Screening and the positive correlation between risk and incentives[*][†]

Dezsö Szalay

September 17, 2010

Abstract

Empirical findings of a positive correlation between risk and incentives have challanged the traditional agency model. Relying on a trade-off between risk sharing and incentive provision, the traditional model gives rise to a negative correlation. I develop two simple models of screening, where the agent has private information about the distribution of returns and show that a positive correlation arises naturally in these models.

JEL Classification: D82, J33

1 Introduction

The CARA-normal model of linear contracts is a corner stone of modern incentive theory. Its basic structure is simple. A risk neutral principal shares profits with a risk averse agent; the agent's utility displays constant absolute risk aversion. The profit depends on the agent's effort and a normally distributed random factor. The principal can only observe profit, but not the agent's effort. A first-best allocation of risk requires that the principal bears all the risk. However, if the principal bears all the risk, the agent has no incentive to work. Consequently, the second-best optimal profit sharing arrangement shifts too much risk to the agent. The cost of this inefficiency

---

[*] University of Bonn, Dept of Economics, Adenauerallee 24-42, 53113 Bonn, Germany. email: szalay@uni-bonn.de

[†] I thank Mark Armstrong, Charles Blackorby, Yeon-Koo Che, Benoit Crutzen, Piero Gottardi, Stephen Hansen, Massimo Morelli, Jean-Charles Rochet, Fabiano Schivardi, Bauke Visser as well as seminar participants at EIEF Rome, EUI Florence, the Tinbergen Institute in Rotterdam, and UPF in Barcelona for comments. All errors are mine.

is the larger the greater the agent's distaste for risk. Therefore, the agent's profit share is the smaller the greater is risk.

The model's comparative statics prediction lends itself to empirical investigation. However, the empirical evidence on the validity of the incentive contracting model is mixed. Prendergast (2002) provides a survey of empirical studies of the risk sharing/incentive predictions. Studies that test the dependence of the power of executives' performance pay on measures of business risk yield mixed results. More strikingly, studies that use data on sharecropping (Rao (1971), Allen and Lueck (1992, 1995)) find a *positive* dependence of measures of output variability on farmers' output shares.

I offer two simple theories that generate a positive correlation between risk and incentives; the essential assumptions are that the agent has some private information about his ability and that the principal has a preference for low risk. The first assumption is self-evident. The second assumption captures the following situations. Obviously, the principal has a preference for low risk if he is risk averse. A risk neutral principal has a preference for low risk if he is facing costs of financial distress that arise if the firms profits fall short of a minimum level. Assuming the principal is risk averse makes sense where principal and agent are demographically similar, e.g. when the principal is a landlord and the agent a tenant, or a senior and a junior partner in a partnership. When it comes to large firms, we usually think of principals as of a group of well diversified shareholders, so they are (approximately) risk neutral. So, for the analysis of optimal contracts between shareholders and CEOs, the model with a risk neutral principal and costs of financial distress is more reasonable. I now describe the economics in the two models in more details.

Suppose a senior lawyer seeks to employ a junior partner. The senior's contribution is the name of the law firm and the client list. The junior is supposed to solve cases. Both partners are strictly risk averse with utility functions displaying constant absolute risk aversion. The senior offers the junior a compensation with a base salary plus a share of the profits the junior will generate. No effort is required to generate profits; instead profits are equal to the junior's talent plus a normally distributed noise term[1]. Only the junior partner knows his talent; the senior partner only knows

---

[1] This formulation should not be taken literally, but instead as saying that the variation in effort is less important than the variation in talent for explaining the variation in output. One reason is that effort does not vary a lot, e.g., because the agent is hard-working regardless of how steep incentive contracts are, possibly to increase the likelihood of being promoted to a senior status later on. Another justification is when the marginal impact of an increase in talent on output is simply much larger than the marginal impact of effort. I take these justifications to the extreme

the distribution of talent in the population.

This model gives rise to a trade-off between risk sharing and rent extraction. Ideally, the senior partner would like to offer a profit sharing arrangement where the shares of profits correspond to the junior's and the senior's risk bearing capacity relative to the aggregate risk bearing capacity, irrespective of the expected return that the junior can generate. However, the more talented juniors would receive large rents under such a contract, because they are more likely to generate higher profits. To extract these rents, the senior partner distorts the profit shares of the junior partner downwards for all but the most able type.[2] Thus, for all but the most able junior type, the junior's share of profit is below the first-best share of profits. The cost of this arrangement is that the senior has to offer higher base salaries to make the junior willing to participate. The utility cost of this misallocation of risk is the larger the higher is risk. Hence, the senior partner reduces the inefficiency when the underlying business risk becomes larger. However, reducing the gap between the second-best and the first-best risk allocation means to increase the profit share for all but the most able junior types. Hence, a positive correlation between risk and the power of the incentive scheme is the natural consequence.

With a risk neutral principal, the simple risk sharing versus rent extraction trade-off cannot account for the positive correlation between risk and incentives in the data. A risk neutral principal can implement the first-best by shielding the agent from all risk. Since the first-best risk sharing arrangement is independent of the underlying risk, this model would predict that risk and incentives are uncorrelated. So, different forces seem to be at play when risk and incentives are positively correlated in a sample of CEOs working for different firms.

Suppose a risk neutral principal dislikes risk because a higher risk makes it more likely that profits are very low, below a critical value that leads to financial distress. The agent-manager influences the distribution of profits in various ways. First, he increases the mean of the profit by exerting effort and by being more talented. Second, he chooses the riskiness of the profit distribution by choice of his management style. Manager's differ both with respect to the mean ability and with respect to their costs of implementing a low risk strategy and have private information with respect to these parameters. Hence, the principal faces a problem of two dimensional screening. Even though the principal is able to induce different risk choices from managers who differ in their cost of implementing safe projects, the principal chooses to bunch the profit shares of

---

and abstract from the effort choice in this model altogether.

[2] So far, everything is completely standard; that contracts serve to screen (or signal) talent is known since Leland and Pyle (1977). The new element is the comparative statics predictions of the model.

agents together who have the same mean talent but differ in their costs of reducing the variance of returns. Separation would be feasible but its costs exceed the benefits. As a result, the profit share of a manager of given mean talent is optimal against the marginal distribution of mean talent. Increasing the fraction of managers with high costs of reducing variance increases the equilibrium risk in the economy and at the same time affects the marginal distribution of mean talents in a way that makes larger profit shares optimal. Hence, a positive correlation between risk and incentives arises naturally again.[3]

In the present paper, contracts play a double role of providing incentives and sreening agents of different abilities. Lazear (1986) has already pointed out that performance contracts have selection effects: contracts providing a given strength of incentives are more attractive to more able workers. My contribution is to the analysis of the comparative statics properties of such sorting models.

The majority of the incentive models in the literature feature a negative correlation between risk and incentives. Only a handful of theories can rationalize a negative correlation. Prendergast (2002) was the first to point to this puzzle in the literature and to offer an explanation for it[4]. Prendergast (2002) argues the standard theory neglects an endogenous delegation decision. Suppose there are two essential inputs in production, effort and information that is used to make decisions, and suppose that agents have better information than principals. The value of this improved information is the larger the more uncertain the environment. Consequently, the larger is business risk, the more likely are principals to delegate decision making to the agent. But to ensure that the agent acts in the principal's interest, the principal makes the agent's pay depend on his performance. Hence, the agent's pay is the more dependent on performance the higher is risk. Thus, essentially Prendergast argues that the existing theories, and their empirical tests, suffer from an omitted variable bias.

Raith (2003) argues that empirical tests of the principal agent model fail to distinguish variability in profits and measurement error in contracting. If this distinction is made, then a positive correlation of performance pay and business risk can be rationalized. In particular, he studies

---

[3] This problem of multi-dimensional screening is related to a recent strand in the literature on screening. Che and Gale (2000), Iyengar and Kumar (2006), Malakov and Vohra (2005), Beaudry, Blackorby, and Szalay (2009), and Szalay (2009) look at two dimensional problems where agents have only one sided deviations available in one of the dimensions. Most results known so far apply to the case where the constrainst ruling out such one sided deviations are slack. The main exception (to my knowledge) is my own work Szalay (2009), where I treat the case of one sided deviations with binding constraints extensively. In this paper I solve the case of two sided deviations in both dimensions.

[4] See also Prendergast (1999) for a survey of the literature.

a model of oligopolistic competition, where a manager's role is to reduce his firm's costs of production. Like in the traditional model, the dependence of the manager's pay on realized cost reductions is the smaller the larger is the measurement error for these same cost reductions. On the other hand, uncertainty about rivals' costs makes firms' profits stochastic. Although the power of the manager's performance pay and the variability of firms' profits are not causally linked to each other, a change in a third factor, e.g. the degree of competition, increases both profit risk and the power of the manager's incentive scheme. Thus, the agent's pay is again more performance dependent when business risk is greater, but there is no causal link between the two effects.

Inderst and Müller (2009) develop a model of managerial entrenchment featuring a positive correlation between risk and incentives. A manager with private information as to whether continuation is profitable or not must be given incentives to use this information in the owner's interest. Incentive contracts are used to align the incentives of owner and manager with respect to the continuation decision, and the incentive contract is the steeper the higher is the underlying cash-flow risk.

The present paper complements these approaches; since only a handful of theories are consistent with a positive correlation between risk and incentives, additional explanations seem still useful at this point. I stay firmly within the confines of the linear contracting, CARA-normal model and show that small variations around the standard model allow me to rationalize a positive correlation between risk and incentives. Thus, empirical findings of a positive correlation do not allow us to question the validity of the incentive model; however, they do allow us to reject particular specifications of the model; that is, they allow us to reject the model of pure moral hazard, but do not allow us to reject models of combined moral hazard and adverse selection.

An interesting strand of the recent literature emphasizes omitted selection biases in the empirical contracting literature. Ackerberg and Botticini (2002) argue that the mostly negative conclusions in the early empirical work on sharecropping is due to an omitted matching process; roughly agents who differ with respect to their risk aversion are matched to principals who differ with respect to the riskiness of the business they operate in. Taking account of this matching they show that their data are consistent with the predictions of their incentive model. Bandiera et al. (2009) follow up on this idea of endogenous matching, develop a theory of endogenous matching between heterogenous managers and different type of firms - family firms versus public firms - and show that all the predictions of their model are confirmed in their data. Although matching and screening are quite different issues, the prediction they generate are largely the same. More able agents get steeper incentive contracts, exert more effort, and obtain higher expected utility; these are the predictions of the matching model, exactly the predictions that one obtains from the present model of screening. Hence, a fruitful avenue for future research would be to integrate both theories.[5]

The paper is structured as follows. In section 2 I briefly review the standard linear contracting model and show that it cannot account for a positive correlation between profit shares and underlying business risk. In section 3 I study a simple two type model of adverse selection. In section 4 I offer its generalization to the case of a continuum of types. In section 5, I study the second model with two-dimensional types. I have relegated all proofs to an appendix.

2 Review of the Classical Model

In this section I review the "reduced form" version of the Holmström-Milgrom (1987) model, where the agent's effort controls the drift and diffusion of a continuous time stochastic process. The central finding of Holmström and Milgrom (1987) is a linearity result stating that the optimal contract can be expressed as an affine function of some accounting statistic that relates to the agent's output (usually profit)[6]. Central assumptions are that the principal's and the agent's absolute risk aversion is constant and the output (profit-) follows a Brownian motion.

These basic results have been applied in a large body of research on contracts. The standard approach in this literature takes a static reduced form of the dynamic model assuming CARA preferences, that the profit follows a normal distribution, and that the agent's pay is a linear function of profit[7]. The purpose of this section is to show that this set of assumptions generates the robust prediction that the agent's share of profits is a decreasing function of the variance of the profit distribution.

The agent produces a profit π whose distribution depends on the agent's effort choice, e. The return follows a normal distribution with mean equal to e and variance equal to σ^2, i.e., $\tilde{\pi} \sim N\left(e, \sigma^2\right)$. If the agent exerts effort e, he bears a cost of effort $c\left(e\right)$. The cost function satisfies

---

[5] On a more technical side this paper is related to the literature on screening with risk averse agents; see Salanié (1990), Laffont and Rochet (1998), Matthews (1983), and Maskin and Riley (1984); for papers in finance that are related in this technical sense, see Biais et al. (2000) and Garcia (2008). The novel element of my first model relative to this literature is that both principal and agent are strictly risk averse.

[6] See Hellwig and Schmidt (2002) for discrete time approximations of this model.

[7] Obviously, although linear contracts are optimal in the dynamic model, they are not optimal in the static model.

$c_e(e) \geq 0$ and $c_{ee}(e) > 0$ to make the agent's maximization problem concave in effort[8]. The agent's effort choice is unobservable to the principal. The principal can only observe π and offers contracts that are linear in profits. The agent's salary is equal to $W^A = \beta + \alpha\pi$. The principal receives $W^B = -\beta + (1-\alpha)\pi$. To render the principal's problem of contract design concave in the parameters of the contract, I assume that the cost-of-effort function satisfies $c_{eee}(e) \geq 0$.

The salary the agent can obtain in an alternative employment is equal to ω. The principal and the agent have von Neumann-Morgenstern utility functions, that display constant absolute risk aversion. The principal has the utility function

$$V\left(W^B\right) = -\exp\left[-bW^B\right]$$

where b measures his degree of constant absolute risk aversion. In the limiting case where b goes to zero, the principal becomes risk neutral; in that case I take his objective as $V\left(W^B\right) = W^B$.

The agent's utility function is defined on wealth net of costs of effort

$$U\left(W^A, e\right) = -\exp\left[-a\left(W^A - c(e)\right)\right]$$

where a measures the agent's degree of absolute risk aversion.

Expected utilities have a convenient certainty equivalent representation. Let E denote the expectation operator. One can write $EU\left(W^A, e\right) = U\left[\left(w^A(\cdot) - c(e)\right)\right]$ and $EV\left(W^B\right) = V\left(w^B(\cdot)\right)$ where the certainty equivalent levels of wealth, $w^A(\cdot)$ and $w^B(\cdot)$ satisfy

$$w^A\left(\alpha, \beta, e, \sigma^2\right) = \beta + \alpha e - a\frac{\alpha^2}{2}\sigma^2 \tag{1}$$

and

$$w^B\left(\alpha, \beta, e, \sigma^2\right) = -\beta + (1-\alpha)e - b\frac{(1-\alpha)^2}{2}\sigma^2 \tag{2}$$

respectively[9]. The certainty equivalent wealth levels are completely described by the moments of the normal distribution and the parameters of the contract.

The principal's problem is to write a contract that maximizes his expected utility such that the agent is willing to exert the effort level the principal wants to induce (constraint IC below) and

---

[8] Throughout the paper, subscripts denote derivatives of functions with respect to their argument.

[9] Throughout the paper, I suppress arguments of functions that are kept constant - as a and b are in $w^A(\cdot)$ and $w^B(\cdot)$, respectively.

that the agent is willing to participate (constraint IR). Formally, his problem is the following:

$$\max_{\alpha,\beta} V\left(w^{B}\left(\alpha,\beta,e,\sigma^{2}\right)\right) \tag{P}$$

$$s.t.$$

$$w_{e}^{A}\left(\alpha,\beta,e,\sigma^{2}\right)=c_{e}\left(e\right) \tag{IC}$$

$$w^{A}\left(\alpha,\beta,e,\sigma^{2}\right)-c\left(e\right)=\omega. \tag{IR}$$

Given that $c_{eee}\left(e\right)\geq 0$, this problem is strictly concave in α. It has a unique solution, that I denote α^{*} and state formally in the following proposition:

Proposition 1 *In the CARA-linear-normal model with effort subject to moral hazard, the agent's optimal share of profits is given by*

$$\alpha^{*}=\frac{b\sigma^{2}c_{ee}\left(c_{e}^{-1}\left(\alpha^{*}\right)\right)+1}{\left(a+b\right)\sigma^{2}c_{ee}\left(c_{e}^{-1}\left(\alpha^{*}\right)\right)+1}$$

The optimal share is the smaller the higher is business risk; formally

$$\frac{d\alpha^{*}}{d\sigma^{2}}<0.$$

The intuition for this result is quite simple. If effort were contractible, the principle would offer a first-best risk sharing to the agent. The principal and the agent, respectively, would bear a fraction of risk that corresponds to their risk bearing capacity relative to the aggregate risk bearing capacity, so $\alpha^{fb}=\frac{\frac{1}{a}}{\frac{1}{a}+\frac{1}{b}}$. It is easy to see that $\alpha^{*}>\alpha^{fb}$, so that the agent bears too much risk. The agent would have incentives to choose the efficient level of effort if $\alpha=1$. So, the solution is a compromise between efficient risk sharing and efficient provision of incentives. Straightforward calculus shows that the optimal share of profits is the smaller the higher is σ^{2}. The higher is σ^{2}, the higher is the cost of the inefficient allocation of risk. In other words, the higher is the fixed part of the compensation needed to make the agent participate for any given share in profits. Therefore, the principal decreases the agent's share in profits as σ^{2} is increased. Thus, a model of pure moral hazard cannot explain a positive correlation between risk and incentives.

I now show that a simple model of incentive compatible risk sharing generates a positive correlation between risk and incentives. I will do so restricting attention to the same class of contracts, that is, linear contracts, as the model above has.[10]

---

[10] Similar to the Holmström-Milgrom (1987) result for pure moral hazard, Sung (2005) provides a continous time justification for linear contracts in a model with adverse selection and moral hazard. However, his model assumes a risk neutral principal. So his justification applies to the model I study in section 5 below; for the partnership

3 A partnership-model with two types

3.1 The model

Let $\tilde{\pi} \sim N\left(\theta, \sigma^2\right)$, where θ measures the agent's ability or competence. The agent knows his ability but the principal does not. The principal only knows the distribution of the the agent's ability; that is the principal knows that $\theta = \overline{\theta}$ with probability λ and $\theta = \underline{\theta}$ with probability $1 - \lambda$, where $\underline{\theta} > \overline{\theta} > 0$. σ^2 is exogenous and known to the agent and the principal.

In contrast to the moral hazard model, it is now optimal to offer a menu of contracts. The principal offers a contract $\overline{\alpha}, \overline{\beta}$ to the high ability type and a contract $\underline{\alpha}, \underline{\beta}$ to the low ability type. The optimal menu of contracts solves the following problem:

$$\max_{\overline{\alpha},\overline{\beta},\underline{\alpha},\underline{\beta}} \lambda V\left(w^B\left(\overline{\alpha}, \overline{\beta}, \overline{\theta}, \sigma^2\right)\right) + (1-\lambda) V\left(w^B\left(\underline{\alpha}, \underline{\beta}, \underline{\theta}, \sigma^2\right)\right) \tag{P^2}$$

$$s.t.$$

$$w^A\left(\overline{\alpha}, \overline{\beta}, \overline{\theta}, \sigma^2\right) \geq w^A\left(\underline{\alpha}, \underline{\beta}, \overline{\theta}, \sigma^2\right) \tag{$\overline{IC}$}$$

$$w^A\left(\underline{\alpha}, \underline{\beta}, \underline{\theta}, \sigma^2\right) \geq w^A\left(\overline{\alpha}, \overline{\beta}, \underline{\theta}, \sigma^2\right) \tag{$\underline{IC}$}$$

$$w^A\left(\overline{\alpha}, \overline{\beta}, \overline{\theta}, \sigma^2\right) \geq \omega \tag{$\overline{IR}$}$$

$$w^A\left(\underline{\alpha}, \underline{\beta}, \underline{\theta}, \sigma^2\right) \geq \omega, \tag{$\underline{IR}$}$$

where $\overline{IC}$ and $\underline{IC}$ are the incentive constraints for the high and the low ability type respectively; each of these types should prefer to accept the contract designed for himself rather than the contract the principal offers to the other type. $\overline{IR}$ and $\underline{IR}$ are the participation constraints of both types. For $\underline{\theta}$ sufficiently high it is indeed optimal to offer contracts that induce participation by both types. Formally, I solve this problem under the following assumption:

Assumption 1: $\underline{\theta} - \max\{a, b\} \frac{\sigma^2}{2} > \omega$.

Assumption 1 guarantees that it is always optimal to hire all agents; the certainty equivalent generated when the least able agent is hired and risks are shared in the least efficient way is still greater than the certainty equivalent generated in the agent's outside activity. If the condition in assumption I fails to hold, then it may become optimal to exclude the less competent agent. The analysis of this case is straightforward but less interesting than the one I present.

If the principal knew the agent's ability, then he would offer first-best optimal risk sharing

---

model in sections 3 and 4, I restrict attention to linear contracts for reasons of comparability with respect to the benchmark model in section 2. A further advantage is that in the linear model, there is exactly one measure that captures the power of the incentive scheme.

arrangements to both types, $\overline{\alpha}^* = \underline{\alpha}^* = \frac{b}{a+b}$, and would set the base pay levels of both agents so as to extract all rents from both types. Formally, imposing $\overline{IR}$ and $\underline{IR}$ as binding and using (1), I obtain $\overline{\beta}^* = -\overline{\alpha}^*\overline{\theta} + \frac{a(\overline{\alpha}^*)^2\sigma^2}{2} + \omega$ and $\underline{\beta}^* = -\underline{\alpha}^*\underline{\theta} + \frac{a(\underline{\alpha}^*)^2\sigma^2}{2} + \omega$. Observe that $\overline{\beta}^* < \underline{\beta}^*$; since the high ability agent generates a return distribution that dominates the return distribution of a low ability agent in the sense of First Order Stochastic Dominance, the principal can offer a lower base pay level to the high ability type.

Obviously, this arrangement is not incentive compatible when ability is only known to the agent but not to the principal. The high ability type can get a higher total payment by claiming to be of low ability, because that raises his base pay without affecting his share of profits. Therefore, on top of the base pay, the principal has to make the profit shares dependent on the agent's ability. I now turn to solve the principal's problem under asymmetric information.

3.2 Incentive compatible risk sharing

As is easy to show (see the appendix), at the optimum $\overline{IC}$ and $\underline{IR}$ hold as equalities, while the other two constraints are slack. Imposing $\underline{IR}$ as an equality, I obtain the base pay level of the low type

$$\underline{\beta}\left(\underline{\alpha}\right) = \omega - \underline{\alpha}\underline{\theta} + a\frac{\underline{\alpha}^2}{2}\sigma^2. \tag{3}$$

Substituting (3) into $\overline{IC}$ I obtain the base pay level of the high type as a function of the shares $\overline{\alpha}$ and $\underline{\alpha}$

$$\overline{\beta}\left(\overline{\alpha}, \underline{\alpha}\right) = \omega - \overline{\alpha}\overline{\theta} + a\frac{\overline{\alpha}^2}{2}\sigma^2 + \underline{\alpha}\left(\overline{\theta} - \underline{\theta}\right). \tag{4}$$

Substituting these expressions into the principal's objective function, I obtain the following maximization problem in the shares $\overline{\alpha}$ and $\underline{\alpha}$.

$$\max_{\underline{\alpha}\geq 0, \overline{\alpha}\geq 0} \lambda V\left(w^B\left(\overline{\alpha}, \overline{\beta}\left(\overline{\alpha}, \underline{\alpha}\right), \overline{\theta}, \sigma^2\right)\right) + (1-\lambda)\, V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right) \tag{5}$$

The problem is strictly concave in the shares, so the solution is obtained by taking first-order conditions. I get

$$\overline{\alpha}^* = \frac{b}{a+b}$$

and

$$\underline{\alpha}^* = \max\left\{0, \frac{b}{a+b} - \frac{\lambda}{1-\lambda}\frac{V\left(w^B\left(\overline{\alpha}, \overline{\beta}\left(\overline{\alpha}, \underline{\alpha}\right), \overline{\theta}, \sigma^2\right)\right)}{V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right)}\frac{\overline{\theta} - \underline{\theta}}{\sigma^2}\right\},$$

where $\underline{\alpha}^* > 0$ for sufficiently low λ.

The high ability type is offered a first-best risk sharing arrangement. On the other hand, the low ability type's share of profits is smaller than the first-best share. If he obtained the first-best share, then the high ability type would capture all the rents arising from his high ability. To extract some of these rents, the principal distorts the low ability type's share downwards. Although the base pay of the low ability type is higher than the base pay of the high ability type, the low ability type receives a smaller share in profits. As a high ability type is more likely to generate high returns this is particularly unattractive for an able type. Hence, the able type is discouraged from mimicking the low ability type this way. The downwards distortion of the low ability type's share is the larger the higher is λ; the higher is λ the more important it is for the principal to extract rents from the high ability type and the less important it becomes to share risks efficiently with the low ability type. Hence, if the fraction of high types becomes large enough, the principal just offers the constant outside wage to the low ability type and bears all the risk himself.[11]

3.3 Comparative Statics

Consider now the comparative statics of the optimal risk sharing arrangement with respect to changes in σ^2 for the case where the principal offers an interior share to the low type. Since $\overline{\alpha}^*$ is equal to the agent's first-best share in profits, and the first-best risk sharing arrangement is independent of the level of risk, $\overline{\alpha}^*$ is independent of σ^2. On the other hand, $\underline{\alpha}^*$ depends on σ^2 in various ways. To determine the balance of the various effects requires some care; I have relegated the details of the calculations to the appendix. However, a simple intuition goes as follows:

The optimal risk sharing arrangement with the low ability type shifts too little risk to the agent relative to the first-best risk sharing arrangement, $\underline{\alpha}^* < \frac{b}{a+b}$. This allows the principal to extract some rents from the high ability type, but it obviously also involves a cost in terms of inefficient risk sharing. The departure from optimal risk sharing is the more costly the higher is risk. Hence, as risk is increased, the principal finds it relatively less important to extract rents and relatively more important to share risks efficiently. Hence, the higher is σ^2 the higher is $\underline{\alpha}^*$, moving closer to its first-best level.[12] I state this result formally as

Proposition 2 *In the CARA-linear-normal model with ability subject to adverse selection, the able agent's optimal share of profits is independent of business risk and the incompetent agent's share is non-decreasing in business risk. Whenever the incompetent able receives a strictly positive*

---

[11] Clearly, offering a constant wage is stricly better than offering a negative share of profits to the agent, which would involve a huge inefficiency in terms of the allocation of risks. This is explained in more detail in the continuum version of the model below.

[12] Formally, the simple intuition is incomplete because the ratio $\frac{V\left(w^B(\overline{\alpha},\overline{\beta}(\overline{\alpha},\underline{\alpha}),\overline{\theta},\sigma^2)\right)}{V\left(w^B\left(\underline{\alpha},\underline{\beta}(\underline{\alpha}),\underline{\theta},\sigma^2\right)\right)}$ depends on σ^2 as well. See the formal proof in the appendix for the complete argument.

share of profits, then

$$\frac{d\underline{\alpha}^*}{d\sigma^2} > 0.$$

4 The partnership model with a continuum of types

I now generalize the two type model to a continuum of types. The agent knows his ability θ. The principal only knows that $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$ and that θ is distributed with density $f(\theta)$ and cdf $F(\theta)$. The density is differentiable and the distribution has full support. Although the analysis is of course more intricate, the results remain as clear-cut as they are in the two type model. To analyze the optimal design of contracts I can invoke the revelation principle. The principle states that it is without loss of generality to restrict attention to direct, incentive compatible, revelation games. The principal commits to a menu of contracts $\left\{\alpha\left(\hat{\theta}\right), \beta\left(\hat{\theta}\right)\right\}$ for all $\hat{\theta}$, the agent announces an ability level, $\hat{\theta} \in \left[\underline{\theta}, \overline{\theta}\right]$, and is given incentives to do this truthfully.

Recall that the certainty equivalent levels of wealth are given

$$w_B\left(\beta\left(\hat{\theta}\right), \alpha\left(\hat{\theta}\right), \theta, \sigma^2\right) = -\beta\left(\hat{\theta}\right) + \left(1-\alpha\left(\hat{\theta}\right)\right)\theta - \frac{b\left(1-\alpha\left(\hat{\theta}\right)\right)^2\sigma^2}{2}$$

and

$$w_A\left(\beta\left(\hat{\theta}\right), \alpha\left(\hat{\theta}\right), \theta, \sigma^2\right) = \beta\left(\hat{\theta}\right) + \alpha\left(\hat{\theta}\right)\theta - \frac{a\alpha\left(\hat{\theta}\right)^2\sigma^2}{2}$$

The principal's problem is the following:

$$\max_{\alpha(\cdot),\beta(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} V\left(w_B\left(\beta(\theta), \alpha(\theta), \theta, \sigma^2\right)\right) f(\theta)\, d\theta \tag{6}$$

s.t. for all $\theta, \hat{\theta}$:

$$w_A\left(\beta(\theta), \alpha(\theta), \theta, \sigma^2\right) \geq w_A\left(\beta\left(\hat{\theta}\right), \alpha\left(\hat{\theta}\right), \theta, \sigma^2\right) \tag{7}$$

and for all θ :

$$w_A\left(\beta(\theta), \alpha(\theta), \theta, \sigma^2\right) \geq \omega \tag{8}$$

Constraint (7) guarantees that the agent reports his type truthfully, and constraint (8) guarantees that the agent wishes to participate.

To solve problem (6) subject to (7) and (8), I derive a tractable characterization for the menus of contracts that satisfy the two constraints. Let such contracts be denoted implementable.

Lemma 1 *A contract is implementable iff*

$$\beta(\theta) = \omega + \frac{a\sigma^2}{2}\left(\alpha(\theta)\right)^2 - \theta\alpha(\theta) + \int_{\underline{\theta}}^{\theta} \alpha(\tau)\, d\tau, \tag{9}$$

and $\alpha(\theta)$ *is non-negative and non-decreasing in* θ *for all* θ*.*

Incentive compatibility requires the usual monotonicity condition on $\alpha(\theta)$. Since monotonic functions are differentiable almost everywhere, $\alpha(\theta)$ is differentiable almost everywhere, and by (9) $\beta(\theta)$ is differentiable almost everywhere as well. Hence, for almost all θ, the first-order condition

$$\beta_{\hat{\theta}}\left(\hat{\theta}\right) + \alpha_{\hat{\theta}}\left(\hat{\theta}\right)\left(\theta - a\alpha\left(\hat{\theta}\right)\sigma^2\right)\Big|_{\hat{\theta}=\theta} = 0$$

must hold. Since the term in brackets is positive under the maintained assumptions, it is obvious that $\beta_{\hat{\theta}}\left(\hat{\theta}\right)$ and $\alpha_{\hat{\theta}}\left(\hat{\theta}\right)$ must have opposing signs. Since the truthful report $\hat{\theta} = \theta$ must also satisfy the second-order condition, we can conclude that $\alpha_{\hat{\theta}}\left(\hat{\theta}\right) \geq 0$ and $\beta_{\hat{\theta}}\left(\hat{\theta}\right) \leq 0$ are necessary for truthtelling; more able agents receive larger shares of profits, but their base pay levels are lower. Invoking the envelope theorem and the first-order condition for truthtelling, I conclude that the change of the agent's utility with respect to an increase in θ is equal to $\alpha(\theta)$. Imposing the individual rationality constraint at the low bound, integrating between $\underline{\theta}$ and θ, and equating the resulting expression to the manager's certainty equivalent wealth, I obtain condition (9).By the standard reasoning the monotonicity condition, that $\alpha(\theta)$ is non-decreasing in θ, makes these local conditions also sufficient for truthtelling to be optimal in the global sense. Since reporting is truthful, I will identify messages and types henceforth, i.e., $\hat{\theta} \equiv \theta$. Economically, the profit share can be used as a sorting device because the more able agents value a marginal increase in α by relatively more than the less able agents.

Consider now the condition that $\alpha(\theta)$ should be non-negative. To understand this condition consider the agent's indirect certainty equivalent wealth, $\mathbf{w}^A(\theta) \equiv \max_{\hat{\theta}} w^A\left(\beta\left(\hat{\theta}\right), \alpha\left(\hat{\theta}\right), \theta\right)$. By the envelope theorem $\mathbf{w}^A_\theta(\theta) = \alpha(\theta)$. Thus, more able agents receive a higher certainty equivalent wealth only if the share of profits is non-negative for all types. Suppose, contrary to the claim in the lemma, that some types receive negative profit shares. By the monotonicity condition on the shares, it would have to be the incompetent agents in some set $\left[\underline{\theta}, \tilde{\theta}\right)$ who receive strictly negative profit shares. But then $\mathbf{w}_A(\theta)$ would reach its minimum at $\theta = \tilde{\theta}$ and the principal would have to leave positive rents to all $\theta \in \left[\underline{\theta}, \tilde{\theta}\right)$. In addition, his income from profit sharing with these types would be more variable than profits themselves. But then risk sharing could be improved if the principal sets $\alpha(\theta) = 0$ for all $\theta \in \left[\underline{\theta}, \tilde{\theta}\right]$. This change would allow the principal in addition to extract all the rents from the low ability types. Since these changes increase the principal's expected income and decrease the variance of his profit share, the candidate menu of contracts that had $\alpha(\theta) < 0$ for some types must have been suboptimal.

Finally, observe that the individual rationality constraint of some agent must be binding. Otherwise the principal can decrease all base salaries by the same amount until the certainty equivalent level of wealth is exactly equal to zero for some θ. Since the principal has CARA utility, such a change of the contract increases his expected utility. By the monotonicity conditions, it is in fact type $\underline{\theta}$ who will be indifferent between participating and not.

4.1 Optimal Risk Sharing

To avoid technical complications related to the monotonicity condition $\alpha_\theta(\theta) \geq 0$, I impose the following assumption on the primitives of the model:

Assumption 2: the density and the parameters satisfy for all θ

$$\frac{f_\theta(\theta)}{f(\theta)} \geq b - \frac{1}{\sigma^2 b} \tag{10}$$

Any density that decreases less fast than the exponential with coefficient $b - \frac{1}{\sigma^2 b}$ satisfies Assumption 2. As usual, these conditions are sufficient to rule out the standard problems of bunching, that would complicate the analysis without adding much in terms of insights.

Using condition (9), I can write the principal's certainty equivalent wealth from offering an implementable contract as

$$-\omega + \theta - \frac{a\alpha(\theta)^2 + b(1-\alpha(\theta))^2 \sigma^2}{2} - \int_{\underline{\theta}}^{\theta} \alpha(\tau)\, d\tau.$$

Notice that $\frac{\partial}{\partial\theta}\left(\int_{\underline{\theta}}^{\theta} \alpha(\tau)\, d\tau\right) = \alpha(\theta)$. Therefore, I can formulate the principal's problem as a problem of optimal control with control variable $\alpha(\theta)$ and state variable $r(\theta) \equiv \int_{\underline{\theta}}^{\theta} \alpha(\tau)\, d\tau$. It takes the following form:

$$\max \int_{\underline{\theta}}^{\overline{\theta}} -\exp\left[-b\left(-w + \theta - \frac{a\alpha^2 + b(1-\alpha)^2}{2}\sigma^2 - r\right)\right] f(\theta)\, d\theta \tag{11}$$

$$s.t. \dot{r} = \alpha;\ r(\underline{\theta}) = 0;\ \alpha(\theta) \geq 0$$

Condition (10) guarantees that the solution to problem (11) automatically satisfies the monotonicity condition $\alpha_\theta(\theta) \geq 0$.

Let

$$J(\theta, \alpha, r) \equiv -\exp\left[-b\left(-w + \theta - \frac{a\alpha^2 + b(1-\alpha)^2}{2}\sigma^2 - r\right)\right] f(\theta)$$

The Hamiltonian for problem (11) is

$$\mathcal{L} = J(\theta, \alpha, r) + (\gamma + \delta)\,\alpha$$

In this expression, γ is the costate variable and δ the Kuhn-Tucker multiplier on the restriction that $\alpha(\theta)$ must be nonnegative. Let $\{\alpha^*(\theta)\}\,\forall\theta$ denote a solution of problem (11). To ease the presentation, it is convenient to define $\{\underline{\alpha}(\theta)\}\,\forall\theta$ as a solution to problem (11) without the constraint $\alpha(\theta) \geq 0$. The solution has the following structure:

Proposition 3 *Suppose the distribution of types satisfies condition* (10). *Then, a contract is optimal if and only if*

$$\alpha^*(\theta) = \max\{0, \underline{\alpha}(\theta)\} \tag{12}$$

where $\underline{\alpha}(\theta)$ *solves the integral equation*

$$\underline{\alpha}(\theta) = \frac{b}{(a+b)} - \frac{\int_{\theta}^{\overline{\theta}} J(\tau, \underline{\alpha}, r)\, d\tau}{(a+b)\,\sigma^2 J(\theta, \underline{\alpha}, r)} \tag{13}$$

Equivalently, $\underline{\alpha}(\theta)$, *is the solution to the first-order differential equation*

$$1 = \sigma^2\left((a+b)\,\underline{\alpha} - b\right)\left[\frac{f_\theta(\theta)}{f(\theta)} - b + b\underline{\alpha}\right] + b\left[\sigma^2\left((a+b)\,\underline{\alpha} - b\right)\right]^2 \dot{\underline{\alpha}} + \sigma^2(a+b)\,\dot{\underline{\alpha}} \tag{14}$$

The fixed payment associated to the optimal profit share satisfies

$$\beta^*(\theta) = \frac{a\sigma^2}{2}\left(\alpha^*(\theta)\right)^2 - \theta\alpha^*(\theta) + \omega + \int_{\underline{\theta}}^{\theta} \alpha^*(\tau)\, d\tau \tag{15}$$

Recall that risk sharing is efficient if the agent bears a share $\frac{b}{a+b}$ of aggregate risk. From condition (13) one can observe that risk sharing is efficient only with the most able type. Less competent agents bear inefficiently little risk. The reason is the familiar one, that incentive compatibility constraints link the rents of the more able agents to the contracts that the less able agents receive. The smaller the downward distortion of the less able agents' profit shares, the more attractive it is for the more able agents to understate their type. Consequently, the larger is the rent they must obtain so that they are indifferent between understating their type and revealing it truthfully. Therefore, to extract the rents of the able agents, the principal distorts the profit shares of the incompetent agents downwards.

An equivalent way to write the solution is by way of the differential equation (14). The regularity condition (10) ensures that any solution to equation (14) is strictly monotonic in θ. The contract may involve corner solutions for the relatively unable types at the low end of the support. Whether the solutions are interior or not depends primarily on the degree of the principal's risk aversion, b. However, before I deal with the rather intricate question of interior solutions, I turn to the central question this model is designed to address, the comparative statics with respect to σ^2.

4.2 Comparative Statics

4.2.1 A positive correlation between risks and incentives

I now turn to the comparative statics analysis of the menu of optimal contracts with respect to changes in the underlying variance, σ^2. I can write the differential equation as

$$\dot{\underline{\alpha}} = \frac{\frac{1}{\sigma^2} - ((a+b)\underline{\alpha} - b)\left[\frac{f_\theta(\theta)}{f(\theta)} - b + b\underline{\alpha}\right]}{\sigma^2 b\left[(a+b)\underline{\alpha} - b\right]^2 + (a+b)}$$

and define the expression on the right-hand side of this equation as $M\left(\alpha\left(\cdot,\sigma^2\right),\sigma^2\right)$. Differentiating with respect to σ^2 on both sides, I can write

$$\alpha_{\theta\sigma^2}\left(\theta,\sigma^2\right) = M_\alpha\left(\alpha\left(\theta,\sigma^2\right),\sigma^2\right)\alpha_{\sigma^2}\left(\theta,\sigma^2\right) + M_{\sigma^2}\left(\alpha\left(\theta,\sigma^2\right),\sigma^2\right)$$

Defining $\kappa(\theta) \equiv \frac{\partial\alpha\left(\cdot,\sigma^2\right)}{\partial\sigma^2}$, I can write this condition as a new differential equation of the form

$$\dot{\kappa}(\theta) = M_\alpha\left(\theta,\sigma^2\right)\kappa(\theta) + M_{\sigma^2}\left(\theta,\sigma^2\right),$$

a nonhomogeneous linear differential equation with boundary condition $\kappa\left(\overline{\theta}\right) = 0$. The boundary condition is deduced from the fact that $\alpha\left(\overline{\theta},\sigma^2\right) = \frac{b}{a+b}$, independent of σ^2. By straightforward calculus I observe that

$$M_{\sigma^2}\left(\theta,\sigma^2\right) = -\frac{\frac{1}{\sigma^4} + b\left[\left((a+b)\underline{\alpha}\left(\theta,\sigma^2\right) - b\right)\right]^2 M\left(\theta,\sigma^2\right)}{\sigma^2 b\left[\left((a+b)\underline{\alpha}\left(\theta,\sigma^2\right) - b\right)\right]^2 + (a+b)} < 0.$$

The inequality follows because the regularity condition (10) implies that $M\left(\theta,\sigma^2\right) = \dot{\underline{\alpha}}\left(\theta,\sigma^2\right) \geq 0$. Moreover, $M_\alpha\left(\theta,\sigma^2\right)$ is bounded for all θ. Therefore $\dot{\kappa}\left(\overline{\theta}\right) = M_{\sigma^2}\left(\overline{\theta},\sigma^2\right) < 0$, where the equality follows from the fact that $\kappa\left(\overline{\theta}\right) = 0$. More generally, for any θ' such that $\kappa\left(\theta'\right) = 0$, I have $\dot{\kappa}\left(\theta'\right) < 0$. Hence, there is at most one such θ'. As I have already observed, I have $\kappa\left(\overline{\theta}\right) = 0$; hence, it follows that $\kappa(\theta) > 0$ for $\theta < \overline{\theta}$. I summarize the result in the following proposition:

Proposition 4 *For any σ^2, the principal shares risks efficiently with the most able type, that is $\alpha^*_{\sigma^2}\left(\overline{\theta},\sigma^2\right) = 0$. For any $\theta < \overline{\theta}$, the agent's share of profits is non-decreasing in σ^2 and strictly increasing in σ^2 if $\alpha^*\left(\theta,\sigma^2\right) > 0$.*

The economics of this result are simple and intuitive. The *principal* bears too much risk except when he contracts with the most able type. The efficiency cost of the suboptimal risk allocation is the larger the larger is the variance of profits. Therefore, the principal reduces the deviation from the efficient risk bearing arrangement as the variance of profits is increased.

A different way to look at this result is to interpret it in terms of the principal's taste for mean and variance of profits. For a given distribution of competence, risk sharing is relatively more important if the risk in the distribution of returns is higher. Therefore, the principal decides to use the agent more for risk sharing at the expense of offering him higher rents. It follows that according to my model there is a strictly *positive correlation* between profit shares and business risk in a cross section of data whenever $\alpha^* (\theta) > 0$ for some types. This correlation is the stronger the larger the subset of types which is offered a strictly positive share of profits. I now show how the portion of types with strictly positive profit shares depends on the principal's degree of risk aversion.

4.2.2 Interior solutions and the principal's degree of risk aversion

The solution schedule $\underline{\alpha} (\theta)$ depends on the principal's degree of risk aversion in so many ways that the direct approach I use for the influence of σ^2 can only be used at a certain cost; in particular I would have to make stronger assumptions on the density $f (\theta)$ to obtain a clear-cut direct result. To avoid such assumptions, I follow an alternative route. As part of the proof of the next proposition below, I show, exploiting the monotonicity properties of the integrand on the right-hand side of (13), that the solution schedule satisfies

$$\frac{b}{a+b} - \frac{1-F(\theta)}{(a+b)\,\sigma^2 f(\theta)} \leq \underline{\alpha}(\theta) \leq \frac{b}{a+b} - \exp\left[-b\left(\overline{\theta} - \theta + \frac{\sigma^2}{2}\frac{b^2}{a+b}\right)\right]\frac{(1-F(\theta))}{(a+b)\,\sigma^2 f(\theta)}. \quad (16)$$

Consider first the lower bound on $\underline{\alpha} (\theta)$. For any $b > 0$, there always exists a $\theta' < \overline{\theta}$, such that the lower bound is positive for all $\theta \in \left(\theta', \overline{\theta}\right]$. So, there is always risk sharing with some types at the high end of the support. If b is sufficiently large, the lower bound is strictly positive for all θ; hence, if the principal is sufficiently risk averse, there is risk sharing with all agents, including the least able ones. Consider now the upper bound. For any $\theta < \overline{\theta}$, there is some b' such that the upper bound is strictly negative for all $b \in [0, b')$. Moreover, for $b = 0$, the upper bound is non-positive for all θ. Hence, in that case there cannot be any risk sharing with any type. I summarize these findings in the following proposition:

Proposition 5 *For $b = 0$, the optimal contract involves $\alpha^* (\theta) = 0$ for all θ. For any $b > 0$, the optimal contract involves $\alpha^* (\theta) > 0$ for some types close to $\overline{\theta}$.For $b > \frac{1-F(\theta)}{\sigma^2 f(\theta)}$ for all θ, the optimal contract involves $\alpha^* (\theta) > 0$ for all θ.*

The parameter b measures the principal's relative taste for the safe payment $\beta (\theta)$ relative to the variable part of his income. In the limit as b goes to zero, the principal's desire to share risks disappears. Therefore, he offers full insurance to the agent at a price which equals the value of the agent's outside option. As a result, the principal can implement the first-best outcome in this special case. For any positive b, the first-best is not implementable and the principal chooses to share risks, at least with some types at the high end of the support. The higher is $\alpha(\theta)$, the more rents the principal gives up. Therefore, b has to be rather large to justify risk sharing with all types, even those at the lower end of the support.[13] As the condition $b > \frac{1-F(\theta)}{\sigma^2 f(\theta)}$ requires that b is sufficiently large, while condition (10) places an upper bound on b, it is worth demonstrating that both conditions can be met simultaneously. For a uniform type distribution, this is the case for $b \in \left(\frac{\overline{\theta}-\underline{\theta}}{\sigma^2}, \frac{1}{\sigma}\right)$, which is a non-empty interval for $\sigma < \overline{\theta} - \underline{\theta}$.

4.3 Extensions

One can extend the model in several dimensions. Most notably, there can be moral hazard in addition to adverse selection. Suppose profits follow a normal distribution $N\left(e+\theta, \sigma^2\right)$, where σ^2 is known and e is the agent's effort. For a strictly positive b one obtains that $\alpha\left(\overline{\theta}\right)$ is an efficient solution to the (moral hazard) incentive provision- risk sharing problem, so in particular, $\alpha\left(\overline{\theta}\right)$ is larger than the agent's profit share in the first-best, $\frac{b}{a+b}$. On the other hand, depending on how important incentives for effort are (that is, depending on the cost-of-effort function), the downward distortion at the low end of the support is so pronounced that $\alpha\left(\underline{\theta}\right)$ is below $\frac{b}{a+b}$. An increase in σ^2 increases the cost of these departures from first-best. Hence, they are reduced at the optimum. Overall, this model gives rise to mixed results. Depending on the distribution of types, the average profit-share (more precisely, the expected profit share, where the expectation is taken with respect to θ) may increase or decrease with an increase in σ^2. Clearly, the positive correlation requires a strictly positive b; for $b = 0$, the agent's share of profits is above first-best for all types; hence, an increase in σ^2 reduces the share for all types. Finally, one can allow for different sources of asymmetric information, e.g. the agent's degree of risk aversion; or for different technologies that allow for interactions between e and θ. The results are robust to these extensions.

The crucial property that drives the positive correlation in the model above is the principal's risk aversion. If that is dropped, then the model clearly predicts a negative correlation between risk and incentives. Thus, to explain a positive correlation between risk and incentives in a model with a risk neutral principal, a different approach is required. I develop such an approach in the next section.

---

[13] Since we are dealing with differential equations, the results are much harder to get than the simple intuition suggests. If one is willing to assume a uniform distribution of types, then one can solve the model in closed form. Details are available from the author upon request.

5 A model of managerial style

Risk aversion of the principal is unrealistic in the context of large firms, where the principal represents a group of (well diversified) shareholders. Thus, to capture such a situation, I assume from now on that the principal is risk neutral. The agent is called a manager, henceforth, and to simplify the analysis, I assume the manager is risk neutral as well. I let $\tilde{\pi} \sim N\left(e+\theta, \sigma^2\right)$, where e is the manager's effort, the parameters θ and σ capture the manager's style. I assume that θ is a given parameter which is known to the agent but not to the principal. σ is a choice made by the agent; σ can take two values, σ_l and σ_h with $\sigma_h > \sigma_l$. The cost of choosing a high variance is normalized to zero. The cost of choosing a low variance depends on the manager's type. If a manager with mean parameter θ is diligent, then his cost of choosing the low variance is $q\left(\sigma_l\right) = \underline{\eta}$; if he is negligent, then his cost is $q\left(\sigma_l\right) = \overline{\eta}$. The manager's choice is observable; however, the manager's cost of choosing the variance is not observable. Without loss of generality, I normalize the value of the agent's outside option and the parameter $\underline{\eta}$ to zero. Henceforth, I denote the manager's type by the tupe (θ, η).

Suppose there are costs of financial distress D that accrue when profits fall short of some level $\pi^{\min}$. The expected costs of distress are equal to product of the probability of π being smaller than $\pi^{\min}$ and D, and thus a function of both θ and σ. Let these expected costs of financial distress be denoted $\tilde{D}(\theta, \sigma)$ and suppose they enter the principal's payoff with a nonnegative weight ε measuring their importance.[14]

I let $f\left(\theta \mid \eta\right)$ and $F\left(\theta \mid \eta\right)$, respectively, denote the density and the cdf, respectively, of the conditional distribution of θ given η and let p denote the probability that $\eta = \underline{\eta}$. The marginal distribution of θ has density $f(\theta) = pf\left(\theta \mid \underline{\eta}\right) + (1-p) f\left(\theta \mid \overline{\eta}\right)$ and cdf $F(\theta) = pF\left(\theta \mid \underline{\eta}\right) + (1-p) F\left(\theta \mid \overline{\eta}\right)$. To simplify notation, I let $\underline{\beta}(\theta) \equiv \beta\left(\theta, \underline{\eta}\right)$, $\underline{\alpha}(\theta) \equiv \alpha\left(\theta, \underline{\eta}\right)$, $\underline{e}(\theta) \equiv e\left(\theta, \underline{\eta}\right)$, and $\underline{\sigma}(\theta) = \sigma\left(\theta, \underline{\eta}\right)$, and use the analogous notation for the negligent's type's contracts.

The manager's expected utility for given contracts takes the form

$$\beta\left(\hat{\theta}\right) + \alpha\left(\hat{\theta}\right)(e+\theta) - c(e).$$

Since this expression is concave in e, an optimal effort choice (from the manager's perspective) satisfies the first-order condition $\alpha\left(\hat{\theta}\right) = c_e(e)$. Inverting this expression, I obtain

$$e\left(\hat{\theta}\right) = c_e^{-1}\left(\alpha\left(\hat{\theta}\right)\right). \tag{17}$$

---

[14] All the results I present below remain valid if $\varepsilon = 0$. In that case the choice of risk is relevant only as a signal allowing for more precise inference about the unknown parameter θ.

This condition can be used to eliminate the moral hazard problem altogether and solve the principal's problem as a problem of pure adverse selection.

I argue below that for ε small, more precisely for $\varepsilon < \frac{\overline{\eta}}{\tilde{D}(\theta,\overline{\sigma})-\tilde{D}(\theta,\underline{\sigma})}$ for all θ, it is optimal to implement low variance for diligent types and a high variance for negligent types, i.e. $\underline{\sigma}(\theta) = \sigma_l$ and $\overline{\sigma}(\theta) = \sigma_h$. It is easiest to explain this result after spelling out the problem. Simply assuming that $\underline{\sigma}(\theta) = \sigma_l$ and $\overline{\sigma}(\theta) = \sigma_h$ for now, the principal's problem is

$$\max p \int_{\underline{\theta}}^{\overline{\theta}} \left(-\underline{\beta}(\theta) + (1-\underline{\alpha}(\theta))\left(c_e^{-1}(\underline{\alpha}(\theta)) + \theta\right) - \varepsilon \tilde{D}(\theta,\underline{\sigma})\right) f\left(\theta \middle| \underline{\eta}\right) d\theta \tag{18}$$

$$+ (1-p) \int_{\underline{\theta}}^{\overline{\theta}} \left(-\overline{\beta}(\theta) + (1-\overline{\alpha}(\theta))\left(c_e^{-1}(\overline{\alpha}(\theta)) + \theta\right) - \varepsilon \tilde{D}(\theta,\overline{\sigma})\right) f\left(\theta \middle| \overline{\eta}\right) d\theta$$

s.t. for all θ and all $\hat{\theta}$

$$\underline{\beta}(\theta) + \underline{\alpha}(\theta)\left(c_e^{-1}(\underline{\alpha}(\theta)) + \theta\right) - c\left(c_e^{-1}(\underline{\alpha}(\theta))\right) \geq \underline{\beta}\left(\hat{\theta}\right) + \underline{\alpha}\left(\hat{\theta}\right)\left(c_e^{-1}\left(\underline{\alpha}\left(\hat{\theta}\right)\right) + \theta\right) - c\left(c_e^{-1}\left(\underline{\alpha}\left(\hat{\theta}\right)\right)\right) \tag{19}$$

$$\underline{\beta}(\theta) + \underline{\alpha}(\theta)\left(c_e^{-1}(\underline{\alpha}(\theta)) + \theta\right) - c\left(c_e^{-1}(\underline{\alpha}(\theta))\right) \geq \overline{\beta}\left(\hat{\theta}\right) + \overline{\alpha}\left(\hat{\theta}\right)\left(c_e^{-1}\left(\overline{\alpha}\left(\hat{\theta}\right)\right) + \theta\right) - c\left(c_e^{-1}\left(\overline{\alpha}\left(\hat{\theta}\right)\right)\right) \tag{20}$$

$$\overline{\beta}(\theta) + \overline{\alpha}(\theta)\left(c_e^{-1}(\overline{\alpha}(\theta)) + \theta\right) - c\left(c_e^{-1}(\overline{\alpha}(\theta))\right) \geq \overline{\beta}\left(\hat{\theta}\right) + \overline{\alpha}\left(\hat{\theta}\right)\left(c_e^{-1}\left(\overline{\alpha}\left(\hat{\theta}\right)\right) + \theta\right) - c\left(c_e^{-1}\left(\overline{\alpha}\left(\hat{\theta}\right)\right)\right) \tag{21}$$

$$\overline{\beta}(\theta) + \overline{\alpha}(\theta)\left(c_e^{-1}(\overline{\alpha}(\theta)) + \theta\right) - c\left(c_e^{-1}(\overline{\alpha}(\theta))\right) \geq \underline{\beta}\left(\hat{\theta}\right) + \underline{\alpha}\left(\hat{\theta}\right)\left(c_e^{-1}\left(\underline{\alpha}\left(\hat{\theta}\right)\right) + \theta\right) - c\left(c_e^{-1}\left(\underline{\alpha}\left(\hat{\theta}\right)\right)\right) - \overline{\eta} \tag{22}$$

$$\underline{\beta}(\theta) + \underline{\alpha}(\theta)\left(c_e^{-1}(\underline{\alpha}(\theta)) + \theta\right) - c\left(c_e^{-1}(\underline{\alpha}(\theta))\right) \geq 0 \tag{23}$$

$$\overline{\beta}(\theta) + \overline{\alpha}(\theta)\left(c_e^{-1}(\overline{\alpha}(\theta)) + \theta\right) - c\left(c_e^{-1}(\overline{\alpha}(\theta))\right) \geq 0 \tag{24}$$

In this problem,(19) and (21) are standard incentive compatibility constraints stating that no type should have an incentive to deviate by announcing any different value of θ. (20) is an additional incentive constraint reflecting that the diligent type can choose a higher variance at the same cost and announce a different value of θ at the same time. (22) is an additional constraint ruling out any double deviation by the negligent manager. Finally, constraints (23) and (24) require that all manager types should receive an expected payoff weakly exceeding the value of their outside option which is normalized to zero.

Notice that the values on the right-hand side of (21) and the right-hand side of (20) are identical. Moreover, (21) ensures that this value is in fact maximized at $\hat{\theta} = \theta$. Hence, the incentive constraint

(20) is met whenever the indirect utility of any type $\left(\theta, \underline{\eta}\right)$ is weakly larger than the indirect utility of type $(\theta, \overline{\eta})$. Similarly, the right-hand sides of (19) and (20) differ only by an additive constant. Hence, the incentive constraint (22) is met whenever the difference between the indirect utility of any type $\left(\theta, \underline{\eta}\right)$ and the indirect utility of type $(\theta, \overline{\eta})$ is at most $\overline{\eta}$. This simple observation is what makes this problem tractable.[15]

It is now obvious why implementing $\underline{\sigma}(\theta) = \sigma_l$ for all diligent types and $\overline{\sigma}(\theta) = \sigma_h$ for all negligent types is optimal for ε small enough. If instead the principal implemented $\underline{\sigma}(\theta) = \sigma_h$ for some diligent types, then (20) would become tighter because it becomes cheaper for some negligent type to imitate the diligent type. Since no additional benefit arises from such a change of policy, it cannot increase the principal's payoff. If the principal implemented $\overline{\sigma}(\theta) = \sigma_l$ for some negligent types then he would have to increase payments to all such types (diligent and negligent) by $\overline{\eta}$. For ε small, more specifically for $\varepsilon\left(\tilde{D}(\theta, \overline{\sigma}) - \tilde{D}(\theta, \underline{\sigma})\right) < \overline{\eta}$ for all θ, this is again suboptimal.

5.1 Optimal contracts

I now solve the principal's problem. To do so, I begin by bringing the problem into an equivalent, but more tractable form:

Lemma 2 *The principal's problem is equivalent to*

$$\max_{\underline{\alpha}(\theta), \overline{\alpha}(\theta), \rho} p \int_{\underline{\theta}}^{\overline{\theta}} \left(c_e^{-1}\left(\underline{\alpha}(\theta)\right) + \theta - c\left(c_e^{-1}\left(\underline{\alpha}(\theta)\right)\right) - \frac{1 - F\left(\theta \mid \underline{\eta}\right)}{f\left(\theta \mid \underline{\eta}\right)} \underline{\alpha}(\theta) - \varepsilon \tilde{D}(\theta, \underline{\sigma}) \right) f\left(\theta \mid \underline{\eta}\right) d\theta - p\rho \tag{25}$$

$$+ (1-p) \int_{\underline{\theta}}^{\overline{\theta}} \left(\left(c_e^{-1}\left(\overline{\alpha}(\theta)\right) + \theta\right) - c\left(c_e^{-1}\left(\overline{\alpha}(\theta)\right)\right) - \frac{1 - F(\theta \mid \overline{\eta})}{f(\theta \mid \overline{\eta})} \overline{\alpha}(\theta) - \varepsilon \tilde{D}(\theta, \overline{\sigma}) \right) f(\theta \mid \overline{\eta})$$

s.t. for all θ

$$\underline{\alpha}(\theta), \overline{\alpha}(\theta) \text{ non-decreasing in } \theta,$$

$$\rho \geq 0,$$

$$\rho + \int_{\underline{\theta}}^{\theta} \underline{\alpha}(\tau)\, d\tau \geq \int_{\underline{\theta}}^{\theta} \overline{\alpha}(\tau)\, d\tau, \text{ and} \tag{26}$$

$$\rho - \overline{\eta} + \int_{\underline{\theta}}^{\theta} \underline{\alpha}(\tau)\, d\tau \leq \int_{\underline{\theta}}^{\theta} \overline{\alpha}(\tau)\, d\tau. \tag{27}$$

---

[15] For a model that exploits the same tractability in a different context, see Beaudry, Blackorby, and Szalay (2009).

Up to the constraints (26) and (27), the result is obvious to the reader familiar with one dimensional incentive models. In fact, the objective is simply the sum over two maximization problems, each of which reflecting the principal's expected profit from contracting with a diligent and a negligent manager type, respectively. (26) requires that the indirect utility of any type $(\theta, \underline{\eta})$ is weakly larger than the indirect utility of type $(\theta, \overline{\eta})$. Clearly, this requires that type $(\underline{\theta}, \underline{\eta})$ receives a non-negative rent, ρ.

The formal proof uses the standard methods that I have already employed in the analysis of the preceding model. So, I just sketch the argument here. Using the envelope theorem, and the manager's incentive compatibility condition for truthtelling in θ, I can deduce that the diligent manager's indirect utility increases with θ at the rate $\underline{\alpha}(\theta)$, while the negligent manager's indirect utility increases at rate $\overline{\alpha}(\theta)$. Clearly, a negligent manager with the lowest mean parameter, $\underline{\theta}$, does not receive any positive rent at the optimum. However, we cannot rule out, on a priori grounds, that the diligent manager with the lowest mean receives a positive rent. Hence, I leave this rent as a non-negative parameter ρ for the time being. Letting $\underline{u}(\theta)$ and $\overline{u}(\theta)$ denote the indirect utility schedules of the diligent and the negligent type, I can integrate up to get $\underline{u}(\theta) = \rho + \int_{\underline{\theta}}^{\theta} \underline{\alpha}(\tau)\, d\tau$ and $\overline{u}(\theta) = \int_{\underline{\theta}}^{\theta} \overline{\alpha}(\tau)\, d\tau$. Substituting for the manager's indirect utility into the objective function allows me to eliminate the base pay schedules $\underline{\beta}(\theta)$ and $\overline{\beta}(\theta)$ from the problem. Integrating by parts, I obtain the new objective function (25). The monotonicity conditions are necessary for incentive compatibility; together with the local incentive constraints, they are also sufficient for incentive compatibility. Finally, (26) requires that the indirect utility of the diligent manager be weakly higher than the indirect utility of a negligent manager. This constraint completely takes care of any deviation in both dimensions at the same time, because I have argued above that due to (21), the best possible report about θ, conditional on misrepresenting diligence, is to state the truth about θ. By the same argument, constraint (27) takes care of double deviations of managers of the negligent type. Since their utility differs from the utility of manager of the same mean type only by a constant, the optimal report in the θ dimension, conditional on having misreported the parameter η, is the same as the optimal report of the diligent manager with the same mean parameter. Hence, even conditional on misreporting η, the optimal report in the θ dimension remains truthful.

As is customary, I will solve my problem imposing conditions that make the monotonicity conditions on $\underline{\alpha}(\theta)$ and $\overline{\alpha}(\theta)$ nonbinding. In that case, problem (25) subject to (26) is a problem of optimal control with control variables $\underline{\alpha}(\theta)$ and $\overline{\alpha}(\theta)$, and state variables $\int_{\underline{\theta}}^{\theta} \underline{\alpha}(\tau)\, d\tau$ and $\int_{\underline{\theta}}^{\theta} \overline{\alpha}(\tau)\, d\tau$. Moreover, the state variables are subject to inequality constraints.

Solution techniques for control problems with inequality constraints on state variables are known (see e.g. Seierstad and Sydsaeter (1993)) but pretty involved. Unfortunately, in the present context, the problem is interesting - in the sense of offering interesting comparative statics - only when at least one of the constraints is binding for some types.

I now introduce two assumptions. The first, to generate the right comparative statics prediction, the second to eliminate the monotonicity constraints from the problem. Note that assumptions 1 and 2 are no longer needed in this model.

Assumption 3: θ and $-\eta$ are affiliated, that is $\frac{f(\theta|\underline{\eta})}{f(\theta|\overline{\eta})}$ is increasing in θ.

For differentiable densities, Assumption 3 is equivalent to the requirement that

$$\frac{f_\theta\left(\theta|\,\underline{\eta}\right)}{f\left(\theta|\,\underline{\eta}\right)} > \frac{f_\theta\left(\theta|\,\overline{\eta}\right)}{f\left(\theta|\,\overline{\eta}\right)} \text{ for all } \theta,$$

which has important implications on the slope of the candidate solution schedules for problem (25). Moreover, as is well known (see Shaked and Shantikumar (2007)), assumption 3 also implies that

$$\frac{1-F\left(\theta|\,\underline{\eta}\right)}{f\left(\theta|\,\underline{\eta}\right)} > \frac{1-F\left(\theta|\,\overline{\eta}\right)}{f\left(\theta|\,\overline{\eta}\right)} \text{ for all } \theta,$$

which has important implications on the level of candidate solution schedules for problem (25).

To guarantee that the monotonicity constraints are automatically satisfied at the optimum, I impose:

Assumption 4: $\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{f(\theta)} \leq 0$ for all θ.

Monotonicity of the inverse hazard rate is the standard condition imposed in most of the screening literature. Many well known distributions satisfy the condition.

Under these assumptions the solution takes the following form:

Proposition 6 *If θ and $-\eta$ are affiliated and $\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{f(\theta)} \leq 0$ for all θ, then $\rho^* = 0$, and the optimal share schedules satisfy $\underline{\alpha}(\theta) = \overline{\alpha}(\theta) = \alpha^*(\theta)$, where $\alpha^*(\theta)$ solves the condition*

$$\frac{1-\alpha^*(\theta)}{c_{ee}\left(c_e^{-1}\left(\alpha^*(\theta)\right)\right)} - \frac{1-F(\theta)}{f(\theta)} = 0, \tag{28}$$

the optimal base-pay schedules satisfy

$$\underline{\beta}^*(\theta) = \overline{\beta}^*(\theta) = \int_{\underline{\theta}}^{\theta} \alpha^*(\tau)\, d\tau - \alpha^*(\theta)\left(c_e^{-1}\left(\alpha^*(\theta)\right) + \theta\right) + c\left(c_e^{-1}\left(\alpha^*(\theta)\right)\right),$$

and the optimal effort schedules satisfy $\underline{e}^*(\theta) = \overline{e}^*(\theta) = c_e^{-1}\left(\alpha^*(\theta)\right)$.

The complete proof is in the appendix; here is a sketch of the argument. First, I conjecture that only one of the constraints (26) and (27) is binding at the optimum; moreover, I conjecture that (27) is slack at the optimum. Likewise, I neglect the monotonicity constraints. I solve the problem taking these properties for granted, and check from the solution whether the conjectures are justified.

To solve the reduced problem, I proceed as follows. First, I show that the inequality constraint (26) needs to be binding for some θ at the optimum. Suppose not; then the solution schedules would satisfy the first-order conditions

$$\frac{1-\underline{\alpha}(\theta)}{c_{ee}\left(c_e^{-1}\left(\underline{\alpha}(\theta)\right)\right)}-\frac{1-F\left(\theta|\,\underline{\eta}\right)}{f\left(\theta|\,\underline{\eta}\right)}=0 \tag{29}$$

and

$$\frac{1-\overline{\alpha}(\theta)}{c_{ee}\left(c_e^{-1}\left(\overline{\alpha}(\theta)\right)\right)}-\frac{1-F\left(\theta|\,\overline{\eta}\right)}{f\left(\theta|\,\overline{\eta}\right)}=0. \tag{30}$$

However, due to assumption 3, these schedules satisfy $\underline{\alpha}(\theta)<\overline{\alpha}(\theta)$ for all $\theta<\overline{\theta}$. Hence, the principal needs to set a large value of ρ to make this scheme incentive compatible. However, the shadow cost of reducing ρ is zero when the inequality constraint is non-binding; on the other hand there is a first-order gain in terms of reduced rents when ρ is reduced.

The second step is to show that the constraint must be binding at the high end of the support. The proof is by contradiction. If the constraint is non-binding at the high end, then the transversality conditions (which have to hold if the constraint is slack at the high end) imply again that the schedules satisfy (29) and (30) at the high end. Letting $R(\theta)\equiv\underline{u}(\theta)-\overline{u}(\theta)$, I have $R_\theta(\theta)=\underline{\alpha}(\theta)-\overline{\alpha}(\theta)$. So, since $\underline{\alpha}(\theta)<\overline{\alpha}(\theta)$, $R_\theta(\theta)<0$ at the high end of the support. However, we already know that there must be some value of θ where $R(\theta)=0$. Taken together, this implies that $R(\theta)<0$ for some θ at the high end of the support, a contradiction.

The next step is to show that the region over which $R(\theta)=0$ at the optimum must be convex. The reason is that assumption 3 implies that the solution schedules over an interval $[\theta_1,\theta_2]$ where $R(\theta_1)=R(\theta_2)=0$ and $R(\theta)>0$ for $\theta\in(\theta_1,\theta_2)$ satisfy a single-crossing condition: if the schedules $\underline{\alpha}(\theta)$ and $\overline{\alpha}(\theta)$ cross at all over that interval, $\underline{\alpha}(\theta)$ must cross $\overline{\alpha}(\theta)$ from below. But this implies that either $R_\theta(\theta_1)<0$ - so that the constraint is violated for θ larger than θ_1 - or that $R_\theta(\theta)>0$ for all $\theta>\theta_1$ - so that $R(\theta_2)>0$. Hence, $R(\theta)=0$ holds over a convex region.

The final elements in the picture are to prove that at the optimum $\rho=0$, and that this implies that $R(\theta)=0$ over the entire support. To show this I demonstrate that the optimal schedules for the best positive value of ρ satisfy $\underline{\alpha}(\theta)>\overline{\alpha}(\theta)$ over the region where the inequality constraint is non-binding. However, this implies that the diligent type's rent increases in θ over that region, implying that the constraint would not be binding at all over the whole support. As I have shown above, this cannot be optimal.

5.2 Comparative Statics

I can now turn to comparative statics. In particular, I will consider the effect of a reduction in p on the optimal share schedule. This is interesting because the average risk of firms in the model is $\tilde{\sigma}(p) \equiv p\underline{\sigma} + (1-p)\overline{\sigma}$. Hence, a reduction in p increases the average riskiness of firms. Since the characterization of optimal contracts is not affected by a change in p, the effect of a change in p is fully captured by the changes in the solution schedule (28) . I have the following result:

Proposition 7 *Under assumption 3, a reduction in p increases the share of profits for all $\theta < \overline{\theta}$; a manager of type $\overline{\theta}$ receives share $\alpha\left(\overline{\theta}\right) = 1$ for any p.*

Proof. Differentiating (28) totally, I obtain

$$\frac{d\alpha}{dp} = \frac{\frac{\partial}{\partial p}\frac{1-F(\theta)}{f(\theta)}}{\frac{\partial}{\partial \alpha}\frac{1-\alpha^{*}(\theta)}{c_{ee}\left(c_{e}^{-1}(\alpha^{*}(\theta))\right)}}$$

Under the maintained assumption that $c_{eee}(e) \geq 0$, the denominator is strictly negative. So, $\frac{d\alpha}{dp} < 0$ (implying that a reduction in p increases α) if the numerator is non-positive (strictly negative for $\theta < \overline{\theta}$). Using the definition of the marginal distribution, I obtain

$$\frac{\partial}{\partial p}\frac{\left(1 - pF\left(\theta|\,\underline{\eta}\right) - (1-p)\,F\left(\theta|\,\overline{\eta}\right)\right)}{\left(pf\left(\theta|\,\underline{\eta}\right) + (1-p)\,f\left(\theta|\,\overline{\eta}\right)\right)} > 0$$

if and only if

$$\frac{\left(1 - F\left(\theta|\,\underline{\eta}\right)\right)}{f\left(\theta|\,\underline{\eta}\right)} > \frac{\left(1 - F\left(\theta|\,\overline{\eta}\right)\right)}{f\left(\theta|\,\overline{\eta}\right)}.$$

■

6 Conclusions

This paper explores the predictions of the standard linear contracting principal agent model when the agent has some private information with respect to the distribution of profits. I show that a positive correlation between risk and incentives arises quite naturally in the model. One source of this positive correlation is risk aversion and the desire to share risks. The trade-off between risk sharing and rent extraction distorts optimal contracts below the first-best level, while moral hazard distorts the second-best optimal contract to a level above the first-best level. The higher is risk, the closer is the solution to the first-best benchmark. Hence, the risk-sharing/rent-extraction model reverses the comparative statics results of the risk-sharing/moral hazard model.

In the second model, the principal has a preference for low risk because lower risk reduces the probability of financial distress. Moreover, the manager's choice of riskiness is informative about his mean talent. However, although manager types who differ in their cost of reducing variance are induced to choose different levels of risk, at the optimum this information is not used to offer them different profit shares. Separating types with respect to that instrument as well would be too costly relative to its benefit. Under the same assumption that generates the bunching in the first place, the optimal incentive schedules and the average risk in the model covary: higher risk is associated with higher powered incentive schemes.

The overall conclusion from these models is that an empirical finding of a positive correlation between risk and incentives is perfectly consistent with the linear contracting model; however, it allows us to reject the simple version of the contracting model that relies on pure moral hazard.

7 Appendix

7.1 The classical model

Proof of Proposition 1. The principal's problem in explicit form is

$$\max_{\alpha,\beta} -\exp\left(-b\left(-\beta+(1-\alpha)e-b\frac{(1-\alpha)^2}{2}\sigma^2\right)\right) \tag{P}$$

$$s.t.$$

$$\alpha = c_e(e) \tag{IC}$$

$$\beta+\alpha e-a\frac{\alpha^2}{2}\sigma^2-c(e)=\omega \tag{IR}$$

Solving for β from the IR constraint and using the IC constraint to substitute for e I obtain the unrestricted problem:

$$\max_{\alpha} -\exp\left(-b\left(-\omega+c_e^{-1}(\alpha)-a\frac{\alpha^2}{2}\sigma^2-b\frac{(1-\alpha)^2}{2}\sigma^2-c\left(c_e^{-1}(\alpha)\right)\right)\right)$$

Provided that $c_{eee}(e)\geq 0$ for all e this problem is strictly concave in α. The optimal contract, α^*, satisfies the first-order condition

$$\left.\frac{1-\alpha}{c_{ee}\left(c_e^{-1}(\alpha)\right)}-\alpha(a+b)\sigma^2+b\sigma^2\right|_{\alpha=\alpha^*}=0$$

By the implicit function theorem

$$\left(\frac{-c_{ee}(\cdot)-(1-\alpha^*)\frac{c_{eee}(\cdot)}{c_{ee}(\cdot)}}{\left(c_{ee}(\cdot)\right)^2}-(a+b)\,\sigma^2\right)d\alpha^*+\left(-\alpha^*(a+b)+b\right)d\sigma^2=0$$

and hence

$$\frac{d\alpha^*}{d\sigma^2}=\frac{\alpha^*(a+b)-b}{\frac{-c_{ee}(\cdot)-(1-\alpha)\frac{c_{eee}(\cdot)}{c_{ee}(\cdot)}}{\left(c_{ee}(\cdot)\right)^2}-(a+b)\,\sigma^2}$$

Using the first-order condition I note that

$$\alpha^*(a+b)-b=\frac{1-\alpha^*}{c_{ee}\left(c_e^{-1}(\alpha)\right)\sigma^2}>0$$

Hence, I have

$$\frac{d\alpha^*}{d\sigma^2}<0.$$

■

7.2 The two type partnership model

The neglected constraints are slack. I first show that constraint $\left(\overline{IR}\right)$ is slack whenever constraints $\left(\overline{IC}\right)$ and $(\underline{IR})$ hold as equalities. In particular, one has

$$\begin{aligned}
w^A\left(\overline{\alpha},\overline{\beta},\overline{\theta},\sigma^2\right) &= w^A\left(\underline{\alpha},\underline{\beta},\overline{\theta},\sigma^2\right)\\
&= w^A\left(\underline{\alpha},\underline{\beta},\underline{\theta},\sigma^2\right)+\underline{\alpha}\left(\overline{\theta}-\underline{\theta}\right)\\
&= \omega+\underline{\alpha}\left(\overline{\theta}-\underline{\theta}\right),
\end{aligned}$$

where the first equality uses $\left(\overline{IC}\right)$, the second follows from straightforward algebra, and the third one uses $(\underline{IR})$. Since, $\underline{\alpha}\left(\overline{\theta}-\underline{\theta}\right)\geq 0$, constraint $\left(\overline{IR}\right)$ is indeed implied by $\left(\overline{IC}\right)$ and $(\underline{IR})$ holding as equalities.

Consider now constraint $(\underline{IC})$. I can write

$$\begin{aligned}
w^A\left(\underline{\alpha},\underline{\beta},\underline{\theta},\sigma^2\right) &\geq w^A\left(\overline{\alpha},\overline{\beta},\underline{\theta},\sigma^2\right)\\
&= w^A\left(\overline{\alpha},\overline{\beta},\overline{\theta},\sigma^2\right)-\overline{\alpha}\left(\overline{\theta}-\underline{\theta}\right)\\
&= \omega+\left(\underline{\alpha}-\overline{\alpha}\right)\left(\overline{\theta}-\underline{\theta}\right),
\end{aligned}$$

where first equality follows from simple algebra, and the second equality uses the binding constraint $\left(\overline{IC}\right)$. Clearly, constraint $(\underline{IC})$ is implied by $\left(\overline{IC}\right)$ and $(\underline{IR})$ holding as equalities whenever $\underline{\alpha}\leq\overline{\alpha}$.

■

Proof of Proposition 2. Starting from the first-order condition,

$$\lambda \frac{\partial}{\partial \underline{\alpha}} V\left(w^B\left(\overline{\alpha}, \overline{\beta}\left(\overline{\alpha}, \underline{\alpha}\right), \overline{\theta}, \sigma^2\right)\right) + (1-\lambda) \frac{\partial}{\partial \underline{\alpha}} V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right) = 0,$$

I use the implicit function theorem to compute

$$\frac{d\underline{\alpha}}{d\sigma^2} = -\frac{\lambda \frac{\partial^2}{\partial \underline{\alpha} \partial \sigma^2} V\left(w^B\left(\overline{\alpha}, \overline{\beta}\left(\overline{\alpha}, \underline{\alpha}\right), \overline{\theta}, \sigma^2\right)\right) + (1-\lambda) \frac{\partial^2}{\partial \underline{\alpha} \partial \sigma^2} V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right)}{\lambda \frac{\partial^2}{\partial \underline{\alpha}^2} V\left(w^B\left(\overline{\alpha}, \overline{\beta}\left(\overline{\alpha}, \underline{\alpha}\right), \overline{\theta}, \sigma^2\right)\right) + (1-\lambda) \frac{\partial^2}{\partial \underline{\alpha}^2} V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right)}.$$

By the second-order condition of the principal's problem, the denominator is negative. Hence, the sign of $\frac{d\underline{\alpha}}{d\sigma^2}$ is given by the sign of the numerator. To determine the sign of the numerator, I first write the first-order condition in explicit form:

$$\begin{aligned}
& -\lambda \exp\left[-b\left(\overline{\theta} - \omega - \underline{\alpha}\left(\overline{\theta} - \underline{\theta}\right) - a\frac{\overline{\alpha}^2}{2}\sigma^2 - b\frac{\left(1-\overline{\alpha}\right)^2}{2}\sigma^2\right)\right]\left(\overline{\theta} - \underline{\theta}\right) \\
& -(1-\lambda) \exp\left[-b\left(\underline{\theta} - \omega - a\frac{\underline{\alpha}^2}{2}\sigma^2 - b\frac{\left(1-\underline{\alpha}\right)^2}{2}\sigma^2\right)\right]\left(-b\sigma^2\left(1-\underline{\alpha}\right) + a\underline{\alpha}\sigma^2\right) \\
= \; & 0
\end{aligned}$$

or again

$$\lambda V\left(w^B\left(\overline{\alpha}, \overline{\beta}\left(\overline{\alpha}, \underline{\alpha}\right), \overline{\theta}, \sigma^2\right)\right)\left(\overline{\theta} - \underline{\theta}\right) + (1-\lambda) V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right)\left(-b\sigma^2\left(1-\underline{\alpha}\right) + a\underline{\alpha}\sigma^2\right) = 0$$

For future reference, I note this implies

$$\left(-b\left(1-\underline{\alpha}\right) + a\underline{\alpha}\right)\sigma^2 = -\frac{\lambda V\left(w^B\left(\overline{\alpha}, \overline{\beta}\left(\overline{\alpha}, \underline{\alpha}\right), \overline{\theta}, \sigma^2\right)\right)\left(\overline{\theta} - \underline{\theta}\right)}{(1-\lambda) V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right)} < 0.$$

Differentiating the first-order condition with respect to σ^2, I get

$$\begin{aligned}
& \lambda \frac{\partial^2}{\partial \underline{\alpha} \partial \sigma^2} V\left(w^B\left(\overline{\alpha}, \overline{\beta}\left(\overline{\alpha}, \underline{\alpha}\right), \overline{\theta}, \sigma^2\right)\right) + (1-\lambda) \frac{\partial^2}{\partial \underline{\alpha} \partial \sigma^2} V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right) \\
= \; & b\left(a\frac{\overline{\alpha}^2}{2} + b\frac{\left(1-\overline{\alpha}\right)^2}{2}\right)\left(\overline{\theta} - \underline{\theta}\right) \lambda V\left(w^B\left(\overline{\alpha}, \overline{\beta}\left(\overline{\alpha}, \underline{\alpha}\right), \overline{\theta}, \sigma^2\right)\right) \\
& + b\left(a\frac{\underline{\alpha}^2}{2} + b\frac{\left(1-\underline{\alpha}\right)^2}{2}\right)(1-\lambda) V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right)\left(-b\sigma^2\left(1-\underline{\alpha}\right) + a\underline{\alpha}\sigma^2\right) \\
& + (1-\lambda) V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right)\left(-b\left(1-\underline{\alpha}\right) + a\underline{\alpha}\right)
\end{aligned}$$

The expression on the third line is positive. To see this, note from the preliminary observation that $-b + \underline{\alpha}^*(a+b) < 0$; as $V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right) < 0$, this proves the claim.

Consider now the expressions on line one and two. From the first-order condition one can easily see that the expression on line one is negative and the expression on line two is positive, again because $\left(-b\sigma^2\left(1-\underline{\alpha}\right) + a\underline{\alpha}\sigma^2\right) < 0$. I can write the sum of the two effects as

$$b\left(a\frac{\underline{\alpha}^2}{2} + b\frac{\left(1-\underline{\alpha}\right)^2}{2}\right)\left\{\begin{array}{c}
\frac{\left(a\frac{\overline{\alpha}^2}{2} + b\frac{\left(1-\overline{\alpha}\right)^2}{2}\right)}{\left(a\frac{\underline{\alpha}^2}{2} + b\frac{\left(1-\underline{\alpha}\right)^2}{2}\right)}\left(\overline{\theta} - \underline{\theta}\right) \lambda V\left(w^B\left(\overline{\alpha}, \overline{\beta}\left(\overline{\alpha}, \underline{\alpha}\right), \overline{\theta}, \sigma^2\right)\right) \\
+ (1-\lambda) V\left(w^B\left(\underline{\alpha}, \underline{\beta}\left(\underline{\alpha}\right), \underline{\theta}, \sigma^2\right)\right)\left(-b\sigma^2\left(1-\underline{\alpha}\right) + a\underline{\alpha}\sigma^2\right)
\end{array}\right\}.$$

The sign of the overall expression equals the sign of the expression $\{\cdot\}$, which is - up to the factor $\frac{\left(a\frac{\overline{\alpha}^2}{2}+b\frac{(1-\overline{\alpha})^2}{2}\right)}{\left(a\frac{\underline{\alpha}^2}{2}+b\frac{(1-\underline{\alpha})^2}{2}\right)}$ - just the first-order condition. Since $\overline{\alpha}^* = \frac{b}{a+b} > \underline{\alpha}^*$ and $\frac{b}{a+b} = \arg\min_\alpha \left(a\frac{\alpha^2}{2} + b\frac{(1-\alpha)^2}{2}\right)$, I have

$$\frac{\left(a\frac{\overline{\alpha}^2}{2}+b\frac{(1-\overline{\alpha})^2}{2}\right)}{\left(a\frac{\underline{\alpha}^2}{2}+b\frac{(1-\underline{\alpha})^2}{2}\right)} < 1.$$

Hence,

$$\begin{aligned} & \frac{\left(a\frac{\overline{\alpha}^2}{2}+b\frac{(1-\overline{\alpha})^2}{2}\right)}{\left(a\frac{\underline{\alpha}^2}{2}+b\frac{(1-\underline{\alpha})^2}{2}\right)} \left(\overline{\theta}-\underline{\theta}\right)\lambda V\left(w^B\left(\overline{\alpha},\overline{\beta}\left(\overline{\alpha},\underline{\alpha}\right),\overline{\theta},\sigma^2\right)\right)+(1-\lambda)V\left(w^B\left(\underline{\alpha},\underline{\beta}\left(\underline{\alpha}\right),\underline{\theta},\sigma^2\right)\right)\left(-b\sigma^2\left(1-\underline{\alpha}\right)+a\underline{\alpha}\sigma^2\right) \\ > & \left(\overline{\theta}-\underline{\theta}\right)\lambda V\left(w^B\left(\overline{\alpha},\overline{\beta}\left(\overline{\alpha},\underline{\alpha}\right),\overline{\theta},\sigma^2\right)\right)+(1-\lambda)V\left(w^B\left(\underline{\alpha},\underline{\beta}\left(\underline{\alpha}\right),\underline{\theta},\sigma^2\right)\right)\left(-b\sigma^2\left(1-\underline{\alpha}\right)+a\underline{\alpha}\sigma^2\right) \\ = & \ 0, \end{aligned}$$

where the inequality follows because the negative term in the first-order condition gets a smaller weight and the equality follows from the first-order condition itself.

I now show that $\underline{\alpha}^*$ is strictly positive for λ small enough, and $\underline{\alpha}^* = 0$ as λ gets large. Using again the implicit function theorem, I obtain

$$\frac{d\underline{\alpha}}{d\lambda} = -\frac{\frac{\partial}{\partial\underline{\alpha}}V\left(w^B\left(\overline{\alpha},\overline{\beta}\left(\overline{\alpha},\underline{\alpha}\right),\overline{\theta},\sigma^2\right)\right)-\frac{\partial}{\partial\underline{\alpha}}V\left(w^B\left(\underline{\alpha},\underline{\beta}\left(\underline{\alpha}\right),\underline{\theta},\sigma^2\right)\right)}{\lambda\frac{\partial^2}{\partial\underline{\alpha}^2}V\left(w^B\left(\overline{\alpha},\overline{\beta}\left(\overline{\alpha},\underline{\alpha}\right),\overline{\theta},\sigma^2\right)\right)+(1-\lambda)\frac{\partial^2}{\partial\underline{\alpha}^2}V\left(w^B\left(\underline{\alpha},\underline{\beta}\left(\underline{\alpha}\right),\underline{\theta},\sigma^2\right)\right)} < 0,$$

because both terms in the numerator are negative.

I now compute the critical λ where $\underline{\alpha}^*$ is equal to zero. Using the expression for $\underline{\alpha}^*$, $\underline{\alpha}^* = 0$ is equivalent to solving the equality

$$\frac{1-\lambda}{\lambda} = \frac{V\left(w^B\left(\overline{\alpha}^*,\overline{\beta}\left(\overline{\alpha}^*,0\right),\overline{\theta},\sigma^2\right)\right)}{V\left(w^B\left(0,\underline{\beta}\left(0\right),\underline{\theta},\sigma^2\right)\right)}\frac{\overline{\theta}-\underline{\theta}}{\sigma^2}\frac{a+b}{b}.$$

Hence, there is a unique value of λ that solves this equation. ■

7.3 The partnership model with a continuum of types

Proof of Lemma 1. Note that the agent's utility is an increasing and concave function of his certainty equivalent wealth. Therefore, I can analyze the incentive compatibility of reports in terms of a maximization of certainty equivalent wealth.

I begin showing that monotonicity of $\alpha(\theta)$ in θ is necessary for truthtelling. Truthfulness requires that any type θ should not want to mimic type θ', that is send report $\hat{\theta} = \theta'$ in a direct mechanism. Formally, this requires that

$$\beta(\theta) + \alpha(\theta)\theta - \frac{a\alpha(\theta)^2\sigma^2}{2} \geq \beta(\theta') + \alpha(\theta')\theta - \frac{a\alpha(\theta')^2\sigma^2}{2}.$$

Likewise, a type θ' should not want to send report $\hat{\theta} = \theta$. Formally,

$$\beta\left(\theta'\right) + \alpha\left(\theta'\right)\theta' - \frac{a\alpha\left(\theta'\right)^2\sigma^2}{2} \geq \beta\left(\theta\right) + \alpha\left(\theta\right)\theta' - \frac{a\alpha\left(\theta\right)^2\sigma^2}{2}.$$

Adding the two inequalities and simplifying, I obtain

$$\left(\alpha\left(\theta\right) - \alpha\left(\theta'\right)\right)\left(\theta - \theta'\right) \geq 0,$$

which proves the claim.

Since monotonic functions are differentiable almost everywhere, the first-order condition

$$\beta_{\hat{\theta}}\left(\hat{\theta}\right)\Big|_{\hat{\theta}=\theta} = \left(\left(a\sigma^2\alpha\left(\hat{\theta}\right) - \theta\right)\alpha_{\hat{\theta}}\left(\hat{\theta}\right)\right)\Big|_{\hat{\theta}=\theta} \tag{31}$$

must necessarily hold almost everywhere. Let $w^A\left(\theta\right) = \max_{\hat{\theta}} \beta\left(\hat{\theta}\right) + \alpha\left(\hat{\theta}\right)\theta - \frac{a\alpha\left(\hat{\theta}\right)^2\sigma^2}{2}$. Using the first order condition and the envelope theorem, one observes that the agent's equilibrium certainty equivalent $w^A\left(\theta\right)$ changes at rate $\alpha\left(\theta\right)$ with an increase in θ, that is $w^A_\theta\left(\theta\right)$. Hence, I can write $w^A\left(\theta\right) = w^A\left(\underline{\theta}\right) + \int_{\underline{\theta}}^{\theta} w^A_\tau\left(\tau\right)d\tau$. Imposing individual rationality at $\underline{\theta}$, i.e., $\omega = \beta\left(\underline{\theta}\right) + \underline{\theta}\alpha\left(\underline{\theta}\right) - a\frac{\sigma^2}{2}\alpha\left(\underline{\theta}\right)^2$, this implies that

$$\beta\left(\theta\right) = \omega + a\frac{\sigma^2}{2}\alpha\left(\theta\right)^2 - \theta\alpha\left(\theta\right) + \int_{\underline{\theta}}^{\theta}\alpha\left(\tau\right)d\tau.$$

These conditions are also sufficient for truthtelling to be a global optimum. The global truth-telling condition can be written as

$$w^A\left(\theta\right) - w^A\left(\hat{\theta}\right) \geq \alpha\left(\hat{\theta}\right)\left(\theta - \hat{\theta}\right)$$

Using the individual rationality constraint at $\theta = \underline{\theta}$ and the condition $w^A_\theta\left(\theta\right) = \alpha\left(\theta\right)$, the global truthtelling condition is equivalent to the condition

$$\int_{\hat{\theta}}^{\theta}\left(\alpha\left(\tau\right) - \alpha\left(\hat{\theta}\right)\right)d\tau \geq 0$$

The monotonicity condition $\alpha_\theta\left(\theta\right)$ renders the integrand on the left hand side of this inequality pointwise nonnegative.

The argument proving $\alpha\left(\theta\right) \geq 0$ is necessary for an optimal contract is in the text. ■

Proof of Proposition 3. Let α^*, r^* denote a solution to the problem. By the maximum principle, a solution satisfies the conditions

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\alpha} &= 0 \Leftrightarrow J\left(\theta, \alpha^*, r^*\right)b\sigma^2\left(\left(a+b\right)\alpha^* - b\right) + \left(\lambda + \mu\right) = 0 \\
\frac{\partial\mathcal{L}}{\partial r} &= -\dot{\lambda} \Leftrightarrow J\left(\theta, \alpha^*, r^*\right)b = -\dot{\lambda} \\
\lambda\left(\overline{\theta}\right) &= 0 \\
\alpha &\geq 0; \mu \geq 0; \alpha\mu = 0
\end{aligned}$$

These conditions are necessary and sufficient for an optimal contract, because problem (11) satisfies the conditions of Mangasarian's sufficiency theorem. The solution to the problem without the constraint $\alpha \geq 0$ is obtained from these conditions setting $\mu \equiv 0$. By the conditions of the maximum principle

$$\begin{aligned} J\left(\theta, \underline{\alpha}, \underline{r}\right) b\sigma^2 \left(\left(a+b\right)\underline{\alpha} - b\right) + \underline{\lambda} &= 0 \qquad (32) \\ J\left(\theta, \underline{\alpha}, \underline{r}\right) b &= -\dot{\underline{\lambda}} \qquad (33) \end{aligned}$$

Integrating condition (33) I obtain $\underline{\lambda}$. From the transversality condition I get

$$\underline{\lambda}\left(\theta\right) = \underline{\lambda}\left(\overline{\theta}\right) - \int_{\theta}^{\overline{\theta}} \dot{\underline{\lambda}} d\tau = b \int_{\theta}^{\overline{\theta}} J\left(\tau, \underline{\alpha}, \underline{r}\right) d\tau$$

Combining this condition with condition (32) I obtain condition (13). If the expression on the right-hand side of (13) is nonnegative, then $\alpha^*\left(\theta\right) = \underline{\alpha}\left(\theta\right)$. Otherwise, $\alpha^*\left(\theta\right) = 0$.

Equivalently, one may express the solution as a differential equation. A total differentiation of (32) delivers

$$\begin{aligned} J\left(\theta, \underline{\alpha}, \underline{r}\right) \frac{f_\theta\left(\theta\right)}{f\left(\theta\right)} b\sigma^2 \left(\left(a+b\right)\underline{\alpha} - b\right) + J\left(\theta, \underline{\alpha}, \underline{r}\right) b\sigma^2 \left(a+b\right)\dot{\underline{\alpha}} \\ + J\left(\theta, \underline{\alpha}, \underline{r}\right) b\sigma^2 \left(\left(a+b\right)\underline{\alpha} - b\right)\left[-b + b\sigma^2\left(\left(a+b\right)\underline{\alpha} - b\right)\dot{\underline{\alpha}} + b\underline{\alpha}\right] = -\dot{\underline{\lambda}} \end{aligned}$$

This condition can be combined with (33) to eliminate $\dot{\underline{\lambda}}$. After dividing the resulting equation on both sides by the common factor $J\left(\theta, \underline{\alpha}, \underline{r}\right) b$ one has

$$\begin{aligned} 1 \;=\; & \frac{f_\theta\left(\theta\right)}{f\left(\theta\right)} \sigma^2 \left(\left(a+b\right)\underline{\alpha} - b\right) + \sigma^2 \left(a+b\right)\dot{\underline{\alpha}} + \\ & \sigma^2 \left(\left(a+b\right)\underline{\alpha} - b\right)\left[-b + b\sigma^2\left(\left(a+b\right)\underline{\alpha} - b\right)\dot{\underline{\alpha}} + b\underline{\alpha}\right] \end{aligned}$$

Simplification delivers the expression (14).

The solution is strictly monotonic if condition (10) holds. To see this, rearrange condition (14) to obtain

$$1 - \sigma^2\left(\left(a+b\right)\underline{\alpha} - b\right)\left[\frac{f_\theta\left(\theta\right)}{f\left(\theta\right)} - b + b\underline{\alpha}\right] = b\left[\sigma^2\left(\left(a+b\right)\underline{\alpha} - b\right)\right]^2 \dot{\underline{\alpha}} + \sigma^2\left(a+b\right)\dot{\underline{\alpha}} \qquad (34)$$

Clearly, $\dot{\underline{\alpha}} \geq 0$ if the left hand side of equation (34) is positive. Since there is too little risk sharing relative to the first-best outcome, the solution satisfies $\left(a+b\right)\underline{\alpha} - b \leq 0$. For any sign of the expression $\frac{f_\theta(\theta)}{f(\theta)} - \left(1 - \underline{\alpha}\right) b$ I can derive a lower bound of the expression on the left hand side. The function $g\left(\alpha\right) = 1 - \sigma^2\left(\left(a+b\right)\alpha - b\right)\left[\frac{f_\theta(\theta)}{f(\theta)} - b + b\alpha\right]$ is strictly concave in α. Therefore, it attains its minimum over any interval either at the lower or the upper bound. At $\underline{\alpha} = \frac{b}{a+b}$ I have

$g(\underline{\alpha})|_{\underline{\alpha}=\frac{b}{a+b}} = 1 > 0$. At $\underline{\alpha} = 0$, I have $g(\underline{\alpha})|_{\underline{\alpha}=0} = 1 + \sigma^2 b\left[\frac{f_\theta(\theta)}{f(\theta)} - b\right]$, which is positive under condition (10) . ■

Proof of Proposition 5. Consider the condition for $\underline{\alpha}(\theta)$:

$$\underline{\alpha}(\theta) = \frac{b}{(a+b)} - \frac{\int_\theta^{\overline{\theta}} J(\tau, \underline{\alpha}(\tau), \underline{r}(\tau))\, d\tau}{(a+b)\sigma^2 J(\theta, \underline{\alpha}, \underline{r})}$$

I derive a lower and an upper bound on $\underline{\alpha}(\theta)$, in that sequence.

Let $X(\theta, \underline{\alpha}, \underline{r}) = \frac{J(\theta,\underline{\alpha},\underline{r})}{f(\theta)}$. I find that

$$\frac{\partial}{\partial\theta} X(\theta, \underline{\alpha}, \underline{r}) = -bX(\theta, \underline{\alpha}, \underline{r})\left(1 - \underline{\alpha} + (b - (a+b)\underline{\alpha})\sigma^2 \dot{\underline{\alpha}}\right)$$

since $X(\theta, \underline{\alpha}, \underline{r}) < 0$, $\underline{\alpha} \leq \frac{b}{a+b} \forall\theta$, and $\dot{\underline{\alpha}} \geq 0$, I note that $\frac{\partial}{\partial\theta} X(\theta, \underline{\alpha}, \underline{r}) \geq 0 \forall\theta$. Hence

$$\int_\theta^{\overline{\theta}} X(\tau, \underline{\alpha}, \underline{r}) f(\tau)\, d\tau \geq X(\theta, \underline{\alpha}, \underline{r}) \int_\theta^{\overline{\theta}} f(\tau)\, d\tau \tag{35}$$

Dividing inequality (35) by $-X(\theta, \underline{\alpha}, \underline{r})(a+b)\sigma^2 f(\theta) > 0$ and adding $\frac{b}{(a+b)}$ on both sides I get

$$\frac{b}{(a+b)} - \frac{\int_\theta^{\overline{\theta}} X(\tau, \underline{\alpha}, \underline{r}) f(\tau)\, d\tau}{X(\theta, \underline{\alpha}, \underline{r})(a+b)\sigma^2 f(\theta)} \geq \frac{b}{(a+b)} - \frac{X(\theta, \underline{\alpha}, \underline{r}) \int_\theta^{\overline{\theta}} f(\tau)\, d\tau}{X(\theta, \underline{\alpha}, \underline{r})(a+b)\sigma^2 f(\theta)}$$

and thus

$$\underline{\alpha}(\theta) \geq \frac{b}{a+b} - \frac{1 - F(\theta)}{(a+b)\sigma^2 f(\theta)}$$

Hence, for

$$b\sigma^2 > \frac{1 - F(\theta)}{f(\theta)}$$

I have $\underline{\alpha}(\theta) > 0$.

I now derive an upper bound on $\alpha(\theta)$.

$$\int_\theta^{\overline{\theta}} X(\tau, \underline{\alpha}, \underline{r}) f(\tau)\, d\tau \leq X(\overline{\theta}, \underline{\alpha}, \underline{r}) \int_\theta^{\overline{\theta}} f(\tau)\, d\tau. \tag{36}$$

Dividing (36) by $X(\theta, \underline{\alpha}, \underline{r})(a+b)\sigma^2 f(\theta) < 0$, I get

$$\frac{\int_\theta^{\overline{\theta}} X(\tau, \underline{\alpha}, \underline{r}) f(\tau)\, d\tau}{X(\theta, \underline{\alpha}, \underline{r})(a+b)\sigma^2 f(\theta)} \geq \frac{X(\overline{\theta}, \underline{\alpha}, \underline{r}) \int_\theta^{\overline{\theta}} f(\tau)\, d\tau}{X(\theta, \underline{\alpha}, \underline{r})(a+b)\sigma^2 f(\theta)}.$$

Thus,

$$\frac{b}{(a+b)} - \frac{\int_\theta^{\overline{\theta}} X(\tau, \underline{\alpha}, \underline{r}) f(\tau)\, d\tau}{X(\theta, \underline{\alpha}, \underline{r})(a+b)\sigma^2 f(\theta)} = \underline{\alpha}(\theta) \leq \frac{b}{(a+b)} - \frac{X(\overline{\theta}, \underline{\alpha}, \underline{r}) \int_\theta^{\overline{\theta}} f(\tau)\, d\tau}{X(\theta, \underline{\alpha}, \underline{r})(a+b)\sigma^2 f(\theta)}.$$

Recalling that

$$X(\theta, \underline{\alpha}, \underline{r}) \equiv -\exp\left[-b\left(-w + \theta - \frac{a\alpha^2 + b(1-\alpha)^2}{2}\sigma^2 - \int_{\underline{\theta}}^{\theta} \alpha(\tau)\, d\tau\right)\right],$$

it is easy to see that

$$\frac{X\left(\overline{\theta},\underline{\alpha},\underline{r}\right)}{X\left(\theta,\underline{\alpha},\underline{r}\right)}=\exp\left[-b\left(\overline{\theta}-\theta-\frac{ab}{2\left(a+b\right)}\sigma^{2}+\frac{a\alpha\left(\theta\right)^{2}+b\left(1-\alpha\left(\theta\right)\right)^{2}}{2}\sigma^{2}-\int_{\theta}^{\overline{\theta}}\alpha\left(\tau\right)d\tau\right)\right]$$

where I have used $\underline{\alpha}\left(\overline{\theta}\right)=\frac{b}{a+b}$. Clearly, the expression on the right-hand side of this equality decreases if I set $\alpha\left(\theta\right)=0$ and omit the term $\int_{\theta}^{\overline{\theta}}\alpha\left(\tau\right)d\tau$. So,

$$\begin{aligned}\frac{X\left(\overline{\theta},\underline{\alpha},\underline{r}\right)}{X\left(\theta,\underline{\alpha},\underline{r}\right)} &\geq \exp\left[-b\left(\overline{\theta}-\theta-\frac{ab}{2\left(a+b\right)}\sigma^{2}+\frac{b}{2}\sigma^{2}\right)\right]\\ &= \exp\left[-b\left(\overline{\theta}-\theta+\frac{\sigma^{2}}{2}\frac{b^{2}}{a+b}\right)\right].\end{aligned}$$

Hence, obviously also

$$-\frac{X\left(\overline{\theta},\underline{\alpha},\underline{r}\right)\int_{\theta}^{\overline{\theta}}f\left(\tau\right)d\tau}{X\left(\theta,\underline{\alpha},\underline{r}\right)\left(a+b\right)\sigma^{2}f\left(\theta\right)}\leq-\exp\left[-b\left(\overline{\theta}-\theta+\frac{\sigma^{2}}{2}\frac{b^{2}}{a+b}\right)\right]\frac{\left(1-F\left(\theta\right)\right)}{\left(a+b\right)\sigma^{2}f\left(\theta\right)}$$

and therefore

$$\underline{\alpha}\left(\theta\right)\leq\frac{b}{\left(a+b\right)}-\exp\left[-b\left(\overline{\theta}-\theta+\frac{\sigma^{2}}{2}\frac{b^{2}}{a+b}\right)\right]\frac{\left(1-F\left(\theta\right)\right)}{\left(a+b\right)\sigma^{2}f\left(\theta\right)}.$$

Taken together I have now that

$$\frac{b}{a+b}-\frac{1-F\left(\theta\right)}{\left(a+b\right)\sigma^{2}f\left(\theta\right)}\leq\underline{\alpha}\left(\theta\right)\leq\frac{b}{\left(a+b\right)}-\exp\left[-b\left(\overline{\theta}-\theta+\frac{\sigma^{2}}{2}\frac{b^{2}}{a+b}\right)\right]\frac{\left(1-F\left(\theta\right)\right)}{\left(a+b\right)\sigma^{2}f\left(\theta\right)}$$

As b becomes large, eventually

$$b>\frac{1-F\left(\theta\right)}{\sigma^{2}f\left(\theta\right)}\text{ for all }\theta$$

and hence $\underline{\alpha}\left(\theta\right)>0$ for all θ.

For any $\tilde{\theta}<\overline{\theta}$, the upper satisfies

$$\lim_{b\rightarrow0}\left[\frac{b}{\left(a+b\right)}-\exp\left[-b\left(\overline{\theta}-\theta+\frac{\sigma^{2}}{2}\frac{b^{2}}{a+b}\right)\right]\frac{\left(1-F\left(\tilde{\theta}\right)\right)}{\left(a+b\right)\sigma^{2}f\left(\tilde{\theta}\right)}\right]=-\frac{\left(1-F\left(\tilde{\theta}\right)\right)}{a\sigma^{2}f\left(\tilde{\theta}\right)}<0.$$

Hence, in the limit as b goes to zero, the principal does not share risks any more. ■

7.4 The model of managerial style

Proof of Proposition 6. The proof is given in several steps. First, I show that under the assumption, the state variable constraint is binding for some types. Second, I spell out the problem. Then I prove that the optimal schedule involves bunching between the diligent and the negligent type over a convex region at the high end of the support. Finally, I show that the bunching region extends over the whole support.

I) The state variable constraint must be binding at the optimum.

Suppose the contrapositive were true. Then we can characterize the optimal schedules $\widetilde{\underline{\alpha}}(\theta)$ and $\widetilde{\overline{\alpha}}(\theta)$ by the conditions

$$\frac{1-\widetilde{\underline{\alpha}}(\theta)}{c_{ee}\left(c_e^{-1}\left(\widetilde{\underline{\alpha}}(\theta)\right)\right)}-\frac{1-F\left(\theta|\,\underline{\eta}\right)}{f\left(\theta|\,\underline{\eta}\right)}=0 \tag{37}$$

and

$$\frac{1-\widetilde{\overline{\alpha}}(\theta)}{c_{ee}\left(c_e^{-1}\left(\widetilde{\overline{\alpha}}(\theta)\right)\right)}-\frac{1-F\left(\theta|\,\overline{\eta}\right)}{f\left(\theta|\,\overline{\eta}\right)}=0. \tag{38}$$

If these schedules are monotonic, then they describe the optimum. However, irrespective of whether they are monotonic or not, I now argue that they violate the state variable constraint. As is well known (see Shaked and Shantikumar (2007)), affiliation of θ and $-\eta$ implies that

$$\frac{1-F\left(\theta|\,\underline{\eta}\right)}{f\left(\theta|\,\underline{\eta}\right)}>\frac{1-F\left(\theta|\,\overline{\eta}\right)}{f\left(\theta|\,\overline{\eta}\right)}\text{ for all }\theta, \tag{39}$$

and this in turn implies from (37) and (38) that $\widetilde{\underline{\alpha}}(\theta)<\widetilde{\overline{\alpha}}(\theta)$ for all θ. But then, in order to render constraint (26) slack, the principal needs to leave a rent $\overline{\rho}$ to the diligent type, where

$$\overline{\rho}\equiv\int_{\underline{\theta}}^{\overline{\theta}}\widetilde{\overline{\alpha}}(\tau)\,d\tau-\int_{\underline{\theta}}^{\overline{\theta}}\widetilde{\underline{\alpha}}(\tau)\,d\tau.$$

However, if (26) is slack, then the shadow cost of the constraint is zero; on the other hand the marginal benefit of reducing ρ slightly is p. Hence, it is optimal to reduce ρ below $\overline{\rho}$.

II) Statement of the problem: I maximize sequentially with respect to the schedules and with respect to ρ. The first step problem (for given ρ) can be viewed as a control problem with Hamiltonian of the following form (up to constants):

$$\begin{aligned}H &= p\left(c_e^{-1}(\underline{\alpha})+\theta-c\left(c_e^{-1}(\underline{\alpha})\right)-\frac{1-F\left(\theta|\,\underline{\eta}\right)}{f\left(\theta|\,\underline{\eta}\right)}\underline{\alpha}\right)f\left(\theta|\,\underline{\eta}\right)\\ &\quad+(1-p)\left(\left(c_e^{-1}(\overline{\alpha})+\theta\right)-c\left(c_e^{-1}(\overline{\alpha})\right)-\frac{1-F\left(\theta|\,\overline{\eta}\right)}{f\left(\theta|\,\overline{\eta}\right)}\overline{\alpha}\right)f\left(\theta|\,\overline{\eta}\right)\\ &\quad+\underline{\kappa}\underline{\alpha}+\overline{\kappa}\overline{\alpha}+\mu\left(\rho-\left(\overline{z}-\underline{z}\right)\right)\end{aligned}$$

where $\overline{z}\equiv\int_{\underline{\theta}}^{\theta}\underline{\alpha}(\tau)\,d\tau$ and $\underline{z}\equiv\int_{\underline{\theta}}^{\theta}\overline{\alpha}(\tau)\,d\tau$ are the state variables, $\underline{\kappa}$ and $\overline{\kappa}$ are the costate variables associated to them, and μ is the multiplier on the inequality constraint. Of course, all these variables depend on θ; I do not make this dependence explicit to keep notation compact.

Differentiating with respect to state variables, I get the conditions of optimality

$$\begin{aligned}\frac{\partial H}{\partial\overline{z}} &= -\mu=-\overline{\kappa}_{\theta}\\ \frac{\partial H}{\partial\underline{z}} &= \mu=-\underline{\kappa}_{\theta};\end{aligned}$$

differentiating with respect to the controls I get

$$\begin{aligned} \frac{\partial H}{\partial \underline{\alpha}} &= \left(\frac{1-\underline{\alpha}\left(\theta\right)}{c_{ee}\left(c_e^{-1}\left(\underline{\alpha}\left(\theta\right)\right)\right)} - \frac{1-F\left(\theta|\,\underline{\eta}\right)}{f\left(\theta|\,\underline{\eta}\right)} \right) pf\left(\theta|\,\underline{\eta}\right) + \underline{\kappa} = 0 \\ \frac{\partial H}{\partial \overline{\alpha}} &= \left(\frac{1-\overline{\alpha}\left(\theta\right)}{c_{ee}\left(c_e^{-1}\left(\overline{\alpha}\left(\theta\right)\right)\right)} - \frac{1-F\left(\theta|\,\overline{\eta}\right)}{f\left(\theta|\,\overline{\eta}\right)} \right) \left(1-p\right) f\left(\theta|\,\overline{\eta}\right) + \underline{\kappa} = 0 \end{aligned} \tag{40}$$

The Kuhn-Tucker conditions are

$$\rho - \left(\overline{z} - \underline{z}\right) \leq 0,\ \ \mu \geq 0,\ \text{ and } \mu\left(\rho - \left(\overline{z} - \underline{z}\right)\right) = 0.$$

For the transversality conditions, I have to distinguish two cases. If $\mu\left(\overline{\theta}\right) = 0$, then $\overline{z}\left(\overline{\theta}\right)$ and $\underline{z}\left(\overline{\theta}\right)$ are both free and the transversality conditions are

$$\overline{\kappa}\left(\overline{\theta}\right) = \underline{\kappa}\left(\overline{\theta}\right) = 0.$$

If $\mu\left(\overline{\theta}\right) > 0$, then $\overline{z}\left(\overline{\theta}\right)$ is fully determined once $\underline{z}\left(\overline{\theta}\right)$ is given and vice versa. Hence, I do not impose any transversality condition in this case.

Suppose that $\mu\left(\overline{\theta}\right) = 0$ and that $\mu\left(\theta\right) = 0$ on a set of positive measure $\left[\theta', \overline{\theta}\right]$. From conditions (40) it is clear that $\overline{\kappa}$ and $\underline{\kappa}$ are continuously differentiable in θ whenever $\overline{\alpha}$ and $\underline{\alpha}$ are continuously differentiable in θ. Using the conditions of optimality for the state variables, $\overline{\kappa}_\theta = \mu$ and $\underline{\kappa}_\theta = -\mu$, and the transversality conditions - which must hold if $\mu\left(\overline{\theta}\right) = 0$ - I have for $\theta \geq \theta'$

$$\overline{\kappa}\left(\theta\right) = \overline{\kappa}\left(\overline{\theta}\right) - \int_{\theta}^{\overline{\theta}} \overline{\kappa}_\tau d\tau = -\int_{\theta}^{\overline{\theta}} \mu\left(\tau\right) d\tau = 0$$

and

$$\underline{\kappa}\left(\theta\right) = \underline{\kappa}\left(\overline{\theta}\right) - \int_{\theta}^{\overline{\theta}} \overline{\kappa}_\tau d\tau = \int_{\theta}^{\overline{\theta}} \mu\left(\tau\right) d\tau = 0.$$

Hence, for $\theta \in \left[\theta', \overline{\theta}\right]$, (40) is equivalent to conditions (37) and (38). However, from the argument made in part I above we know that this would imply $\underline{\alpha}\left(\theta\right) > \overline{\alpha}\left(\theta\right)$ for $\theta \in \left[\theta', \overline{\theta}\right]$. Let $R\left(\theta\right) \equiv \rho + \int_{\underline{\theta}}^{\theta} \underline{\alpha}\left(\tau\right) d\tau - \int_{\underline{\theta}}^{\theta} \overline{\alpha}\left(\tau\right) d\tau$ denote the excess rent of the diligent type over the negligent one. If $\mu\left(\theta'\right) > 0$, then $R\left(\theta'\right)$. On the other hand, I have just shown that $R_\theta\left(\theta\right) = \underline{\alpha}\left(\theta\right) - \overline{\alpha}\left(\theta\right) < 0$ for $\theta > \theta'$. However, taken together this would imply that $R\left(\theta\right) < 0$ for $\theta > \theta'$, violating the state variable constraint and thus leading to a contradiction. Hence, the state variable constraint must be binding at the high end of the support.

III) If the state variable constraint is binding over an interval, then $\underline{\alpha}\left(\theta\right) = \overline{\alpha}\left(\theta\right)$ over that interval.

Differentiating the definition of the excess rent, I get $R_\theta(\theta) = \underline{\alpha}(\theta) - \overline{\alpha}(\theta)$. Clearly, if $R(\theta) = 0$ over an interval, then $R_\theta(\theta) = 0$ over the same interval, and hence $\underline{\alpha}(\theta) = \overline{\alpha}(\theta)$.

IV) The value of costate variables at the upper bound:

I now argue that $\overline{\kappa}\left(\overline{\theta}\right) + \underline{\kappa}\left(\overline{\theta}\right) = 0$.

To see this, notice that the costate variables are allowed to jump at points where the state variable constraint switches from being binding to nonbinding and vice versa. This means that the schedule $\alpha(\theta) \equiv \underline{\alpha}(\theta) = \overline{\alpha}(\theta)$ must maximize the value of the objective function absent the state variable constraint over the bunching region (since the state variable constraint is automatically satisfied if $\underline{\alpha}(\theta) = \overline{\alpha}(\theta)$). This is consistent with (40) only if $\overline{\kappa}\left(\overline{\theta}\right) + \underline{\kappa}\left(\overline{\theta}\right) = 0$, in which case adding up the two conditions in (40) (and using the definition of $F(\theta)$ and $f(\theta)$) gives

$$\frac{1-\alpha(\theta)}{c_{ee}\left(c_e^{-1}(\alpha(\theta))\right)} - \frac{1-F(\theta)}{f(\theta)} = 0.$$

V) Convexity of the bunching region:

Suppose there is an interval $[\theta_1, \theta_2]$, with $\underline{\theta} < \theta_1 < \theta_2 < \overline{\theta}$, over which the state variable constraint is non-binding. I show that this is not possible, proving that the bunching region is convex.

Again, because the state variables are allowed to jump at points where the state variable constraint switches from being binding to nonbinding or vice versa, the schedules $\underline{\alpha}(\theta)$ and $\overline{\alpha}(\theta)$ cannot be optimal, unless they maximize the problem over the interval $[\theta_1, \theta_2]$ in isolation. This subproblem corresponds to the problem of maximizing (25) with θ_1, θ_2 replacing $\underline{\theta}, \overline{\theta}$ subject to the constraint

$$\int_{\theta_1}^{\theta_2} \underline{\alpha}(\tau)\, d\tau - \int_{\theta_1}^{\theta_2} \overline{\alpha}(\tau)\, d\tau = 0.$$

This is an isoperimetric problem. Letting k denote the multiplier attached to the integral constraint (note that k is independent of θ), the solution satisfies the first-order conditions:.

$$\frac{1-\underline{\alpha}(\theta)}{c_{ee}\left(c_e^{-1}(\underline{\alpha}(\theta))\right)} = \frac{1-F\left(\theta|\,\underline{\eta}\right)}{f\left(\theta|\,\underline{\eta}\right)} - \frac{k}{pf\left(\theta|\,\underline{\eta}\right)} \tag{41}$$

and

$$\frac{1-\overline{\alpha}(\theta)}{c_{ee}\left(c_e^{-1}(\overline{\alpha}(\theta))\right)} = \frac{1-F(\theta|\,\overline{\eta})}{f(\theta|\,\overline{\eta})} + \frac{k}{(1-p)\, f(\theta|\,\overline{\eta})} \tag{42}$$

Notice that at least one of the schedules cannot take values in excess of unity. Hence, any point of intersection must be at a value of α below unity.

I now show that the schedule $\underline{\alpha}(\theta)$ increases faster than the schedule $\overline{\alpha}(\theta)$ at points where

$\underline{\alpha}(\theta) = \overline{\alpha}(\theta)$. Differentiating (41) and (42) totally I obtain

$$\frac{d\underline{\alpha}}{d\theta} = \frac{1 + \frac{f_\theta\left(\theta|\underline{\eta}\right)}{f\left(\theta|\underline{\eta}\right)}\left(\frac{\left(1-F\left(\theta|\underline{\eta}\right)\right)}{f\left(\theta|\underline{\eta}\right)} - \frac{k}{pf\left(\theta|\underline{\eta}\right)}\right)}{-\frac{\partial}{\partial\alpha}\frac{1-\underline{\alpha}(\theta)}{c_{ee}\left(c_e^{-1}(\underline{\alpha}(\theta))\right)}}$$

and

$$\frac{d\overline{\alpha}}{d\theta} = \frac{1 + \frac{f_\theta(\theta|\overline{\eta})}{f(\theta|\overline{\eta})}\left(\frac{(1-F(\theta|\overline{\eta}))}{f(\theta|\overline{\eta})} + \frac{k}{(1-p)f(\theta|\overline{\eta})}\right)}{-\frac{\partial}{\partial\alpha}\frac{1-\overline{\alpha}(\theta)}{c_{ee}\left(c_e^{-1}(\overline{\alpha}(\theta))\right)}}$$

We know that the relevant range in which the two schedules can cross is below unity. At points where $\underline{\alpha}(\theta) = \overline{\alpha}(\theta)$, the denominators are equal and the expression in brackets in the numerators are also equal and positive.. Moreover, from the affiliation assumption, $\frac{f_\theta\left(\theta|\underline{\eta}\right)}{f\left(\theta|\underline{\eta}\right)} > \frac{f_\theta(\theta|\overline{\eta})}{f(\theta|\overline{\eta})}$, which completes the argument.

From $\frac{d\underline{\alpha}}{d\theta} > \frac{d\overline{\alpha}}{d\theta}$ at $\underline{\alpha}(\theta) = \overline{\alpha}(\theta)$ it follows that the schedules cross at most once over the interval $[\theta_1, \theta_2]$. Two cases are possible: i) $\underline{\alpha}(\theta_1) \geq \overline{\alpha}(\theta_1)$ and ii) $\underline{\alpha}(\theta_1) < \overline{\alpha}(\theta_1)$. In case i) the schedules can cross if at all only at θ_1 but not again over the interval $[\theta_1, \theta_2]$. However, this implies that $R_\theta(\theta) = \underline{\alpha}(\theta) - \overline{\alpha}(\theta) > 0$ over the entire interval $(\theta_1, \theta_2]$ and hence $R(\theta_2) > 0$, contradicting the supposition that the state variable constraint becomes binding at θ_2. In case ii) we $R_\theta(\theta_1) = \underline{\alpha}(\theta_1) - \overline{\alpha}(\theta_1) < 0$ and $R(\theta_1) = 0$, which implies that the state variable constraint is violated for some values of θ.

VI) As a simple corollary of the previous argument, note that for $\rho = 0$, the bunching region extends over the whole interval $\left[\underline{\theta}, \overline{\theta}\right]$ (because $R(\underline{\theta}) = 0$).

VII) The previous steps established that the solution is of the form that the state variable constraint is non-binding over some (possibly empty) interval $\left[\underline{\theta}, \theta'\right)$, switches to being binding at θ', and continues to be binding over the remainder of the interval.

$$\max_{\underline{\alpha}(\theta),\overline{\alpha}(\theta),\alpha(\theta)} p\int_{\underline{\theta}}^{\theta'}\left(c_e^{-1}(\underline{\alpha}(\theta)) + \theta - c\left(c_e^{-1}(\underline{\alpha}(\theta))\right) - \frac{1-F\left(\theta|\underline{\eta}\right)}{f\left(\theta|\underline{\eta}\right)}\underline{\alpha}(\theta)\right) f\left(\theta|\underline{\eta}\right) d\theta - p\rho \tag{43}$$

$$+(1-p)\int_{\underline{\theta}}^{\theta'}\left(\left(c_e^{-1}(\overline{\alpha}(\theta)) + \theta\right) - c\left(c_e^{-1}(\overline{\alpha}(\theta))\right) - \frac{1-F(\theta|\overline{\eta})}{f(\theta|\overline{\eta})}\overline{\alpha}(\theta)\right) f(\theta|\overline{\eta})\, d\theta + k\left(\rho + \int_{\underline{\theta}}^{\theta'}(\underline{\alpha}(\theta) - \overline{\alpha}(\theta))\, d\theta\right)$$

$$+\int_{\theta'}^{\overline{\theta}}\left(c_e^{-1}(\alpha(\theta)) + \theta - c\left(c_e^{-1}(\alpha(\theta))\right) - \frac{1-F(\theta)}{f(\theta)}\alpha(\theta)\right) f(\theta)\, d\theta \tag{44}$$

A small variation of the switching point θ' must not change the value of the objective. This requires a value matching condition:

$$\begin{aligned}
& p\left[c_e^{-1}\left(\underline{\alpha}\left(\theta'\right)\right)+\theta'-c\left(c_e^{-1}\left(\underline{\alpha}\left(\theta'\right)\right)\right)\right] f\left(\theta'\middle|\,\underline{\eta}\right)-p\left(1-F\left(\theta'\middle|\,\underline{\eta}\right)\right)\underline{\alpha}\left(\theta'\right) \\
& +(1-p)\left[c_e^{-1}\left(\overline{\alpha}\left(\theta'\right)\right)+\theta'-c\left(c_e^{-1}\left(\overline{\alpha}\left(\theta'\right)\right)\right)\right] f\left(\theta'\middle|\,\overline{\eta}\right)-(1-p)\left(1-F\left(\theta'\middle|\,\overline{\eta}\right)\right)\overline{\alpha}\left(\theta'\right)+k\left(\underline{\alpha}\left(\theta'\right)-\overline{\alpha}\left(\theta'\right)\right) \\
= \; & \left[c_e^{-1}\left(\alpha\left(\theta'\right)\right)+\theta'-c\left(c_e^{-1}\left(\alpha\left(\theta'\right)\right)\right)\right] f\left(\theta'\right)-\left(1-F\left(\theta'\right)\right)\alpha\left(\theta'\right)
\end{aligned}$$

Clearly, $\overline{\alpha}\left(\theta'\right)=\underline{\alpha}\left(\theta'\right)=\alpha^*\left(\theta'\right)$ satisfies the value matching condition. I now show that there is no other solution. To see this, note from (41) and (42) that any adjustment in $\overline{\alpha}\left(\theta'\right)$ and $\underline{\alpha}\left(\theta'\right)$ must come through a change in k. Now, invoking the envelope theorem, the change in the left-hand side value with respect to k is simply $\underline{\alpha}\left(\theta'\right)-\overline{\alpha}\left(\theta'\right)$. Hence, increasing k increases the value of the left-hand side, decreasing k decreases the value. Hence, there is exactly one solution.

VIII) The optimal value of ρ is zero.

Using the value matching condition and the invoking the envelope theorem again, the change in the objective with respect to a change in ρ is

$$-p+k.$$

Hence, either $\rho^*=0$ or $\rho^*>0$ and $k=p$. However, the latter case is impossible, since (41) and (42) would imply in this case that

$$\frac{1-\underline{\alpha}(\theta)}{c_{ee}\left(c_e^{-1}\left(\underline{\alpha}(\theta)\right)\right)}=\frac{-F\left(\theta\middle|\,\underline{\eta}\right)}{f\left(\theta\middle|\,\underline{\eta}\right)} \tag{45}$$

and

$$\frac{1-\overline{\alpha}(\theta)}{c_{ee}\left(c_e^{-1}\left(\overline{\alpha}(\theta)\right)\right)}=\frac{1-F\left(\theta\middle|\,\overline{\eta}\right)}{f\left(\theta\middle|\,\overline{\eta}\right)}+\frac{p}{(1-p)\,f\left(\theta\middle|\,\overline{\eta}\right)} \tag{46}$$

for $\theta \leq \theta'$. However, as $\underline{\alpha}(\theta)>1>\overline{\alpha}(\theta)$ in this case, the schedules cannot cross and in fact $R_\theta(\theta)=\underline{\alpha}(\theta)-\overline{\alpha}(\theta)>0$ for all θ. But this contradicts the fact that the state variable constraint must become binding over some interval at the high end of the support. ■

8 References

- Allen, D. and Lueck, D. (1992) "Contract Choice in Modern Agriculture: Cash Rent versus Crop Share" J. of Law and Economics 35, 397-426

- Allen, D. and Lueck, D. (1995) "Risk, Uncertainty and Contracts" Manuscript New Haven; Yale Law School

- Ackerberg, D., and Botticini, M. (2002) “Endogenous Matching and the Empirical Determinants of Contract Form” Journal of Political Economy, 564-591
- Bandiera, O. Guiso, L., Prat, A. and Sadun, R. (2009) Matching Firms, Managers,and Incentives, CEPR Discussion paper 7207
- Beaudry, P., C. Blackorby and D. Szalay, (2009) “Taxes and Employment Subsidies in Optimal Redistribution Programs ”, American Economic Review 99(1), , 216-242.
- Biais, B., Martimort, D., and Rochet, J.-C. (2000) “Competing Mechanisms in a Common Value Environment” Econometrica 68 (4), 799-838
- Che, Y. and I. Gale, (2000)“The Optimal Mechanism for Selling to a Budget-Constrained Buyer”, Journal of Economic Theory **92**, 198-233.
- Garcia, D. (2008)”Optimal Contracts with Privately Informed Agents and Active Principals” mimeo, UNC at Chapel Hill,
- Hellwig, M., and Schmidt, K. (2002) “Discrete-Time Approximations of the Holmström-Milgrom Brownian-Motion Model of Intertemporal Incentive Provision”, Econometrica 70, 2225-2264.
- Holmström, B. and Milgrom, P. (1987) “Aggregation and Linearity in the Provision of Intertemporal Incentives” Econometrica 55, 303-328
- Inderst, R, and Müller, H. (2009) “CEO Replacement under Private Information” working paper NYU and Frankfurt
- Iyengar, G. and Kumar, A. (2006) “Optimal Procurement Mechanisms for Divisible Goods with Capacitated Suppliers” working paper Columbia University
- Laffont, J.-J. and Rochet, J.-C. (1998) ”Regulation of a Risk Averse Firm” Games and Economic Behavior 25, 149-173
- Lazear, E. (1986) “Salaries and Piece Rates,” Journal of Business, 59 (3), 405-31.
- Leland, H. E., and D. H. Pyle, (1977), “Informational asymmetries, financial structure, and financial intermediation,” Journal of Finance, 32(2), 371–387
- Malakow, A. and R. Vohra, (2005) “An Optimal Auction for Capacitated Bidders - a Network Perspective”, mimeo University of North Carolina and Northwestern University

- Maskin, E. and Riley, J. (1984) ”Optimal Auctions with Risk Averse Buyers” Econometrica 52, 1473-1518
- Matthews, S.A. (1983) ”Selling to Risk Averse Buyers” Journal of Economic Theory 30, 370-400
- Prendergast, C. (1999) ”The Provision of Incentives in Firms”, Journal of Economic Literature, Vol. 37, No. 1 (Mar., 1999), pp. 7-63
- Prendergast, C. (2002) ”The Tenuous Trade-off between Risk and Incentives” Journal of Political Economy 110, 1071-1102
- Rao, C.H. Hanumantha (1971) “Uncertainty, Entrepreneurship, and Sharecropping in India” Journal of Political Economy 79, 578-595
- Raith, M. (2003) “Competition, Risk and Managerial Incentives”, American Economic Review, 93, p1425-1436.
- Salanié, B. (1990) ”Sélection Adverse et Aversion pour le Risque” Ann. Econ. Statist. 18, 131-150
- Seyerstad, K. and A. Sydsaeter (1993)“Optimal Control Theory with Economic Applications; Second reprint”, North Holland, 1999
- Shaked, M. and Shantikumar, J.G., (2007) “Stochastic Orders”, Springer, New York,
- Sung, Y. (2005) ”Optimal Contracts under Moral Hazard and Adverse Selection: a Continuous Time Approach”, Review of Financial Studies, 1021-1073
- Szalay, D. (2009) “Regulating a multi-attribute/multi-type Monopolist” working paper University of Bonn
- Szalay, D. and Yokeeswaran, V. (2009) “Management style and the positive correlation between risk and incentives”, in preparation