who announced $\hat{\theta}$. The corresponding optimization

[11] Alternatively, we could impose the assumption that the agent cannot be punished for producing an output level in excess of what has been laid down in the contract. Since the principal is better off the higher the production level, it is hard to imagine that a court would enforce a penalty in this case.

[12] We will show below that E^R always coincides with E unless the enforceable level exceeds the first-best effort.

program for the principal under asymmetric information can then be written as[13]

$$\max_{e(\underline{\theta}),e(\bar{\theta})} p[\bar{\theta}+e(\bar{\theta})-\psi(e(\bar{\theta}))-U(\bar{\theta})]+(1-p)[\underline{\theta}+e(\underline{\theta})-\psi(e(\underline{\theta}))-U(\underline{\theta})] \tag{3}$$

s.t.

$$\begin{aligned} t(\theta)-\psi(e(\theta,\theta)) &\geq t(\hat{\theta})-\psi(e(\theta,\hat{\theta})), \quad \theta,\hat{\theta}\in\{\bar{\theta},\underline{\theta}\} \qquad (IC) \\ t(\theta)-\psi(e(\theta,\theta)) &\geq 0, \quad \theta\in\{\bar{\theta},\underline{\theta}\} \qquad (IR). \end{aligned}$$

One can easily show that only the incentive compatibility constraint (IC) of the high-productivity manager $\bar{\theta}$, and the individual rationality constraint (IR) of the low-productivity manager $\underline{\theta}$ are binding while their respective counterparts are slack. We can simplify notation by abbreviating $\bar{e}$ (respectively $\underline{e}$) for $e(\underline{\theta})$ (resp. $e(\bar{\theta})$). Furthermore, define $\Delta\theta=\bar{\theta}-\underline{\theta}>0$ and let $\hat{e}\equiv\underline{e}-\Delta\theta$ be the effort level of the high-productivity agent who shirked, i.e. claimed to be of low ability. Using this notation in what follows, program (3) reduces to

$$\max_{\bar{e},\underline{e}} p\{\bar{\theta}+\bar{e}-\psi(\bar{e})-[\psi(\underline{e})-\psi(\hat{e})]\}+(1-p)\{\underline{\theta}+\underline{e}-\psi(\underline{e})\}. \tag{P_S}$$

The (necessary and sufficient) first order conditions are

$$1 = \psi'(\bar{e}^S) \quad \text{and} \tag{4}$$

$$1 = \psi'(\underline{e}^S)+\frac{p}{1-p}[\psi'(\underline{e}^S)-\psi'(\hat{e}^S)]. \tag{5}$$

Condition (4) is the well known no-distortion-at-the-top property: the principal optimally implements the first best effort level when the manager is highly productive. The effort of the less productive manager, in contrast, is distorted downwards [see (5)]. This reduction in productivity is motivated by the principal's interest to reduce the informational rent of the high-productivity manager (his gain from shirking). Noting that for any effort level of the low-productivity manager $\underline{e}$ under the standard contract, we can write this rent as $U_S\equiv\psi(\underline{e})-\psi(\hat{e})>0$ which is strictly increasing in $\underline{e}$.

2.2 Monitoring and Effort-Contingent Contracts

We now assume that the principal can enforce a positive effort level E through monitoring. In this section, we will take E as exogenously given.[14] Let us first outline

[13] As is easily seen, the agent will produce exactly $y=\tilde{y}$ under a contract $\mathcal{C}_S$. Thus, we can substitute $\tilde{y}(\cdot)$ in the principal's program by $\theta+e(\cdot)$.

[14] This assumption is relaxed in Section 2.3.

the way in which an effort contingency can potentially improve the principal's payoff. Recall that under the standard screening contract, the principal has to concede an informational rent $U_S = \psi(\underline{e}) - \psi(\hat{e}) > 0$ to the agent. Now suppose that the contract includes an effort requirement $E^R \leq E$. If the principal implements an effort level of the low-productivity agent $\underline{e}$ characterized by $\underline{e} \geq E^R$, the informational rent of the more productive agent becomes

$$U_M = \psi(\underline{e}) - \psi(\max\{\hat{e}, E^R\}). \tag{6}$$

Hence, monitoring affects the principal's utility in two respects:

a) Suppose first that the enforceable effort level satisfies $\underline{e} > E > \hat{e}$ and that $E^R = E$. In this case, (6) reduces to[15]

$$U_M = \psi(\underline{e}) - \psi(E) < U_S(\underline{e}) \quad \forall\, \underline{e}.$$

This *rent reduction effect* emerges since the more productive agent now finds it ceteris paribus less attractive to mimic his less productive counterpart, i.e. shirking is less profitable: a high-productivity agent who shirks now must exert at least E (instead of $\hat{e}$). Thus, the minimal effort requirement forces him to work more. Note also that because output produced in this case, $y = E + \bar{\theta}$ exceeds $\tilde{y}(\underline{\theta}) = \underline{e} + \underline{\theta} = \hat{e} + \bar{\theta}$, and the manager has to waste the excess output $y - \tilde{y}(\underline{\theta}$.

b) Second, suppose that $E \geq \underline{e}$. In this case, the principal can set $\underline{e} = E^R$ and the agent's rent becomes

$$U_M = \psi(E^R) - \psi(E^R) = 0.$$

Thus, the principal does not concede positive rents in this case. Furthermore, output increases in $\underline{e}$. Therefore, it is indeed optimal to set $E^R = E \Leftrightarrow \underline{e} = E$ as long as $E \leq e^{FB}$. We will call this the *output effect* of monitoring: the higher E, the higher the implemented effort in the considered interval as long as E is weakly below the first best effort level.

In order to determine the optimal contract for the principal, we can again substitute the (IR)-constraint of the high-productivity agent and the (IC)-constraint of

[15]Since the agent's rent decreases in E^R, it is optimal for the principal to set $E^R = E$.

the low-productivity agent into the objective fuction. It is easy to see that the latter constraint will be binding whenever $\underline{e} \leq e^{FB}$ in which case we know from the reasoning above that $E^R = E$. Thus, if we restrict attention to enforceable effort levels $E < e^{FB}$ and contractual minimum effort levels $E^R = E$, we can write the principal's maximization program as

$$\max_{\underline{e},\bar{e}} p\{\bar{\theta}+\bar{e}-\psi(\bar{e})-[\psi(\max\{\underline{e},E\})-\psi(\max\{\hat{e},E\})]\}+(1-p)\{\underline{\theta}+\underline{e}-\psi(\underline{e})\}. \qquad (P_M)$$

It is immediate that the implemented effort of the high-productive agent still coincides with the first best level, i.e. $\bar{e}^M = e^{FB}$. To determine the optimal effort level of the less productive agent, $\underline{e}^M$, the maximum operator in P_M renders it necessary to analyze different intervals for any given E. Note that, in any of these intervals analyzed below, program P_M is strictly concave such that any interior solution is unique and a local maximum.

- Consider effort levels $\underline{e}$ from the range $\underline{e} > E + \Delta\theta$. Since this interval implies $\hat{e} < E$, neither the rent reduction nor the output effect of monitoring apply, and the program P_M reduces to the standard screening program in that interval. Accordingly, the local optimum in case of an interior solution coincides with $\underline{e}^S$ as implicitely defined by (5)[16]. [*Case 0*]

- Next, consider the interval $E+\Delta\theta \geq \underline{e} > E$. Since this interval is equivalent to the range $\underline{e} > E \geq \hat{\underline{e}}$, the rent reduction effect of monitoring applies. Hence, a corresponding interior solution of P_M must fulfill the the first-order condition

$$1 = \psi'(e^*) + \frac{p}{1-p}\psi'(e^*) \qquad (7)$$

  at $\underline{e} = e^*$. [17] [*Case 1*]

- Finally, consider the range $\underline{e} \leq E$. By our preceding arguments, the local optimum in this interval coincides with E: recall that the principal does not have to concede positive informational rents in this case. For this reason, it

[16] If $\underline{e}^S < E + \Delta\theta$, we have a corner solution $\underline{e} = E + \Delta\theta$. Note, however, that this solution can never be the global optimum: this effort level entails no rent reduction effect of monitoring, such that it is strictly dominated by $\underline{e}_S$.

[17] Alternatively, the following corner solutions can emerge: $\underline{e} = E$ for $e^* < E$, or $\underline{e} = E + \Delta\theta$ for $e^* > E + \Delta\theta$. Again, the latter of these boundary solutions can never be globally optimal; see Case 0.

must be optimal to increase $\underline{e}^M$ as much as possible. Thus, $\underline{e}^M = E$ as long as E does not exceed e^{FB} (which we have ruled out by assumption). [*Case 2*]

Taken together, there are three candidate optima for a global optimum of P_M which we indicate as $\underline{e}^M$: to begin with, the optimal effort level may correspond to that under the standard screening contract [Case 0]. Clearly, monitoring then has no effect on the principal's payoff and he renunciates from proposing an effort contingency. In contrast, the other candidate optima make it necessary to specificy a contractual effort contingency. First, it may be optimal to implement $\underline{e}^M = e^*$ as defined in Case 1, which requires that E and the solution to (7), e^*, are such that $e^* > E \geq \hat{e}^*$ holds. Second, if E weakly exceeds e^*, the implemented effort level $\underline{e}^M$ under a non-trivial effort contingent contract corresponds to E as long as $E \leq e^{FB}$.[18] Recall that we already have ruled out the case $E > e^{FB}$ where the principal can achieve a first-best result by setting $\underline{e}^M = e^{FB}$. Now, we can easily translate these results into the optimal effort contingent contracts for any level of E. These findings are illustrated in Figure 1 below.

Figure 1: Monitoring and Effort

The figure displays the level of managerial effort under the optimal standard and effort contingent contract, respectively, depending upon the level of enforceable effort E. For E smaller than the boundary effort $\hat{e}^* = e^* - \Delta\theta$, any effort contingency must be useless since the associated globally optimal effort level corresponds to that under a standard arrangement, i.e. monitoring does not affect the principal's payoff [Case 0]. In the intermediate interval $e^* > E \geq \hat{e}^*$, the *rent reduction effect* of monitoring explained above arises: a marginal change in E decreases informational rents, leaving equilibrium effort unaffected. As a consequence, effort is constant and equal to e^* over the considered range [Case 1]. In the region where $\underline{e}^M = E$, informational rents have vanished, and the *output effect* of monitoring comes into play [Case 2].

The figure also shows that there exists a nonempty set of enforceable effort levels $E \in [\hat{e}^*, \underline{e}^S)$ where the equilibrium effort of the low-productivity agent is strictly

[18] Note that when the interior solution e^* prevails in Case 1, continuity of P_M at $\underline{e} = E$ guarantees that $P_M(e^*) > P_M(E)$. Otherwise, $\underline{e}^M = E$ must be optimal whenever a non-trivial effort-contingent contract is globally efficient.

lower under the effort contingent contract with monitoring than under a standard screening contract without monitoring. To see this, one can simply compare (5) and (7) which reveals that $e^* < \underline{e}^S$. In other words, monitoring of effort can lead the principal to sacrifice productive efficiency. The proposition below summarizes our results:

Proposition 1 *Let $\underline{e}^M$ be the equilibrium effort of the low-productivity manager whenever an effort contingent contract is globally optimal. Then, for all values of E in the interval $[\hat{e}^*, e^*]$, we have $\underline{e}^M = e^*$. For values $E \in [e^*, e^{FB}]$, the solution is $\underline{e}^M = E$ and the high-productivity manager obtains no informational rents.*
Moreover, the induced effort level of the low-productivity manager under the effort-contingent contract with monitoring, $\underline{e}^M$, is strictly less than his effort level under the standard contract without montitoring, $\underline{e}^S$, for all $E \in [\hat{e}^*, \underline{e}^S)$.

While similar results to those stated in the first part of Proposition 1 can already be found in related models with different monitoring technologies,[19] the last part of the proposition is to our knowledge new. It states that the possibility of monitoring may lead the principal to implement an effort level that is *lower* than the optimal effort without monitoring. Although this result may be surprinsing at first glance, it has an intuitive explanation: consider again the rent of the high-productivity manager in the intermediate range of enforceable effort levels [Case 1] which can be written as $U_M = \psi(\underline{e}) - \psi(E)$ for a given $\underline{e}$. Clearly, U_M is strictly lower than under the standard contract without monitoring and decreases in E. From this observation, one might be tempted to conclude that higher levels of E lead the principal to improve allocative efficiency (increase $\underline{e}^M$). This reasoning, however, fails to take into account the *marginal* increase in rent associated with a higher effort of the low-productivity manager. Comparing the marginal change in rents under both contractual types yields:

$$\begin{aligned}\frac{\partial U_M}{\partial \underline{e}} &= \psi'(\underline{e}) \\ &> \psi'(\underline{e}) - \psi'(\hat{e}) = \frac{\partial U_S}{\partial \underline{e}}\end{aligned}$$

[19] See, e.g., Kofman and Lawarrée (1993). Although monitoring in their model is random and concerns the agent's unknown ability rather than his effort, they also find a rent-reduction and an output effect of monitoring. Accordingly, the optimal contract in their model has similar properties: as monitoring is more and more facilitated, the principal at first leaves effort unaffected. Only after all rents have been extracted, she optimally increases implemented effort (output).

Accordingly, increasing $\underline{e}$ is *more costly* for the principal if she monitors. Thus, the principal's best strategy is to forgo additional productive efficiency when the agent is of the low-productivity type which can help her to save on informational rents. Observe that this downwards-distortion outcome continues to hold if $\underline{e}^M = E$ as long as the the enforceable effort level is smaller than $\underline{e}^S$, the managerial effort under the standard contract.

Some properties of the optimal effort-contingent labor arrangement are worth to be stated:

Corollary 1 *If $E \leq e^{FB}$, the optimal effort-contingent contract for the low-productivty agent specifies an input target $E^R = E$ as well as an output target $y^M(\underline{\theta})$. The associated equilibrium effort $\underline{e}^M$ satisfies $\underline{e}^M \geq E$.*
Conversely, the optimal contract for the high-productivity agent manager is purely output-dependent. The associated equilibrium effort $\bar{e}^M$ satisfies $\bar{e}^M = e^{FB}$.

The corollary asserts that the principal may optimally specify an input requirement for the low-productivity agent even if this target is not binding. The intuition behind this result lies in the fact that even a non-binding input requirement may prevent high-productivity agents from shirking, i.e. mimicking their less productive counterparts. Including an effort requirement for the high-productivity agent, in contrast, can never be profit-enhancing as his rents are independent of his own effort.[20]

In the next section, we will briefly compare the implications of both contract types on the principal's profit. Obviously, the fact that positive effort levels E can be enforced through monitoring can never hurt the principal. As we will see, however, she will not always use the instrument of effort-contingent contracts. This is true even for parameter values of E for which the rent reduction or output effects of monitoring apply.[21]

[20] Note, too, that a low-productivity manager has no incentive to mimic the high-productivity manager due to the latter's (high) output target. Thus, input requirements for highly productive individuals do not affect the incentives of those who are less productive.

[21] Technically, for $\hat{e}^* < E < \hat{e}^S$, program (P_M) may have two local optima, $\underline{e}^M$ and $\underline{e}^S$, respectively. The welfare ranking of these optima is evaluated in the next section.

2.3 Optimal Contracts and Endogenous Monitoring

In the preceding sections, we have derived the optimal standard and effort-contingent contracts, respectively. We can now proceed and elaborate on the following issues in more detail: first, for which range of parameter values E is the effort-contingent contract preferable? Second, which level of E will the principal choose when monitoring is costly?

Optimal Contracting

When is it optimal for P to offer an effort-contingent contract? To deal with this issue, it is useful to start with some observations that straightforwardly emerge from our previous analysis.

- If $E \leq \hat{e}^* = e^* - \Delta\theta$, it is useless to rely upon a contractual specification of effort and the standard contract is optimal: for this parameter values, E does not affect the effort choice of both types of managers. In particular, it is not binding for a high-productivity manager who shirks.
- If $E \geq \hat{e}^S = \underline{e}^S - \Delta\theta$, the effort-contingent contract strictly dominates the standard contract. It is simple to see that for these levels of monitoring, an effort contingency increases the principal's profits: even if she still implements $\underline{e}^S$, the productive manager's informational rents are reduced. Moreover, $\underline{e}^S$ is no longer optimal since the global solution to (P_M) now becomes $\underline{e}^M \in \{e^*, E\}$.

These arguments demonstrate that there exists a threshold value of the enforceable effort, $\tilde{E} \in (\hat{e}^*, \hat{e}^S)$ above which effort-continent contracts dominate standard contracts. This threshold value $\tilde{E}$ may be larger or smaller than e^*. [22]

Proposition 2 *There exists a thereshold level $\tilde{E} \in (\hat{e}^*, \hat{e}^S)$ such that for any enforceable effort larger (smaller) than $\tilde{E}$, the optimal contract for the principal is an effort-contingent contract with monitoring (a standard contract without monitoring). In case that $\tilde{E} < e^*$ applies, the induced effort of both agents strictly exceeds the contractual minimum requirement for all $E \in [\tilde{E}, e^*)$.*

Endogenous Monitoring

[22] It is generally impossible to deduce whether $\tilde{E} \leq e^*$. For $\bar{\theta} = 1$ and the quadratic effort cost function $\psi(e) = e^2/2$, for example, one can show that $\tilde{E} \geq e^*$ as $\Delta\theta \leq 1 - p(1 - \sqrt{1-p})$.

The preceding analysis has taken the level of effort that can be enforced through monitoring, E, as exogenously given. Morevover, there were no costs associated with monitoring for the principal. In this section, we relax these assumptions and show that a) monitoring endogenously arises even if it is costly and b) all relevant cases discussed in the previous section can occur. For this purpose, we now allow the principal to observably invest into an enforceable level of E prior to her contract offer to the agent with the understanding that after P has invested, she can verifiably observe any $e \leq E$ exerted by A. Her associated investment (monitoring) costs are represented by an increasing and convex monitoring cost function $K(E) = \alpha k(E)$, $\alpha \geq 0$.

Let us first calculate the values of E that are locally optimal in the respective intervals $E \in (\hat{e}^*, e^*)$ and $E \geq e^*$, respectively (note that one of the corresponding local optima can be the global optimum only if it exceeds $\tilde{E}$). Maximizing the principal's objective function $P^M(\underline{e}^M(E)) - K(E)$ with respect to E and assuming that $K'(E) > \psi'(E) \,\forall\, E$, one obtains the necessary and sufficient first-order conditions for local interior maxima

$$p\psi'(E) = K'(E) \quad \Leftrightarrow \quad E \in (\hat{e}^*, e^*) \quad \text{(Case 1)} \qquad (8)$$

$$\text{and} \qquad (1-p)[1-\psi'(E)] = K'(E) \quad \Leftrightarrow \quad E \geq e^* \quad \text{(Case 2).} \qquad (9)$$

Now, we can ask which of these solutions is efficient provided that a positive monitoring level is the global optimum. Observe that both interior local solutions cannot (generically) prevail at the same time: defining E_1 (E_2) as the solutions to (8) (resp. (9)) and rewriting (9) to read $p\psi'(E) + (1-p-\psi'(E)) = K'(E)$, we see that $E_1 < E_2$ is necessary for both local optima to exist. This condition, however, is not consistent with $E_2 \geq e^*$.[23] Moreover, the principal's objective function is continuous in E even at $E = e^*$. Thus:

Proposition 3 *There exist parameter values $0 < \underline{\alpha} < \bar{\alpha}$ such that, when monitoring is globally optimal, the principal's optimal choice of E^* is a) $E^* = e^{FB}$ for $\alpha = 0$, b) determined by (9) for $\alpha \in (0, \underline{\alpha}]$, and c) determined by (8) for $\alpha \in [\underline{\alpha}, \bar{a})$.*
Moreover, a positive investment in monitoring is globally optimal whenever $a \leq \tilde{\alpha} \in]0, \bar{a})$ and the associated effort level that can be enforced is characterized by $E^* \geq \tilde{E}$.

[23] Since $\psi'(e^*) = (1-p)$ from (7), the second term in $p\psi'(E) + (1-p-\psi'(E)) = K'(E)$ must be negative which implies $E_2 < E_1$, a contradiction.

Proof: As $\alpha \to 0$, E_2 converges to e^{FB}. For positive but small values of α, the interior solution E_2 still prevails while no interior solution to (8) exists. At some boundary value $\underline{\alpha}$, $E_2 = e^*$; for higher α, he have a corner solution to (9) and the interior local solution to (8) and $E^* = E_1$. Finally, for some level $\bar{\alpha} > \underline{\alpha}$, $E_1 = \hat{e}^*$ by monotonicity. Therefore, monitoring is not optimal for any $\alpha \geq \bar{\alpha}$. Moreover, whenever a positive monitoring level is globally optimal, $E^* \geq \tilde{E}$ (see above). Accordingly, there must exist some $\tilde{\alpha} \in (0, \bar{\alpha})$ such that the respective enforceable levels of E described above are globally optimal whenever $\alpha \geq \tilde{\alpha}$, and $E^* = 0$ otherwise.

The propositions says that monitoring must be optimal whenever the parameterized monitoring costs are smaller than some positive threshold level. In the limit where monitoring is costless, the principal obviously can attain a first-best result by choosing $E^* = e^{FB}$. As monitoring becomes more costly, she will first decide upon a level of E that satisfies (9). For this level, the true effort $\underline{e}^M = E$ exerted by A and his effort requirement are identical. Provided that $\tilde{\alpha} < \underline{\alpha}$, there also exists an additional range in which a relatively small enforcement level that is implicitely determined by (8) is globally optimal. In this case, the agent's equilibrium effort $\underline{e}^M = e^*$ exceeds the enforceable level (the effort requirement) in equilibrium. Finally, for monitoring costs larger than the boundary $\tilde{\alpha}\,[< \bar{\alpha}]$, it becomes too costly to set up a monitoring technology, and the outcome of the overall game coincides with that in the standard model without monitoring.[24]

3 A Continuous Type Space

All previous results for the two-types example qualitatively carry over to any finite distribution of managerial abilities. In particular, it may still be optimal for the principal to induce a negative effort distortion when monitoring. In this section, we will briefly analyze the limit case where managerial types θ are drawn from a continuous distribution function $F(\theta)$ over the interval $[\underline{\theta}, \bar{\theta}]$. For convenience, we impose the monotone hazard rate condition $\partial[(1 - F(\theta))/f(\theta)]/\partial\theta \leq 0$ and suppose $\psi'''(e) \geq 0$ to ensure that the optimization program is well behaved [see, e.g., Laffont-

[24]While it is generally impossible to determine whether $\tilde{\alpha} > \underline{a}$ applies, one can easily find numerical specifications where that relation indeed holds. For example, for $\psi(e) = e^2/2$, $K(E) = \alpha E^3/3$, $\bar{\theta} = 1$, $\Delta\theta = 0.5$ and $p = 0.4$, one has $\tilde{a} \sim 1.79 > 2/3 = \underline{a}$. Hence, a positive monitoring level can be globally optimal even though it falls short of the agent's equilibrium effort for some interval of monitoring costs.

Tirole (1993)]. Clearly, the first best effort is unchanged relative to the discrete case. The second best effort under the standard screening contract is now determined by the first-order conditions[25]

$$\psi'(e^S(\theta)) = 1 - \frac{1-F(\theta)}{f(\theta)}\psi''(e^S(\theta)) \quad \forall\, \theta. \tag{10}$$

These conditions imply that a) the manager with the highest productivity $\bar{\theta}$ works efficiently under the optimal contract and b) the effort of less efficient managers is smaller than optimal and monotonically increasing in θ. Finally, the manager's informational rents are increasing in his type whith $U(\underline{\theta}) = 0$.

Let us now characterize the optimal effort contingent contract.[26] From our earlier arguments, we know that the utility of an agent of type θ is

$$U(\theta, \hat{\theta}) = t(\hat{\theta}) - \psi(\max\{E, \tilde{y}(\hat{\theta}) - \theta\}). \tag{11}$$

Restricting attention to piecewise differentiable mechanisms, the optimal announcement of agent θ when facing an effort-contingent contract offer C_M is determined by the first order conditions

$$\frac{\partial U(\theta, \hat{\theta})}{\partial \hat{\theta}} = 0 = t'(\hat{\theta}) - \begin{cases} 0 & \text{if } \tilde{y}(\hat{\theta}) - \theta = E \\ \psi'(y(\hat{\theta}) - \theta)y'(\hat{\theta}) & \text{if } y(\hat{\theta}) - \theta > E. \end{cases} \tag{12}$$

Incentive-compatibility then requires that (12) holds at $\hat{\theta} = \theta$, i.e., $\partial U(\theta, \theta)/\partial\theta = 0 \,\forall\, \theta :$ the manager's optimal report $\hat{\theta}$ must correspond to his true type θ under the second-best efficient compensation and output scheme. In contrast to the first-order condition in the standard model without monitoring, condition (12) takes into account that an agent θ may submit a report $\hat{\theta}$ which forces him to work just E and waste output [this happens if $E + \theta \geq y(\hat{\theta})$] when the principal's contract offer includes the minimum effort requirement $E^R = E$. Note that equilibrium effort must be increasing in θ. Therefore, it either is the case that each agent works more than E in equilibrium, or that a nonempty subset of agents with low abilities is induced to exert exactly E. Considering first the latter possibility, we can define

$$\dot{\theta}(E) = \sup\{\theta : y(\theta) - \theta = E\} \tag{13}$$

[25] See, for example, Laffont-Tirole (1993) in a setting where the agent's differ in intrinsic production costs.

[26] We confine ourselves to the case where $E < \psi^{-1'}(1)$, i.e. to a situation where a first-best result is not feasible.

as the boundary agent from the set $[\underline{\theta}, \dot{\theta}(E)]$ of agents who exert effort E. From (12), we see that incentive-compatibility requires these agents to obtain the same flat wage rate. Moreover, their equilibrium utility is zero, i.e. they do not earn a positive rent under the optimal contract. As a consequence, agents with higher productivity never have an incentive to submit a wrong report $\hat{\theta} \leq \dot{\theta}(E)$: mimicking a type from this interval forces any more productive agent to work just E, and no informational rent can be gained. Thus, the relevant incentive-compatibility constraint of any of these agents coincides with those under a standard contract [only the lower part of (12) matters], and with respect to the calculation of the equilibrium effort scheme everything is as if the lowest type $\underline{\theta}$ in the standard problem without monitoring is replaced by a type $\dot{\theta}(E)$.[27] As is well-known and shown in the appendix, the monotone hazard rate is invariant with respect to downward truncations of the distribution. As a result, the optimal contract for any type $\theta > \dot{\theta}(E)$ coincides with that in the standard model. In addition, it is easy to demonstrate that the boundary type $\dot{\theta}(E)$ for any given E is determined by $e^S(\dot{\theta}(E)) = E$, i.e., the boundary type under the optimal effort-contingent contract coincides with the agent who works exactly E under the standard contract.

Similar results apply when an agent with lowest ability works more than E under the optimal contract. Again, the informational rent of an agent is simply the integral over the rents of agents with lower productivity. Since, for any type close to $\underline{\theta}$, the minimum effort requirement E is also not binding, the principal cannot profitably reduce expected rents relative to the standard program. Expressed differently, provided that $e^S(\underline{\theta}) > E$ under the standard screening contract, an effort-contingency is of no help: it is optimal to induce every agent to work more than E.

The following proposition summarizes the preceding discussion:

Proposition 4 *Consider a continuous distribution of managerial abilities. Then:*

a) if $e^S(\underline{\theta})) > E$, i.e., if the least productive agent works more than E under the standard screening contract, the optimal effort-contingent contract duplicates the standard contract,

[27] The appendix provides an additional argument that is needed to establish this outcome: note that the principal can, when designing an incentive-compatible effort scheme, even ignore an agent's incentive to submit wrong reports larger than $\dot{\theta}(E)$ which would force the agent to exert the minimum effort E. Nevertheless, we show that this restriction on the agent's relevant strategy set does not decrease his equilibrium utility nor does it affect his effort level.

b) if $e^S(\underline{\theta})) \leq E$, the optimal effort-contingent contract is characterized by:

- *i) all agents $\theta \in [\underline{\theta}, \dot{\theta}(E)]$ provide the same effort $e^M(\theta) = E$ and do not obtain positive rents, i.e. $t(\theta) = \psi(E)$. The optimal contract is purely effort-dependent for all agents in this interval although their output increases in type with unit slope.*
- *ii) all agents $\theta \in (\dot{\theta}(E), \bar{\theta}]$ provide an effort which is equal to that under the standard contract, i.e., $e^M(\theta) = e^S(\theta)$. Compared to a situation without monitoring, the rents of these agents are reduced by the amount $\Delta R(\theta) = \psi(E) - \psi(e^S(\underline{\theta})) > 0$.*

In contrast to the discrete type case analyzed in the previous sections, the proposition states that productive efficiency weakly increases relative to the standard contract for any type of agent. Moreover, if the agent with lowest ability exerts an effort exceeding E under the standard screening contract, the possiblity to monitor neither affects productive efficiency nor the principal's payoff. When E becomes large, the availability of monitoring pushes the effort level of low types up to E without affecting the effort of higher types (although their rents are reduced). Hence, the outcome in the continuous case differs significantly from that in the discrete model. The following argument explains this difference intuitively: the rent of an agent is identical to the rent of his lower-ability neighbour, augmented by his marginal effort costs when picking the contract of this less efficient type. Consider a situation where, under a standard screening contract, every agent works more than E. Then, it is clear that no agent of a type above $\underline{\theta}$ faces a binding constraint $\hat{e}(\theta, \hat{\theta}) \geq E$: since the abilities of an agent slightly above $\underline{\theta}$ are just marginally higher than that of the worst type, he still would have to provide $e(\theta, \underline{\theta}) > E$ after mimicking. Moreover, since any more efficient agent compares his utility from truthtelling with that when mimicking the type just below him, the argument applies to the whole set of agents. Hence, in contrast to the discrete framework no rent-reduction effect can emerge; consequently, it can never pay for the principal to decrease the equilibrium effort of any type under an effort-contingent contract when differences in abilities across neighbouring types become negligible.

4 Discussion and Concluding Remarks

The present paper has shown that the ability of a principal to monitor an agent's effort may reduce productive efficiency. To our knowledge, this rather surprising result is new in the literature on adverse selection and monitoring. In line with existing results, monitoring abilities tend to decrease the informational rents of more efficient types of agents. Apart from the limiting case of a continuous type distribution, however, this decrease in absolute rents may be accompanied by an increase in marginal rents. As a consequence, the principal's basic tradeoff between rent extraction and productive efficiency may be tilted in favor of lower efficiency and lower rents relative to the standard model without monitoring.

We believe that the monitoring technology put forward in this paper bears empirical relevance. Courts often have difficulties to accurately verify the effort level of an agent, be it an employee or a contractor. Since judicial evidence on shirking is collected by the principal who is a self-interested party, courts must be careful when rating this evidence, and a point audit may not be sufficient to impose legal penalties. Therefore, random inspections will in many cases not be the proper instrument for a principal to control an agent. A continuous gathering of evidence may often be required to convince a court on an agent's shirking, in particular, if effort is provided over time. The role of those continuous audits is emphasized in the present paper.

Importantly, and in contrast to the monitoring models found in the literature, the paper postulates a connection between the feasible contractual effort specifications on the one hand, and the costs of monitoring on the other. Although the costs of monitoring clearly do not directly depend on the effort specifications agreed upon in a contract between principal and agent, there is an indirect channel through which such an interrelation may emerge: if the bilateral contract prescribes a small level of managerial effort, it may be easier for the principal to find verifiable evidence that the agent violates his contractual duties. For example, a part-time employee can be controlled at a smaller cost than a full-time worker since this control requires a smaller amount of time devoted to monitoring. Therefore, at given audit costs, monitoring may become more attractive in situations where the optimal (standard) contract assigns a low effort level to less efficient agents.[28] Conversely, when the effort induced under the standard contract does not substantially differ between

[28] In the discrete case, this is the case if p and/or $\Delta\theta$ are sufficiently large.

types of agents (which is optimal when types have similar abilities, or the ex ante probability to attract a bad agent is high), monitoring is less favorable because it calls for a high level of monitoring resources which have to be spent, as well as a relatively low rent reduction.

The effort-monitoring model analyzed by Laffont and Tirole (1992, 1993) suggests that an agent's deviation from *any arbitrary* contractually specified effort agreement is followed by the most extreme penalty. An increase of auditing in their model corresponds to a higher frequency of point observations, and one observation is sufficient to trigger the maximum penalty in case of shirking. Consequently, monitoring costs and precontracted effort are entirely disconnected, which is in stark contrast to the basic idea of the present approach.[29] Since either modeling strategy seems to be appropriate to mirror certain economic situations, it may thus be illuminating to learn that the fundamental outcomes differ significantly.

For a continuous interval of productivities, we found that monitoring induces pooling among low-productivity agents. The firm's owner offers identical and purely effort-contingent contracts to these types of managers. This result is notable since the pooled managers differ in productivity and hence produce different output levels. It is well in line with empirical observations which show that output-independent 'flat' labor or procurement contracts are very common in practice. In an employment context, workers or low-tier managers usually sign contracts that specify only their working hours. Similarly, government procurement contracts are often cost-based. Our model postulates the optimality of this contract type in situations where the principal perceives a low productivity of a manager or a regulated firm, respectively. Conversely, purely output-contingent incentive contracts turn out to be optimal if the agent's productivity is high. In the discrete case, we also found that the optimal contract under monitoring may combine input and output targets. The paper demonstrates that this mixed contractual form can be optimal even when the precontracted effort requirement is not binding for any type of agent. In this case, monitoring is still optimal since it can limit the incentives of high-type managers to shirk and effort-contingent labor arrangements can moderate the informational rents of these agents. Again, this conclusion may provide a rationale for seemingly odd contractual terms that can be observed in practice such as arrangements which combine output targets with input requirements although the contracting parties

[29] To our knowledge, this disconnection is also found in all other existing models on monitoring in adverse selection situations.

rationally expect an excess effort being necessary to reach the output goal.[30]

An interesting extension of the present model would be the introduction of a multi-dimensional effort variable. Because it is often impossible to make a contract contingent on quality dimensions of the agent's input (e.g., his mental engagement to work), this extension would provide a further motivation for partial monitoring and is an interesting topic for future research.

[30] While the labor contracts of many employees specify working hours, their actual working time frequently exceeds this precontracted level (good examples are employed lawyers, or secretaries in free-lance firms). In procurement transactions, suppliers often complain that input demands considerably increase in course of the procurement cycle; see Kovacic (1991).

Appendix

We will show that, under an effort-contingent contract, the equilibrium effort level of any agent with ability θ implicitly defined by $e^S(\theta) \geq E$ is identical to that under a standard contract. Conversely, all agents with lower productivity exert exactly E in equilibrium. For convenience, define $\ddot{\theta}(E) = \max\{\underline{\theta}, \dot{\theta}(E)\}$. Note first that, for any agent with abilities $\theta > \dot{\theta}(E)$, any report $\hat{\theta}$ from the interval $\hat{\theta} \in [\underline{\theta}, \dot{\theta}(E)]$ cannot be optimal since this report would deprive him from all rents. Noting that equilibrium effort and transfer payment of any agent $\theta > \ddot{\theta}$ is increasing in type, we can also ignore announcements where E becomes strictly binding for agent $\theta > \dot{\theta}(E)$: any such report would force him to exert E, while his transfer payment is strictly increasing in the announced type. Formally, we can define the boundary announcement as

$$\tilde{\theta}(\theta, E) \equiv \sup\{\hat{\theta} : y(\hat{\theta}) - \theta = E \,|\theta > \dot{\theta}(E)\}. \tag{14}$$

Note that $\tilde{\theta}(\theta, E) > \dot{\theta}(E)$, since the true productivity of agent θ is higher than $\dot{\theta}$. We will now show that, although the relevant interval of announcements for agent θ is narrowed down to the range $[\tilde{\theta}(\theta, E), \theta]$, his equilibrium utility under an incentive compatible contract is identical to that when the relevant support would be the broader interval $[\ddot{\theta}(E), \theta]$. To verify this claim, consider the utility level that has to be conceded to agent θ to prevent him from submitting a report $\tilde{\theta}(.)$, which is [recall that the agent with true type $\tilde{\theta}(.)$ works more than E by construction]

$$U(\theta, \tilde{\theta}(\theta, E)) = \int_{\tilde{\theta}(\theta,E)}^{\theta} \psi'(e(\tilde{\theta}))d\tilde{\theta} + U(\tilde{\theta}(\theta, E)). \tag{15}$$

Define a sequence $\{\theta^t\}_0^\infty$ where $\theta^t = \tilde{\theta}(\theta^{t-1}, E)$ and a start value $\theta^0 = \theta$. Observe that

$$U(\tilde{\theta}(\theta, E)) = \int_{\tilde{\theta}(\tilde{\theta}(\theta,E),E)}^{\theta(\tilde{\theta},E)} \psi'(e(\tilde{\theta})d\tilde{\theta} + U(\tilde{\theta}(\tilde{\theta}(\theta, E), E)). \tag{16}$$

The sequence defined above yields $\theta^t \to \max\{\dot{\theta}(E), \underline{\theta}\}$ as $t \to \infty$; hence, we can solve the utility function recursively to obtain

$$U(\theta) = \int_{\ddot{\theta}(E)}^{\theta} \psi'(e(\tilde{\theta}))d\tilde{\theta} = \psi(e(\theta)) - \max\{\psi(E), \psi(e(\underline{\theta})\}. \tag{17}$$

As a result, the agent's utility is identical to that under a standard contract with lower boundary $\ddot{\theta}(E)$.

We will now prove the claim that, provided that the distribution function is characterized by the monotone hazard rate condition, the equilibrium effort of any type above $\ddot{\theta}(E)$ coincides with that under a standard contract. To do so, we calculate the expected informational rents to be paid by the principal under a menu of incentive-compatible contracts. Integrating the lower part of (12) yields the informational rent of manager θ, $\theta > \ddot{\theta}(E)$ which is identical to (18). The expected informational rents for the principal are therefore calculated as

$$\begin{aligned}
\mathcal{E}_\theta U(\theta) &= \int_{\ddot{\theta}(E)}^{\bar{\theta}} \int_{\ddot{\theta}(}^{\theta} \psi'(e(\tilde{\theta}))d\tilde{\theta} f(\theta)d\theta \\
&= F(\theta) \int_{\ddot{\theta}(E)}^{\theta} \psi'(e(\tilde{\theta}))d\tilde{\theta}\,]_{\ddot{\theta}(E)}^{\bar{\theta}} - \int_{\ddot{\theta}(E)}^{\bar{\theta}} \frac{\partial U(\theta)}{\partial \theta} F(\theta)d\theta \\
&= [\psi(e(\bar{\theta})) - \psi(e(\ddot{\theta}(E))] - F(\ddot{\theta}(E))[\psi(e(\ddot{\theta}(E))) - \psi(e(\ddot{\theta}(E)] - \int_{\ddot{\theta}(E)}^{\bar{\theta}} \frac{\partial U(\theta)}{\partial \theta} F(\theta)d\theta \\
&= \int_{\ddot{\theta}(E)}^{\bar{\theta}} \psi'(e(\theta))d\theta - \int_{\ddot{\theta}(E))}^{\bar{\theta}} \psi'(e(\theta))F(\theta)d\theta \\
&= \int_{\ddot{\theta}(E)}^{\bar{\theta}} \psi'(e(\theta)) \frac{1-F(\theta)}{f(\theta)} f(\theta)d\theta. \qquad (18)
\end{aligned}$$

Accordingly, the principal's optimation program reads

$$\begin{aligned}
P_M &= \int_{\ddot{\theta}(E)}^{\bar{\theta}} [(\theta + e(\theta)) - \psi(e(\theta)) - \psi'(e(\theta))\frac{1-F(\theta)}{f(\theta)}]f(\theta)d\theta \\
&+ \int_{\underline{\theta}}^{\ddot{\theta}(E)} [(\theta + E) - \psi(E)]f(\theta)d\theta. \qquad (19)
\end{aligned}$$

Pointwise differentiation with respect to $e(\theta)$ yields the first-order condition [31]

$$\psi'(e(\theta)) = 1 - \frac{1-F(\theta)}{f(\theta)}\psi''(e(\theta)) \quad \forall \quad \theta > \max\{\dot{\theta}(E), \underline{\theta}\}, \qquad (20)$$

while managers of type $\theta \leq \ddot{\theta}(E)$ provide an effort identical to the monitoring level E. Condition (20) establishes that the necessary first-order conditions for the equilibrium effort levels of any agent $\theta > \ddot{\theta}(E)$ coincide with those under a standard contract.

Next, we demonstrate that $\dot{\theta}(E)$ is implicitly defined by $e^S(\dot{\theta}(E)) = E$. Assume first that the boundary type $\dot{\theta}(E)$ would be more efficient, i.e., $e^S(\dot{\theta}(E)) > E$. Then,

[31] We can neglect the constraint $e'(\theta) \geq 0$ in program (19) which is always fulfilled if $\psi'''(e) \geq 0$.

it must be optimal to increase the effort level of type $\dot{\theta}$ up to $e^S(\dot{\theta})$ which increases productive efficiency while having only a second-order effect on the rents of the set of more productive agents. Conversely, $e^S(\dot{\theta}(E)) < E$ cannot be optimal since the principal could increase productive efficiency at no costs in terms of additional informational rents.

Taken together, the main statements in Proposition 4 follow.

Finally, we will prove our claim that the informational rents of all agents above $\dot{\theta}(E)$ uniformly decrease relative to the rents under a standard contract. Recall that an optimal annoucement of manager θ must fulfill the first-order condition (12). Accordingly, and invoking the revelation principle, an incentive compatible contract is characterized by

$$\frac{\partial U(\theta,\theta)}{\partial\theta} = \begin{cases} 0 & \text{if } E = y(\theta) - \theta = e(\theta) \\ \psi'(y(\theta) - \theta) = \psi'(e(\theta)) & \text{if } E < y(\theta) - \theta = e(\theta) \end{cases} = 0 \tag{21}$$

Using standard arguments, we can now integrate (22) to obtain

$$U(\theta) = \begin{cases} const. & \text{if } e(\theta) = E \\ const. + \int_{\dot{\theta}(E)}^{\bar{\theta}} \psi'(e(\theta)) & \text{if } e(\theta) > E. \end{cases} \tag{22}$$

Note that $U(\theta) = 0$ is optimal whenever $e(\theta) = E$. When $e(\theta) > E$, the utility level of any agent corresponds to that in the standard model where the upper boundary $\underline{\theta}$ is replaced by $\dot{\theta}(E)$. Accordingly, the informational rent of these agents is uniformly diminished by

$$\int_{\underline{\theta}}^{\dot{\theta}(E)} \psi'(e^S(\theta))d\theta = \psi(E) - \psi(e^S(\underline{\theta}). \tag{23}$$

which establishes our result. □

References

Aghion, P. and J. Tirole (1994): Formal and Real Authority in Organizations, Discussion Paper, University of Toulouse.

Baron, D. and D. Besanko (1984): Regulation, Asymmetric Information, and Auditing, *Rand Journal of Economics 15*, 447 - 470.

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

Dittmann, I. (1996): On Agency and Optimal Auditing, Discussion Paper Nr. 96-05, University of Dortmund.

Kofman, F. and Jaques Lawarrée (1993): Collusion in Hierarchical Agency, *Econometrica 61*, 629 - 656.

Kovacic, W.E. (1991): Defense Contracting and Extensions to Price Caps, *Journal of Regulatory Economics 3*, 219-240.

Laffont, J.-J. and J. Tirole (1986): Using cost Observation to Regulate Firms, *Journal of Political Economy 94*, 614-641.

Laffont, J.-J. and J. Tirole (1992): Cost Padding, Auditing and Collusion, *Annales d'Économi et Statistique 25-26*, 205 - 226.

Laffont, J.-J. and J. Tirole (1993): A Theory of Incentives in Procurement and Regulation, MIT Press.

Melumad, N.D. and D. Mookherjee (1989): Delegation as Commitment: The Case of Income Tax Audits, *Rand Journal of Economics 20*, 139-163.

Nalebuff, B. and D. Scharfstein (1987): Testing in Models of Adverse Selection, *Review of Economic Studies 54*, 265 - 277.

Polinsky, A.M. and S. Shavell (1979): The Optimal Tradeoff Between the Probability and Magnitude of Fines, *American Economic Review 69*, 880 - 891.