Managerial Incentive Problems and Return Distributions*

Dezsö Szalay[†] and Venuga Yokeeswaran

University of Bonn and CEPR

University of Bonn

September 22, 2014

Abstract

We study a model of managerial incentive problems where a manager chooses the first two moments of his firm's profit distribution - mean and volatility - along an efficient frontier. Assuming that managers differ with respect to their marginal cost of effort and their risk aversion we explore our model's comparative statics predictions in full detail. If managers' preference parameters are commonly known and associated, then a positive correlation between expected returns, volatility of profits, and incentives is the natural outcome. Allowing in addition for adverse selection with respect to the managers' preference parameters does not change the predicted correlation if the variation in observed contracts is not too large. Moreover, observed incentive schemes reflect exclusion of some manager types. Neglecting the endogeneity of risk in empirical studies biases estimates towards zero.

JEL Classification: D82, J33

Keywords: Managerial incentive problems, comparative statics, multidimensional heterogeneity, multidimensional screening

*We would like to thank seminar audiences at the University of Bonn, Frankfurt, Rotterdam, Napels, Pompeu Fabra and an SFB-conference meeting where previous versions of this paper were presented. Many thanks in particular to Helmut Bester, Nina Bobkova, Jörg Budde, Inga Deimen, Armin Falk, Guido Friebel, Martin Hellwig, Ian Jewitt, Matthias Kräkel, Salvatore Piccolo, and Klaus Schmidt. Venuga Yokeeswaran acknowledges financial support through a doctoral grant by the DFG.

†Send communications to szalay@uni-bonn.de or by ordinary mail to Dezsö Szalay, Institute for Microeconomics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany.

1 Introduction

The separation of ownership and control (Berle and Means (1932)) makes it vital to understand the optimal design of incentive schemes for managers. The theoretical literature on the subject is vast, leaving no hope to do justice to all contributions. A cornerstone of incentive theory is the Holmström and Milgrom (1987) continuous time model, where a manager's compensation takes the form of a linear compensation scheme; a fixed part plus a variable part that depends linearly on some accounting measure; Hellwig and Schmidt (2002) have provided discrete time approximations for this model. Applied work on contract theory usually starts from a static reduced form version of these models, assuming that a manager receives a compensation package that is linear in profits, and studies how the components of the manager's pay change with the underlying problem. One comparative statics prediction that is shared by the majority of these models is that the sensitivity of the manager's pay to the firm's profits should be the lower the more risky the firm's profits are. Efficient risk sharing between well diversified shareholders and the firm's managers would allocate all the risk to the shareholders, but such an arrangement would give the manager too little incentives to work. Hence, moral hazard induces an inefficiency that is the more costly the larger the underlying risk and so the optimal sensitivity of the manager's performance pay is reduced when the firm's profits become more volatile.

The empirical evidence as to whether the data support this comparative statics prediction is mixed. In the context of executive pay, Core and Guay (1999) and Oyer and Schaefer (2004) find a positive and significant relation between measures of business risk and performance sensitivity of pay; Aggarwal and Samwick (2002) and Lambert and Larcker (1987) find a negative and significant relation between risk and incentives. Quite some studies find results that are statistically not significant: Bushmann et al. (1996) and Ittner et al. (1997) study whether firms are more or less inclined to use individual performance evaluation rather than compensation based on financial performance measures when risk is higher and find a positive result when they take variance in stock returns as the measure of risk; they find a negative result when they take variance in accounting returns as their measure of risk; Ittner et al. (1997) find positive results for various measures of risk (volatility of accounting returns, stock returns and net earnings); Yermack (1994) finds that firms provide more incentives from stock options when accounting earnings contain larger amounts of noise.

We propose a new way to look at this evidence. We develop a theoretical model of performance pay where the manager is given incentives to be diligent in two respects. Firstly, the manager exerts effort which, all else equal, makes higher profits more likely. Secondly, the manager can also choose the firm's strategy, that is, he can select the riskiness of the firm's profits along an efficient frontier. We stick firmly to the applied perspective and assume that the manager faces a compensation package that is linear in profits[1]. The performance sensitivity of the manager's pay determines both his optimal effort choice and the optimal volatility of the firm's profits. The optimal contract is influenced by the manager's underlying characteristics. When these characteristics vary, the observed contract choices vary too and furthermore induce variation in the observed firm characteristics. Hence, our model makes predictions as to the covariation between observed contract choices and firm characteristics, that is, mean and variance of profits. Since we do not in general know whether the characteristics are known to the principals who design the contracts (in practice), we extend our results to allow for adverse selection with respect to the manager's characteristics on top of moral hazard with respect to the choices made by the manager. Under fairly general conditions, we obtain a (pairwise) positive covariance of performance-sensitivity of pay and mean returns and volatility of profits.

If there is a grain of truth to our story, then our model sheds new light on the existing evidence. The hypothesis that risk and incentives should be inversely related is based on a model where risk is exogenous. In contrast, when risk is endogenous through choices made by the managers, then our theoretical model predicts a positive relation. Moreover, in empirical studies endogeneity would not only affect the sign but also the magnitude of the estimated relations, at least when the endogeneity is not entirely accounted for: the resulting correlations between risk as a regressor and the error terms biases the estimates towards zero, explaining why it is difficult to reject the hypothesis that there is no relation between risk and incentives at all.

Our story is closely related to Holmström and Milgrom (1994) and Demski and Dye (1999). Holmström and Milgrom (1994) develop a theory of the joint determination of various elements of contracts. While their theory explains the covariation of choices made by the principal, we wish to explain how choices made by principals (that is, contracts) covary with choices made by the managers (that is, expected level and riskiness of profits). A key element in our theory is an efficient frontier, which introduces a relation between equilibrium expected return and risk. Demski and Dye (1999) also build on the idea that a manager can make mean-variance trade-offs;

[1] This is a standard perspective taken in a sizeable branch of the literature. While restricting contracts to a particular functional form is clearly a restriction, doing so allows us to closely compare our results to those found in the applied literature that works from this hypothesis, which is precisely the aim of the present paper. Thus, the restriction to linear contracts is imposed deliberately, not just for analytical convenience.

however, they address quite different questions with their model.

Thus, the key idea is to allow for more margins of decision making that affect the contracting environment. This idea is also present in Hellwig (2009), Sung (2005), Araujo et al. (2007) and Garcia (2013). All these papers allow for endogenous risk choices, even though the precise trade-offs and who controls the choice of risk differs across the approaches. Hellwig (2009) points out that all moments, mean and risk, are jointly determined as solutions of one incentive problem and thus challenges the way we think of debt contracts as a solution to one incentive problem and equity contracts as a solution to another one. Sung (2005) studies a continuous time principal agent problem with moral hazard and adverse selection which allows for an endogenous choice of volatility by the principal. We use a static model but allow for various sources of heterogeneity among agent types, have all choices except for contracts made by the agent, and explore the comparative statics properties based on the association of random variables as Holmström and Milgrom (1994) do[2]. Araujo et al. (2007) analyze a problem where the manager's effort choice raises means and reduces variance at the same time. In Garcia (2013), risk can again be seen as an additional contracting tool that the principal uses alongside with linear contracts to control the agent's effort choice.

Overall, we believe it is very natural to assume that all the moments of the return distribution are endogenous and find it reassuring that different variations on the same theme share similar results.[3] Many variations and their predictions for empirical work remain unexplored to date.[4]

Part of the empirical contracting literature discusses endogeneity of risk explicitly; see, e.g. Garen (1994) and, in the context of franchising, Lafontaine (1992) and Lafontaine and Slade (2007, 1998). One way to deal with the issue is to find measures of risk that are likely to be exogenous to the firm's choices. Garen (1994) follows this approach and uses R&D intensity as a proxy for the riskiness of a firm's industry. Using that proxy he finds a negative but statistically not significant relation between the pay-performance sensitivity and this proxy. Based on our theory, we propose

[2] Combining Sung's (2005) continuous time with our multidimensional approach is - as we believe - an interesting avenue for future research.

[3] It should also be stressed that incentive problems in practice may depend on the context, ranging from excessive risk taking to excessive conservatism. This paper does not address excessive risk taking by managers, and is therefore clearly not the adequate framework to think about contracts for bank managers, where excessive risk taking is the main concern. For a further discussion, see the final section.

[4] In a recent paper, Weinschenk (2014) studies a model with endogenous project choice to challenge the Marshallian hypothesis that higher incentives lead to higher expected profits - a feature that our model has. He shows that this need not be the case in his model.

an empirical approach that attacks this issue more directly, that is to regress all the choices made by the manager and the firm's owners on the characteristics of the underlying problem. This would allow to estimate the endogenous relation between risk and incentives.

While we are not aware of any study in the context of executive pay that addresses this issue, Ackerberg and Botticini (2002) make a closely related point in the context of sharecropping, pointing out that some characteristics of their underlying contracting problem may be endogenous through tenant/landowner matching. In their context, the landowner decides on what crop to grow; if crops differ in their riskiness, then tenants who differ in their risk aversion feel attracted to different landowners. Similar to their work, we stress that endogeneity is an important issue. However, since the details of optimal choices in a contracting relationship are different from the details in the matching process, our way to address the endogeneity is quite different.

A number of theories can rationalize a positive relation between risk and incentives. The main value added to our exercise is not so much to provide yet another one explaining the same thing but much more to paint a rich picture of the comparative statics predictions of a contracting model allowing for many margins along which managers make choices and for many dimensions of heterogeneity among managers that may or may not be private information within a unified framework. We are not aware of a similar attempt in the literature.

Prendergast (2002) was first to take up the mismatch between theory and empirical work. He argues that 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 a model of oligopolistic competition, where a manager's role is to reduce his firm's costs of production. As 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 managers' 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 managers' incentive schemes. Thus, the agents' pay is more performance dependent when business risk is greater, but there is no causal link between the two effects.

More recently, Inderst and Müller (2010) point to the role of incentive pay-schemes when it comes to inducing exit by bad managers. Comparing severance pay with on the job payment schemes, they find that severance pay makes shirking too attractive for managers; on the other hand, risky pay-for-performance is only attractive to manager who think they are more likely to generate high returns. Moreover, performance pay may be steeper if the underlying firm risk is higher.[5]

As stressed before, the main point of this paper is not so much that the relation between risk and incentives is positive but that it is endogenous and shaped by many factors that may or may not be observed when contracts are written. We develop a framework which allows us to illustrate the implications of this insight for empirical work.

The remainder of this article is structured as follows. In section two, we lay out the model and explain the principal's problem including its solution in the first-best situation in section three. In section four, we study the contracting problem with known characteristics, in section five we extend these results to the case of adverse selection with respect to the manager's characteristics. In section six, we remind the reader of the attenuation problem in empirical studies that arises from endogeneity of regressors, in our case risk. Section seven concludes. All proofs are gathered in the appendix.

2 The Model

An owner of a firm hires a manager to produce output. Henceforth, we call the owner the principal (she) and the manager the agent (he). The distribution of profits, π, depends on the agent's management style, that is two choices the agent makes. In particular, the agent chooses the mean μ and the variance σ^2 of a Gaussian profit distribution, so $\tilde{\pi} \sim N\left(\mu, \sigma^2\right)$. The agent's choices are constrained by an efficient frontier $\mu = \mu\left(e, \sigma\right)$, where e is the agent's effort. The efficient

[5] A positive relation between risk and incentives can be rationalized in a number of other ways, e.g., through endogenous matching between principals and agents (Serfes (2005) and Wright (2004)) or by combining limited liability with risk aversion on the part of the agent (Budde and Kräkel (2011).

frontier describes the maximum expected return the agent can reach for any given variance and effort choice. For a given effort, higher expected returns can only be reached at the cost of higher variance. By increasing his effort, the agent can expand the set of feasible profit distributions; the efficient frontier is increasing in e for any given volatility σ. We assume that $\mu(e, \sigma)$ is jointly concave in e and σ. Finally, there is an upper bound on the volatility, $\overline{\sigma}$. Figure 1 depicts the efficient frontier.[6]

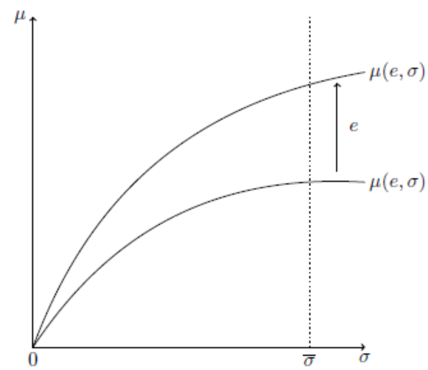

Figure 1: The efficient frontier

Effort is costly to the agent. The cost of effort is $c \cdot e$, where c is a positive parameter[7]. Contracts can only be written on profits; the agent's choices themselves - neither effort nor volatility - are not observable to the principal by the time payments are made. Moreover, the principal is restricted to use linear contracts. So, the principal's wealth is equal to $W_P = -\beta + (1-\alpha)\pi$ and the agent's wealth is equal to $W_A = \beta + \alpha\pi$, where β is a base salary and α the agent's share of profits. The principal is risk neutral while the agent is risk averse. His utility function displays constant absolute risk aversion. More precisely, we have

$$U_A(W_A, e) = -\exp(-a(W_A - ce)),$$

where a is the coefficient of absolute risk aversion. As is well known, the agent's expected utility can be expressed as $\mathbb{E}[U_A(W_A, e)] = U_A(w_A - ce)$ where

$$w_A \equiv \beta + \alpha\mu(e, \sigma) - a\frac{\alpha^2}{2}\sigma^2 \tag{1}$$

[6] For most of the paper, $\overline{\sigma}$ can be taken as ∞. Since the principal is risk neutral, we need a finite $\overline{\sigma}$ to make the first-best allocation well defined.

[7] Making costs linear in effort is a normalization that is without loss of generality.

is the agent's certainty equivalent level of wealth. Clearly, for any given effort choice, the agent will always choose a point on the efficient frontier, so $\mu = \mu(e, \sigma)$. The principal's expected utility is equal to his expected wealth, so

$$\mathbb{E}[W_P] = -\beta + (1 - \alpha)\,\mu(e, \sigma).$$

The agent's outside option gives rise to a certainty equivalent level of wealth of ω. The agent knows his marginal cost of effort and his coefficient of risk aversion. These parameters are distributed with full support on the product set $\mathbf{T} \equiv [\underline{a}, \overline{a}] \times [\underline{c}, \overline{c}]$ where $\overline{a} > \underline{a} > 0$ and $\overline{c} > \underline{c} > 0$. We let $\mathbf{t} \equiv (a, c)$ denote a type and let $k(\mathbf{t})$ and $K(\mathbf{t})$ denote the joint density and cdf of $\mathbf{t}$, respectively.

Apart from the efficient frontier - our key new element - these assumptions are standard in the literature (see, e.g. Holmström and Milgrom (1994)). We explore two variations of our model; in the first version, the agent's type is commonly known so that the only contractual friction is moral hazard arising from the unobservability of the agent's choices; in the second version, the principal only knows the distribution of the agent's type (and this is common knowledge), so there is adverse selection on top of moral hazard.

3 The Principal's Problem

We state the principal's problem for the most general case, where the agent has private information about his level of risk aversion and cost of effort. The case of symmetric information is then a special case of the general formulation.

Invoking the Revelation Principle, an optimal contract can be found restricting attention to a direct revelation game, where the agent is asked to announce his preference parameters $\hat{\mathbf{t}}$, and is given incentives to announce his type truthfully. For any given announced type, $\hat{\mathbf{t}} \in \mathbf{T}$, a contract specifies the quadruple $\left\{\beta(\hat{\mathbf{t}}), \alpha(\hat{\mathbf{t}}), e(\hat{\mathbf{t}}), \sigma(\hat{\mathbf{t}})\right\}$. Our problem is a combined problem of moral hazard and adverse selection. However, once the agent has announced a type, $\beta(\hat{\mathbf{t}})$ and $\alpha(\hat{\mathbf{t}})$ are given from his perspective. So, we can use (1) to compute the optimal choices of effort and standard deviation (from his perspective); let $e(\alpha, \mathbf{t})$ and $\sigma(\alpha, \mathbf{t})$ denote these choices. Since $\mu(e, \sigma)$ is jointly concave in its arguments, incentive compatible choices are completely described by the pair of first-order conditions

$$\alpha(\hat{\mathbf{t}})\,\mu_e\left(e\left(\alpha(\hat{\mathbf{t}}), \mathbf{t}\right), \sigma\left(\alpha(\hat{\mathbf{t}}), \mathbf{t}\right)\right) = c \tag{2}$$

and either

$$\sigma\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right)\leq\overline{\sigma}\ \mathit{and}\ \mu_{\sigma}\left(e\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right),\sigma\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right)\right)-a\alpha\left(\hat{\mathbf{t}}\right)\sigma\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right)=0, \tag{3}$$

or

$$\sigma\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right)=\overline{\sigma}\ \mathit{and}\ \mu_{\sigma}\left(e\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right),\sigma\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right)\right)-a\alpha\left(\hat{\mathbf{t}}\right)\sigma\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right)\geq 0. \tag{4}$$

Given strict concavity of the function $\mu(\cdot,\cdot)$ in its arguments, this system of equations has a unique solution. Taking these choices into account, the principal's problem is reduced to a problem of pure adverse selection. The principal's problem is to

$$\max_{\beta(\cdot),\alpha(\cdot),\tilde{\mathbf{T}}(\omega)}\int_{\tilde{\mathbf{T}}(\omega)}\left(-\beta(\mathbf{t})+(1-\alpha(\mathbf{t}))\,\mu\left(e\left(\alpha(\mathbf{t}),\mathbf{t}\right),\sigma\left(\alpha(\mathbf{t}),\mathbf{t}\right)\right)\right)k(\mathbf{t})\,d\mathbf{t} \tag{5}$$

$$s.t.$$

$$\begin{aligned} & w_A\left(\beta(\mathbf{t}),\alpha(\mathbf{t}),e\left(\alpha(\mathbf{t}),\mathbf{t}\right),\sigma\left(\alpha(\mathbf{t}),\mathbf{t}\right)\right)-ce\left(\alpha(\mathbf{t}),\mathbf{t}\right) \\ \geq\ & w_A\left(\beta\left(\hat{\mathbf{t}}\right),\alpha\left(\hat{\mathbf{t}}\right),e\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right),\sigma\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right)\right)-ce\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right)\ \ \mathit{for\ all}\ \mathbf{t},\hat{\mathbf{t}}\in\tilde{\mathbf{T}}(\omega) \end{aligned} \tag{6}$$

$$w_A\left(\beta(\mathbf{t}),\alpha(\mathbf{t}),e\left(\alpha(\mathbf{t}),\mathbf{t}\right),\sigma\left(\alpha(\mathbf{t}),\mathbf{t}\right)\right)-ce\left(\alpha(\mathbf{t}),\mathbf{t}\right)\geq\omega\ \mathit{for\ all}\ \mathbf{t}\in\tilde{\mathbf{T}}(\omega) \tag{7}$$

$$\max_{\hat{\mathbf{t}}\in\tilde{\mathbf{T}}(\omega)} w_A\left(\beta\left(\hat{\mathbf{t}}\right),\alpha\left(\hat{\mathbf{t}}\right),e\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right),\sigma\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right)\right)-ce\left(\alpha\left(\hat{\mathbf{t}}\right),\mathbf{t}\right)\leq\omega\ \mathit{for\ all}\ \mathbf{t}\in\mathbf{T}\setminus\tilde{\mathbf{T}}(\omega) \tag{8}$$

In this problem, constraint (6) is the incentive constraint that guarantees truthtelling. (7) ensures that agents with characteristics $\mathbf{t}\in\tilde{\mathbf{T}}(\omega)$ are willing to participate, (8) ensures that agents with other characteristics do not participate; the principal chooses the set $\tilde{\mathbf{T}}(\omega)$, i.e., whom to attract and whom to exclude. Note again that the moral hazard part of our problem has been subsumed into the hidden information part of the problem by requiring that $e(\alpha(\mathbf{t}),\mathbf{t})$ and $\sigma(\alpha(\mathbf{t}),\mathbf{t})$ satisfy the conditions (2) and either (3) or (4).[8] Thus, the problem of pure moral hazard corresponds to the problem above when we drop constraint (6). Moreover, in this case, the problem can always be solved pointwise for each $\mathbf{t}$.

The choice of the set $\tilde{\mathbf{T}}(\omega)$ is only interesting in case the characteristics are privately known to the agent; this is due to the absence of wealth effects in the principal's and the agent's utility function. For this reason we study the problem with known characteristics under conditions that ensure that full participation is optimal; formally, we set $\omega=0$ for the first part of the analysis in section 4, which ensures that $\tilde{\mathbf{T}}(\omega)=\mathbf{T}$. In contrast, the optimal allocation in the case of privately known characteristics, analyzed in section 5, features exclusion of a portion of types - who are

[8] With a slight abuse of notation, $e(\mathbf{t})=e(\alpha(\mathbf{t}),\mathbf{t})$ and $\sigma(\mathbf{t})=\sigma(\alpha(\mathbf{t}),\mathbf{t})$ correspond to the "recommended" choices introduced in the definition of contracts above.

particularly risk averse and have very high costs of effort - with strictly positive measure whenever $\omega > 0$.

Before we dive into the analysis of the incentive problems, we shall briefly discuss the case where the principal can observe the agent's type and his choices. If the principal is perfectly informed about the agent's preferences and choices, then there is no need to use the share α to control incentives. Hence, α is set so as to induce an optimal allocation of risk between agent and principal, so $\alpha^* (\mathbf{t}) = 0$ for all $\mathbf{t}$. Since the principal is indifferent towards risk, and the efficient frontier is increasing in σ, he will prefer for any given effort the maximum volatility, so $\sigma^* (\mathbf{t}) = \overline{\sigma}$ for all $\mathbf{t}$. Finally, the optimal level of effort satisfies the first-order condition $\mu_e (e (\alpha^* (\mathbf{t}), \mathbf{t}), \overline{\sigma}) = c$. Notice that both α^* and σ^* are independent of the agent's preference parameters. The optimal level of the mean is decreasing in c, as effort is decreasing in c under complete information.

4 The Problem of Pure Moral Hazard

Even though we can characterize the solution to our model - in the case of commonly known characteristics - for general functions $\mu(e, \sigma)$, clear-cut comparative statics predictions require quite a lot more structure. Thus, to make progress we assume from now on that

$$\mu (e, \sigma) = e^{\lambda} \sigma^{\delta}. \tag{9}$$

It is easy to verify that the agent's problem of choosing e and σ for given contract is jointly concave in the choice variables if $0 \leq \lambda, \delta \leq 1$ and $\lambda + \frac{\delta}{2} \leq 1$,[9] so we impose these restrictions to make the agent's problem well behaved. As we discuss shortly, to make the principal's problem well behaved (that is concave in α), we assume on top of this that $\lambda \leq .5$ and $\delta \leq 2\lambda$.

To solve our problem in the most reader friendly way, we proceed as follows. We demonstrate the important features of the solution for interior volatility choices in the main text. We provide the details of the solution in the appendix along side with a discussion for which parameter values the solution is indeed interior.

It is useful to ease notation defining some statistics of the model parameters. The details are not interesting in any way but are provided for completeness in Definition 1 in the appendix. Let $\eta \equiv \eta(\delta, \lambda)$, $\Delta \equiv \Delta(\delta, \lambda)$, and $\Gamma \equiv \Gamma(\delta, \lambda)$ denote functions of the parameters only, and

[9] The factor .5 stems from the fact that the cost of risk bearing is quadratic in σ. Switching variables from standard deviation to variance in (9) gives rise to the standard restriction that the sum of exponents be smaller than unity.

let $\theta(\mathbf{t}) \equiv a^{\frac{-\delta}{2(1-\lambda)-\delta}} \cdot c^{\frac{-2\lambda}{2(1-\lambda)-\delta}}$. Building on this notation, we can write the mean of the return distribution induced by an agent with characteristics $\mathbf{t}$ as

$$\mu(e(\alpha,\mathbf{t}),\sigma(\alpha,\mathbf{t})) = \Delta\theta(\mathbf{t})\,\alpha^{\eta-1}; \tag{10}$$

the agent's cost of risk bearing and effort as

$$a\frac{\alpha^2}{2}\sigma(\alpha,\mathbf{t})^2 + ce(\alpha,\mathbf{t}) = \Gamma\theta(\mathbf{t})\,\alpha^{\eta};$$

and the agent's indirect certainty equivalent level of wealth as

$$w_A(\beta,\alpha,e(\alpha,\mathbf{t}),\sigma(\alpha,\mathbf{t}),a) - ce(\alpha,\mathbf{t}) = \beta + \alpha\Delta\theta(\mathbf{t})\,\alpha^{\eta-1} - \Gamma\theta(\mathbf{t})\,\alpha^{\eta}. \tag{11}$$

4.1 Optimal Contracts

It is now straightforward to solve the principal's problem. Clearly, when the agent's characteristics are known, the participation constraint has to be binding for each $\mathbf{t}$. Notice, that the indirect certainty equivalent level of wealth - when the agent chooses effort and volatility optimally from his perspective - depends only on $\theta = \theta(\mathbf{t})$, a unidimensional statistic of $\mathbf{t}$. This simplifies the model dramatically. Imposing (7) for each θ and substituting into the principal's objective, we obtain the following unconstrained problem

$$\max_{\alpha} \theta(\mathbf{t})\left(\Delta\alpha^{\eta-1} - \Gamma\alpha^{\eta}\right). \tag{12}$$

It is easy to verify that problem (12) is increasing in α for $\alpha = 0$ and concave in α for $\eta \in (1,2)$, or equivalently for $\lambda \leq .5$ and $\delta \leq 2\lambda$, which is precisely the reason we impose this restriction. In this case, we can characterize the solution by the pointwise first-order conditions

$$\alpha^* = \underline{\alpha} \equiv \frac{\eta-1}{\eta}\frac{\Delta}{\Gamma}. \tag{13}$$

The optimal share α^* is independent of the agent's preference parameters as long as the implied volatility choice is interior. This is due to the Cobb-Douglas technology. Since the agent's volatility choice is the higher the less risk averse the agent is and the smaller his marginal cost of effort is, the volatility choice is indeed interior for relatively high values of the agent's preference parameters. Let $\mathbf{T}_I$ denote the (closed) set of parameters giving rise to an interior solution. Depending on the support of the agent's preference parameters, there is necessarily a set of types for which the upper bound on volatility is a binding constraint. Let $\mathbf{T}_C = \mathbf{T} \setminus \mathbf{T}_I$ denote this (open) set. $\mathbf{T}_C$ is nonempty if some agents are close to risk neutral and/or have very low cost of effort. To capture

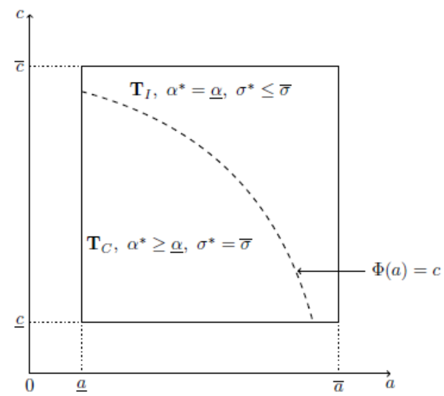

Figure 2: Interior versus corner solutions.

this case, we assume that the lower bounds $\underline{a}$ and $\underline{c}$ are sufficiently low. In this case, there exists a strictly decreasing function $\Phi(a)$ that separates the sets $\mathbf{T}_C$ and $\mathbf{T}_I$, depicted in figure 2 below.

We can now characterize the overall solution to the contracting problem.

Proposition 1 *i) There exists a strictly decreasing function* $\Phi(a)$, *such that* $\mathbf{T}_C \equiv \{\mathbf{t} : c < \Phi(a)\}$. *For* $\underline{a}$ *and* $\underline{c}$ *sufficiently small,* $\mathbf{T}_C$ *is nonempty.*

ii) For relatively risk averse agents with a relatively high cost of effort (formally, for $\mathbf{t} \in \mathbf{T}_I$*), the optimal share* $\alpha^*(\mathbf{t})$ *is independent of* $\mathbf{t}$ *and given by* (13). *The optimal choice of volatility and the expected level of profits are both decreasing in* $\mathbf{t}$.

iii) For relatively risk tolerant agents with a relatively low cost of effort (formally, for $\mathbf{t} \in \mathbf{T}_C$*), the optimal share* $\alpha^*(\mathbf{t})$ *is decreasing in* $\mathbf{t}$. *The optimal choice of volatility is* $\sigma^*(\mathbf{t}) = \overline{\sigma}$. *The expectation of profits is decreasing in* $\mathbf{t}$.

iv) For any $\mathbf{t}, \mathbf{t}'$ *such* $\mathbf{t} \in \mathbf{T}_C$ *and* $\mathbf{t}' \in \mathbf{T}_I$, *we have* $\alpha^*(\mathbf{t}) > \alpha^*(\mathbf{t}')$.

The economics is straightforward. An efficient allocation of risks would require that α be set equal to zero for all $\mathbf{t}$. While a riskless contract would induce the agent to choose the optimal volatility from the principal's perspective, that is $\sigma = \overline{\sigma}$, it would not give the agent any incentive to exert effort. Hence, α is set too high relative to the first-best. As a result, there is a strictly positive cost of risk bearing which is increasing in a. Since the agent's participation constraint always holds as an equality, it is the principal who bears the cost of this inefficiency. The higher is a, the more costly it becomes to convince the agent to participate for a given share of profits α. Hence, the principal weakly reduces α as a is increased. The agent, on the other hand, can reduce the cost of risk bearing by changing the volatility of the project. Hence, the agent (weakly) reduces the volatility of the project as he becomes more risk averse.

Similarly, when c increases, any given level of effort becomes more costly to implement. Hence, the principal finds it optimal to reduce incentives for effort when c increases, so α is reduced. Since volatility and effort are complements along the efficient frontier, the agent has less of an incentive to engage in risk taking. Hence, the optimal volatility is reduced as well.

4.2 Covariance of Contracts and Moments of the Profit Distribution

Inspired by Holmström and Milgrom (1994), we build our comparative statics predictions on the concept of associated random variables. Recall from Esary, Barlow, and Walkup (1967)) that random variables $\mathbf{t}$ are associated if

$$COV\left(x\left(\mathbf{t}\right), y\left(\mathbf{t}\right)\right) = \mathbb{E}\left[x\left(\mathbf{t}\right) y\left(\mathbf{t}\right)\right] - \mathbb{E}\left[x\left(\mathbf{t}\right)\right] \mathbb{E}\left[y\left(\mathbf{t}\right)\right] \geq 0$$

for all non-decreasing functions $x\left(\mathbf{t}\right)$ and $y\left(\mathbf{t}\right)$ (that is, functions that are non-decreasing in each of the arguments) for which $\mathbb{E}\left[x\left(\mathbf{t}\right) y\left(\mathbf{t}\right)\right], \mathbb{E}\left[x\left(\mathbf{t}\right)\right]$, and $\mathbb{E}\left[y\left(\mathbf{t}\right)\right]$ exist.[10] Notice that the functions $\alpha^*\left(\mathbf{t}\right)$, $\mu^*\left(\mathbf{t}\right)$, and $\sigma^*\left(\mathbf{t}\right)$ described in the proposition above are comonotone. Before $\mathbf{t}$ is realized, the values these functions take are random. Let $\tilde{\alpha}^*$, $\tilde{\sigma}^*$, and $\tilde{\mu}^*$ directly denote these random variables.

Proposition 2 *i) The covariance of $\tilde{\alpha}^*$ and $\tilde{\sigma}^*$ is strictly positive.*
ii) If $\mathbf{t}$ is associated, then the covariance of $\tilde{\alpha}^*$ and $\tilde{\mu}^*$ and the covariance of $\tilde{\mu}^*$ and $\tilde{\sigma}^*$ are nonnegative.
iii) If either managerial risk aversion or his/her cost of effort can be controlled for, then the covariance of $\tilde{\alpha}^*$ and $\tilde{\mu}^*$ and the covariance of $\tilde{\mu}^*$ and $\tilde{\sigma}^*$ are both strictly positive.

Part i) follows from the fact that the functions $\alpha^*\left(\mathbf{t}\right)$ and $\sigma^*\left(\mathbf{t}\right)$ are comonotone and moreover that one function is strictly decreasing exactly in the region where the other is constant. Calculating the covariance by separating these regions yields the strictly positive result. Part ii) follows directly from the association property, because $\alpha^*\left(\mathbf{t}\right)$, $\mu^*\left(\mathbf{t}\right)$, and $\sigma^*\left(\mathbf{t}\right)$ are all monotonic in $\mathbf{t}$. Finally, part iii) follows from the association property, the fact that one random variable is always associated, and that one can rewrite $\tilde{\alpha}^*$ and $\tilde{\sigma}^*$ as increasing functions of $\tilde{\mu}^*$.

The predictions of the pure moral hazard model when all moments of the profit distribution are endogenous are remarkably unambiguous: provided that the parameters in the agent's payoff

[10] For the relationship between association and other concepts of dependence, see Esary and Proschan (1972).

function are positively correlated in the sense of association, the solution to the incentive problem and the induced moments reflect this positive correlation. This is in remarkably stark contrast to the predictions of the exact same model when the variance is taken as exogenous.[11] Intuitively, effort and risk are complements in the agent's problem, so both choices tend to increase the stronger are the incentives the agent faces. On the other hand, the principal offers steeper incentives to agents that are easier to incentivize.

5 The Case of Combined Adverse Selection and Moral Hazard

We now analyze the full problem, where the agent has private information about his preference parameters. In this case we obtain a rich set of comparative statics predictions also for the case where the feasibility constraint on the volatility is never binding. For convenience, we focus on this case.

Building on the analysis of the pure moral hazard case, we know that the agent's certainty equivalent level of wealth depends on the underlying parameters only through the statistic θ. Therefore, it is clear that there must necessarily be bunching of types $\mathbf{t}$ with the same level of θ. From, (11) the agent's certainty equivalent level of wealth for any given announced type $\hat{\theta}$ is

$$\beta\left(\hat{\theta}\right) + (\Delta - \Gamma)\,\theta\alpha\left(\hat{\theta}\right)^{\eta}.$$

Note that the cross derivative of this expression with respect to α and θ is positive, so the single crossing condition holds.[12] Moreover, the agent's indirect utility is the higher the higher is θ.

A crucial difference between the present problem and the pure moral hazard problem is that the level of the agent's outside option matters quite a bit; the solution for the case where the agent's outside option, ω, satisfies $\omega > 0$ is qualitatively different from the case where $\omega = 0$, because the principal finds it optimal to exclude some types. Since types with higher θ derive higher utility from participating, the principal excludes types with a low level of θ.

[11] The standard trade-off between risk and incentives arises in the parameter set that gives rise to a corner solution, $\mathbf{T}_C$, when the upper bound on volatility, $\overline{\sigma}$, increases.

[12] See Araujo et al. (2007) for an analysis of the case where the single crossing condition fails to hold. See Biais et al. (2000) for a multidimensional model allowing for a similar reduction of the dimension of the incentive problem.

Building on these insights, we can write the principal's problem formally as follows:

$$\max_{\alpha(\cdot),\beta(\cdot),\theta_m} \int_{\theta_m}^{\overline{\theta}} \left\{ -\beta(\theta) + (1-\alpha(\theta))\,\theta\Delta\alpha(\theta)^{\eta-1} \right\} dF(\theta)$$

$$s.t$$

$$\beta(\theta) + (\Delta-\Gamma)\,\theta\alpha(\theta)^{\eta} \geq \beta\left(\hat{\theta}\right) + (\Delta-\Gamma)\,\theta\alpha\left(\hat{\theta}\right)^{\eta} \;\; for\ all\ \theta, \hat{\theta} \geq \theta_m$$

$$\beta(\theta) + (\Delta-\Gamma)\,\theta\alpha(\theta)^{\eta} \geq \omega \;\; for\ all\ \theta \geq \theta_m$$

$$\max_{\hat{\theta}} \beta\left(\hat{\theta}\right) + (\Delta-\Gamma)\,\theta\alpha\left(\hat{\theta}\right)^{\eta} \leq \omega \;\; for\ all\ \theta < \theta_m.$$

where $F(\theta)$ is the cdf of the distribution of θ and θ_m is the marginal type θ that is included; all types $\theta < \theta_m$ are excluded.

The first step to solve this problem is to bring the incentive and participation constraint into a more tractable form. We call a pair of schedules implementable if they satisfy these two conditions.

Lemma 1 *A pair of schedules $\alpha(\theta)$ and $\beta(\theta)$ is implementable if and only if*

$$\beta(\theta) = \int_{\theta_m}^{\theta} (\Delta-\Gamma)\,\alpha(z)^{\eta}\,dz - (\Delta-\Gamma)\,\theta\alpha(\theta)^{\eta} \;\; for\ all\ \theta \geq \theta_m \tag{14}$$

and $\alpha(\theta)$ is nondecreasing in θ. Exclusion of types $\theta < \theta_m$ is incentive compatible if $\alpha(\theta) = \beta(\theta) = 0$ for $\theta < \theta_m$.

The proof of the lemma is standard and therefore only sketched in the appendix. Only monotonic schedules $\alpha(\cdot)$ can satisfy the incentive constraint. As monotonic schedules are differentiable almost everywhere, the agent's indirect utility function is differentiable almost everywhere. By the envelope theorem, the agent's utility changes with his preference statistic θ at rate $(\Delta-\Gamma)\,\alpha(\theta)^{\eta} \geq 0$. Imposing the participation constraint for the marginal type θ_m and integrating the changes in utility, we get (14). Finally, one shows that when contracts satisfy monotonicity of the schedule $\alpha(\theta)$ and (14), then there is no profitable deviation for the agent. In particular, this argument implies also that deviations for a type $\theta < \theta_m$ to any type $\hat{\theta} \geq \theta_m$ would yield a level of utility that is strictly smaller than what the agent can get elsewhere, ω.

Recall that $\theta = \theta(\mathbf{t}) = a^{\frac{-\delta}{2(1-\lambda)-\delta}} \cdot c^{\frac{-2\lambda}{2(1-\lambda)-\delta}}$ is a statistic of the underlying parameters. Since $\theta(\mathbf{t})$ is a function of random variables, we need to derive its distribution from the underlying distributions. The following lemma gathers the important features. For convenience, define $r \equiv a^{\frac{-\delta}{2(1-\lambda)-\delta}}$ and $s \equiv c^{\frac{-2\lambda}{2(1-\lambda)-\delta}}$, respectively.

Lemma 2 *The distribution of θ is supported on a set $\left[\underline{\theta}, \overline{\theta}\right]$, where $\underline{\theta} \equiv \underline{rs}$ and $\overline{\theta} \equiv \overline{rs}$. Moreover, let $F(\theta)$ denote the cdf of θ and $f(\theta)$ denote the pdf. The density satisfies $f(\underline{\theta}) = 0$ and $f(\theta) > 0$ for $\theta > \underline{\theta}$. Moreover, provided that $g(s|r)$, the conditional density of s given r satisfies $\frac{\partial}{\partial s}(sg(s|r)) \geq 0$, the distribution of θ satisfies $\frac{\partial}{\partial \theta}\frac{1-F(\theta)}{\theta f(\theta)} \leq 0$ for $\theta > \underline{\theta}$.*

Note that the density of θ goes to zero as θ approaches the lower end of the type support. This is a well known property of this sort of problem and the driving force behind the exclusion result that we establish below, replicating Armstrong's (1996) observation for multidimensional screening problems more generally. As we discuss below in greater detail, the extent of exclusion in this particular contexts is simply a question of the level of the agent's outside option.

5.1 Optimal Contracts

It proves convenient to solve the principal's problem in two steps. In the first step, we take the choice of θ_m as given and solve for optimal contracts for given θ_m. In the second step, we address the exclusion problem. Types that are induced to opt out are offered a contract $\alpha(\theta) = \beta(\theta) = 0$. Substituting for $\beta(\theta)$ from (14) into the principal's objective function and integrating by parts, we have

$$V(\theta_m) = \max_{\alpha(\cdot)} \int_{\theta_m}^{\overline{\theta}} \left\{ \left(\theta \Delta \alpha(\theta)^{\eta-1} - \Gamma \theta \alpha(\theta)^{\eta} \right) f(\theta) - (\Delta - \Gamma) \alpha(\theta)^{\eta} (1 - F(\theta)) \right\} d\theta - \omega (1 - F(\theta_m))$$

$$s.t.\ \alpha(\theta) \text{ nondecreasing in } \theta.$$

The single crossing condition ensures that the participation constraint only binds at the low end of the support, in particular at θ_m. For a given choice of θ_m, the principal faces the standard efficiency versus rent extraction trade-off. On the one hand, the principal wishes to raise α for each type so as to improve upon incentives for effort. On the other hand, the higher is α, the higher are the rents the principal needs to give up to agents with a relatively high value of θ. The optimal schedule $\alpha(\theta)$ strikes a balance between these two motives. Under an appropriate regularity condition, the solution can be found by point-wise maximization under the integral. We state these results in the following proposition:

Proposition 3 *Suppose that, for $\theta > \underline{\theta}$, $\frac{1-F(\theta)}{\theta f(\theta)}$ is non-increasing in θ. Then, the optimal share schedule for $\theta \geq \theta_m$ is given by*

$$\alpha(\theta) = \frac{\frac{\eta-1}{\eta}\Delta}{\Gamma + (\Delta - \Gamma)\frac{1-F(\theta)}{\theta f(\theta)}}.$$

The optimal associated schedule $\beta(\theta)$ *is given by* (14). *In the limit as* $\theta_m \to \underline{\theta}$, *we have* $\lim_{\theta_m \to \underline{\theta}} \alpha(\theta_m) = 0$.

We omit a formal proof; the result follows straightforwardly from pointwise maximization. Moreover, it is easy to verify that the regularity condition implies that the solution is monotonic in θ, so that we can indeed use pointwise maximization techniques.

The solution has the classical features. There is no distortion due to adverse selection for the agent with the highest parameter θ; that is, $\alpha(\overline{\theta}) = \frac{\eta-1}{\eta}\frac{\Delta}{\Gamma}$, corresponding exactly to the solution under pure moral hazard. For all $\theta < \overline{\theta}$, the share schedule is distorted downwards so as to extract rents from the agents with high parameters θ. There is no rent at the bottom. Moreover, since the density of types θ goes to zero at the low bound of the support, if such agents are offered a contract, then the shares they are offered become very small and go to zero as $\theta_m \to \underline{\theta}$. The reason is well understood from Armstrong (1996) and Rochet and Choné (1998). The density measures the weight given to the (constrained) efficiency motive in the principal's objective; on the other hand, $1 - F(\theta)$ measures the weight given to the rent-extraction motive. Hence, at the low end of the support, the rent extraction motive becomes infinitely more important than the efficiency motive.

Consider now the optimal choice of types to include or to exclude, respectively. Using the first-order condition for the optimal $\alpha(\theta)$, the derivative of the principal's payoff with respect to θ_m is

$$V'(\theta_m) = \left(\omega - \frac{\mu(\theta_m)}{\eta}\right) f(\theta_m),$$

where, with a slight abuse of notation, $\mu(\theta)$ is short for the induced mean according to (10). The following results are now obvious:

Proposition 4 *It is optimal to exclude a set of types with positive measure if and only if* $\omega > 0$. *The marginal type* θ_m^* *is uniquely defined by the condition*

$$\omega\eta = \mu(\theta_m^*),$$

where θ_m^* *is the higher the higher is* ω. *Moreover, the higher is* ω, *the higher is the lowest incentive share that is offered,* $\alpha^*(\theta_m^*)$, *and the higher is* $\mathbb{E}[\alpha^*(\theta) | \theta \geq \theta_m^*]$, *the "average" observed incentive power of agents that are hired.*

Since $\alpha^*(\theta)$ goes to zero as θ approaches the low end of the support, the expected profit generated by an agent of given type θ goes to zero. Moreover, higher θ types generate higher expected profits. Consequently, there is a uniquely defined marginal type θ_m^* who generates exactly zero net surplus to the principal. Since only monotonic incentive schemes are incentive compatible, the positive implications of exclusion are as follows. The larger is the set of excluded agents, that is the higher is θ_m, the higher is the minimum level of α that is observed in the cross section; in particular, the minimum exposure to risk is bounded away from zero so as to exclude some agents.[13]

5.2 Covariance of Contracts and Moments

We now turn to the comparative statics properties of the optimal contracting arrangement.

Proposition 5 *i) With combined moral hazard and adverse selection, the covariance of $\tilde{\alpha}^*$ and $\tilde{\mu}^*$ is strictly positive whenever $\alpha^*(\theta)$ is increasing in θ on a set of positive measure.*
ii) The covariance of $\tilde{\alpha}^*$ and $\tilde{\sigma}^*$ and of $\tilde{\mu}^*$ and $\tilde{\sigma}^*$, respectively, is in general ambiguous. The covariance of $\tilde{\alpha}^*$ and $\tilde{\sigma}^*$ and of $\tilde{\mu}^*$ and $\tilde{\sigma}^*$, respectively, is strictly positive if $\mathbf{t}$ is associated and the distribution of θ satisfies $-\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{f(\theta)} \leq \frac{\delta+2\lambda}{2(1-\lambda)-\delta}$ for $\theta \geq \tilde{\theta}$.

Part i) of the proposition is due to the fact that $\tilde{\alpha}^*$ and $\tilde{\mu}^*$ are nondecreasing functions in θ. Given θ is unidimensional one can rewrite $\tilde{\alpha}^*$ as a nondecreasing function of $\tilde{\mu}^*$. Since a scalar random variable is always associated, the result follows directly if the optimal α is strictly monotonic on a set of positive measure. Part ii) states that, in general, the model loses its predictive power when it comes to the covariance of $\tilde{\alpha}^*$ and $\tilde{\sigma}^*$. However, one can give simple sufficient conditions for a positive correlation between risk and incentives. The one given in the proposition ensures that the optimal profit share $\alpha^*(\theta)$ does not change too fast as θ changes, ensuring that the agent's optimal choice of σ becomes monotonic in the agent's underlying preference parameters. Since the defining property of associated random variables is precisely that the covariance of any monotonic functions of these random variables is positive, the conclusion follows immediately.

6 Attenuation

In our model, the moments of the profit distribution and the optimal contracts are endogenously determined as functions of the agent's preference parameters $\mathbf{t} = (a, c)$, that is, his degree of absolute risk aversion, a, and his marginal cost of effort, c. While we do not test our model directly,

[13] This should not be taken as a justification for high levels of manager compensation. The level is determined to a large extent by ω, which is exogenous in the present model.

we now discuss how to bring it to the data and the consequences of neglecting the endogeneity of the variance.

If we are merely interested in contracts and risk, then a reduced form of our model is a system of equations for α and σ (both normalized around their means) of the sort

$$\sigma = \gamma_1 a + \gamma_2 c + \varepsilon_1 \tag{15}$$

and

$$\alpha = \delta_1 a + \delta_2 c + \varepsilon_2, \tag{16}$$

where ε_1 and ε_2 are independent of each other and in particular independent of a and c.

What if the endogeneity of σ is neglected and instead σ is treated as a regressor for α? If there is some β_1 such that $\frac{\delta_1}{\gamma_1} = \frac{\delta_2}{\gamma_2} = \beta_1$, then we can find another linear relation between α and σ of the form

$$\alpha = \beta_1 \sigma + \varepsilon_3. \tag{17}$$

However, by implication of (15) and (16), ε_3 is related to ε_1 and ε_2 according to

$$\varepsilon_3 = \varepsilon_2 - \beta_1 \varepsilon_1. \tag{18}$$

Using $\mathbb{E}[\varepsilon_3] = 0$ and conditions (15) and (18), we find that

$$Cov(\sigma, \varepsilon_3) = \mathbb{E}[\varepsilon_3 (\sigma - \mathbb{E}\sigma)] = \mathbb{E}[(\varepsilon_2 - \beta_1 \varepsilon_1)(\varepsilon_1)] = -\beta_1 Var(\varepsilon_1),$$

so that $Cov(\sigma, \varepsilon_3) < (>) \, 0$ iff $\beta_1 > (<) \, 0$. Therefore, specification (17) fails to satisfy the assumptions of the linear regression model. As a consequence, the estimated value of β_1 is biased towards zero, an effect that is known as attenuation (see e.g., Greene (1993); neglecting, the endogeneity of σ biases the estimate of β_1 towards zero. (The estimate is also inconsistent.)

The effects are slightly different but not more reassuring if σ is treated as a regressor alongside with controls a and c. Suppose we specify a linear model of the form

$$\alpha = \kappa_1 \sigma + \kappa_2 a + \kappa_3 c + \varepsilon_4 \tag{19}$$

in a situation where the true model is the system of equations (15) and (16). In fact, a form like (19) is obtained if we start from (16) and add κ_1 times the difference between the left and right side of (15). We obtain the following relation

$$\alpha = \kappa_1 \sigma + (\delta_1 - \kappa_1 \gamma_1) a + (\delta_2 - \kappa_2 \gamma_2) c + \varepsilon_2 - \kappa_1 \varepsilon_1,$$

which is indeed of the same form as (19) with $\varepsilon_4 = \varepsilon_2 - \kappa_1\varepsilon_1$. Exactly the same attenuation problem as above arises if $\delta_1 = \kappa_1\gamma_1$ and $\delta_2 = \kappa_2\gamma_2$. If $\delta_1 \neq \kappa_1\gamma_1$ or $\delta_2 \neq \kappa_2\gamma_2$, then the direction of the bias is no longer clear, but the estimate of κ_1 remains biased.

Summing up, neglecting the endogeneity of risk in simple regressions of contracts on measures of risk biases the estimates towards zero, irrespective of whether the estimated relation between the slope of incentive contracts and risk is positive or negative. In more sophisticated regressions that include risk as a regressor alongside with controls that effectively determine both the left- and the right-hand side of the regression equation, the direction of the bias is less clear; however, the estimation clearly remains biased also in these cases.

Whether, risk is exogenous or endogenous clearly depends on the context, so we cannot settle the question in a theoretical model. However, we simply point out, that attenuation makes it more likely to reject the hypothesis that incentives (α) depend positively (or negatively) on risk (σ).

7 Conclusions

In this paper, we analyze a model of managerial compensation with endogenous risk. Contracts serve a double purpose as providers of effort incentives and to guide the manager's project choices along an efficient frontier. The model offers a rich set of insights that have not been explored in such detail before. The resulting connection between risk and incentives depends on the underlying incentive problem. With pure moral hazard, a positive relation arises very naturally under general assumptions. With combined moral hazard and adverse selection, it is easy to find examples where the correlation between risk and incentives remains positive, but one can also construct cases where the covariance between risk and incentives is negative. However, we do not so much argue for a particular sign of this relation. The main point of the exercise is more that risk may be endogenous and to explore the implications of this variation. Empirically, endogeneity of risk gives rise to an attenuation problem resulting in estimates that are biased towards zero. We believe this may explain why a good part of the empirical studies on the subject produce relatively small (often statistically not significant) relations between risk and incentives. We leave taking our model to the data directly to future work.

We have analyzed a particular incentive problem in this paper where risk averse agents interact with risk neutral principals. As a consequence, our model cannot address excessive risk taking behavior. An incentive problem of this sort would arise, e.g., if both managers and principals are risk neutral and the manager gets some form of convex pay-scheme (e.g. through the use of options); similarly, excessive risk taking arises if managers and principals are risk neutral and firms suffer costs of financial distress - making the principal's payoff effectively concave in profits. Interestingly, even though the incentive problem differs substantially, the models may share the same comparative statics predictions that risk and incentives are positively related.

8 Appendix

To ease notation, we introduce the following variables, that are functions of the underlying parameters.

Definition 1

$$\begin{aligned} \eta &\equiv \frac{2(1-\delta)}{2(1-\lambda)-\delta}; \Delta \equiv \lambda^{\frac{2\lambda}{2(1-\lambda)-\delta}} \delta^{\frac{\delta}{2(1-\lambda)-\delta}}; \\ \Gamma &\equiv \left(\frac{1}{2}+\frac{\lambda}{\delta}\right)\left(\delta^{1-\lambda}\lambda^{\lambda}\right)^{\frac{2}{2(1-\lambda)-\delta}}; \Lambda \equiv \left(\delta^{1-\lambda}\lambda^{\lambda}\right)^{\frac{1}{2(1-\lambda)-\delta}}. \end{aligned}$$

Moreover,

$$\theta \equiv r \cdot s,$$

where

$$r \equiv a^{\frac{-\delta}{2(1-\lambda)-\delta}} \text{ and } s \equiv c^{\frac{-2\lambda}{2(1-\lambda)-\delta}}.$$

Proof of Proposition 1. We first establish part ii, then parts i), iii) and iv).

ii) It is straightforward to compute the interior solution from the first-order conditions. For a given share α, the agent's effort and volatility choice are given by

$$\sigma(\alpha, \mathbf{t}) = \Lambda r^{\frac{1-\lambda}{\delta}} s^{\frac{1}{2}} \alpha^{\frac{2\lambda-1}{2(1-\lambda)-\delta}} \tag{20}$$

and

$$e(\alpha, \mathbf{t}) = \frac{\lambda}{\delta} r^{\frac{2(1-\lambda)-\delta}{-\delta}} s^{\frac{2(1-\lambda)-\delta}{2\lambda}} \alpha^2 \sigma(\alpha, \mathbf{t})^2. \tag{21}$$

As shown in the main text, the optimal share α^* is given by $\alpha^* = \underline{\alpha} \equiv \frac{\eta-1}{\eta}\frac{\Delta}{\Gamma}$.

$\sigma(\alpha, \mathbf{t})$ defined by (20) is decreasing in a and c. The implied mean is

$$\mu(e(\underline{\alpha}, \mathbf{t}), \sigma(\underline{\alpha}, \mathbf{t})) = \left(\frac{\lambda}{\delta}\right)^{\lambda} \Lambda^{2\lambda+\delta} \underline{\alpha}^{\frac{2\lambda-\delta}{2(1-\lambda)-\delta}} rs,$$

a decreasing function of a and c.

i) The function $\Phi(a)$ is defined by the condition $\sigma(\underline{\alpha}, \mathbf{t}) = \overline{\sigma}$, that is

$$\Lambda r^{\frac{1-\lambda}{\delta}} s^{\frac{1}{2}} \left(\frac{2\lambda - \delta}{2(1-\delta)} \frac{\Delta}{\Gamma} \right)^{\frac{2\lambda - 1}{2(1-\lambda)-\delta}} = \overline{\sigma}. \tag{22}$$

Solving for c, we obtain

$$c = \Phi(a) \equiv \Lambda^{\frac{2(1-\lambda)-\delta}{\lambda}} \left(\frac{2\lambda - \delta}{2(1-\delta)} \frac{\Delta}{\Gamma} \right)^{\frac{2\lambda-1}{\lambda}} a^{\frac{\lambda-1}{\lambda}} \overline{\sigma}^{-\frac{2(1-\lambda)-\delta}{\lambda}}.$$

iii) For values of a and c such that $c < \Phi(a)$, it is not optimal to implement $\sigma(\alpha, \mathbf{t})$ defined by (3). Instead, the relevant implementation constraint becomes (4). Two possibilities arise. Firstly, (4) holds as an equality - which is the case for values of $\mathbf{t}$ close to the locus defined by $c = \Phi(a)$. Since effort is always interior, we can compute $\alpha^*(\mathbf{t})$ from (4) as an equality:

$$\left(\frac{\lambda}{\delta} r^{\frac{2(1-\lambda)-\delta}{-\delta}} s^{\frac{2(1-\lambda)-\delta}{2\lambda}} \alpha(\mathbf{t})^2 \overline{\sigma}^2 \right)^{\lambda} \delta \overline{\sigma}^{\delta-1} - a\alpha(\mathbf{t}) \overline{\sigma} = 0,$$

which yields

$$\alpha^*(\mathbf{t}) = \left(\lambda^{\lambda} \delta^{1-\lambda} a^{\lambda-1} c^{-\lambda} \overline{\sigma}^{\delta - 2(1-\lambda)} \right)^{\frac{1}{1-2\lambda}}, \tag{23}$$

a decreasing function of a and c. Clearly also, when substituting for $c = \Phi(a)$ into (23), we obtain $\alpha^*(\mathbf{t}) = \underline{\alpha}$. Thus, the solution is continuous at the boundary $c = \Phi(a)$ separating the parameter values that give rise to interior and corner solutions, respectively. The implied mean is

$$\mu(e(\alpha^*(\mathbf{t}), \mathbf{t}), \overline{\sigma}) = \lambda^{\frac{\lambda}{1-2\lambda}} \delta^{\frac{\lambda}{1-2\lambda}} \overline{\sigma}^{\frac{\delta-2\lambda}{1-2\lambda}} a^{-\frac{\lambda}{1-2\lambda}} c^{-\frac{\lambda}{1-2\lambda}},$$

a decreasing function of a and c.

Secondly, it can be the case that (4) is satisfied automatically - i.e. holds as a strict inequality. Solving the first-order condition for effort for the optimal level of effort for any given contract α, we obtain

$$e(\alpha, \mathbf{t}) = \lambda^{\frac{1}{1-\lambda}} c^{\frac{-1}{1-\lambda}} \overline{\sigma}^{\frac{\delta}{1-\lambda}} \alpha^{\frac{1}{1-\lambda}}.$$

This implies that the equilibrium mean of the return distribution is

$$\mu^* = \mu(e(\alpha, \mathbf{t}), \overline{\sigma}) = \lambda^{\frac{\lambda}{1-\lambda}} c^{\frac{-1}{1-\lambda}} \overline{\sigma}^{\frac{\delta}{1-\lambda}} \alpha^{\frac{\lambda}{1-\lambda}}.$$

Hence, the principal's problem becomes

$$\max_{\alpha} \left\{ \lambda^{\frac{\lambda}{1-\lambda}} c^{\frac{-\lambda}{1-\lambda}} \overline{\sigma}^{\frac{\delta}{1-\lambda}} \alpha^{\frac{\lambda}{1-\lambda}} - \frac{a}{2} \overline{\sigma}^2 \alpha^2 - \lambda^{\frac{1}{1-\lambda}} c^{\frac{-\lambda}{1-\lambda}} \overline{\sigma}^{\frac{\delta}{1-\lambda}} \alpha^{\frac{1}{1-\lambda}} \right\} \tag{24}$$

The solution satisfies the first-order condition

$$\lambda^{\frac{1}{1-\lambda}} \frac{1}{1-\lambda} c^{\frac{-\lambda}{1-\lambda}} \overline{\sigma}^{\frac{\delta}{1-\lambda}} \alpha^*(\mathbf{t})^{\frac{2\lambda-1}{1-\lambda}} - a\overline{\sigma}^2 \alpha^*(\mathbf{t}) - \lambda^{\frac{1}{1-\lambda}} \frac{1}{1-\lambda} c^{\frac{-\lambda}{1-\lambda}} \overline{\sigma}^{\frac{\delta}{1-\lambda}} \alpha^*(\mathbf{t})^{\frac{\lambda}{1-\lambda}} = 0. \tag{25}$$

By the second-order condition and the fact that the left-hand side of (25) is decreasing in a and c, we note that the optimal $\alpha^*(\mathbf{t})$ is decreasing in a and c. Moreover, $\sigma^* = \overline{\sigma}$ over this range. Note again that $\mu(e(\alpha^*(\mathbf{t}), \mathbf{t}), \overline{\sigma})$ is decreasing in a and c.

Finally, suppose that (4) holds as an equality at $\hat{\mathbf{t}}$ and as a strict inequality at $\tilde{\mathbf{t}}$ and choose $\hat{\mathbf{t}}$ and $\tilde{\mathbf{t}}$ arbitrarily close to each other. Then it must be the case that $\alpha^*\left(\tilde{\mathbf{t}}\right) \geq \alpha^*\left(\hat{\mathbf{t}}\right)$. To see this, suppose we had $\alpha^*\left(\tilde{\mathbf{t}}\right) < \alpha^*\left(\hat{\mathbf{t}}\right)$. However, then the agent would have a strictly higher incentive to choose marginal incentive to increase σ at $\hat{\mathbf{t}}$, contradicting that (4) holds as an equality at $\hat{\mathbf{t}}$ and as a strict inequality at $\tilde{\mathbf{t}}$.

iv) This follows from the continuity of the solution at the boundary separating the two sets $\mathbf{T}_C$ and $\mathbf{T}_I$ in conjunction with the comparative statics properties of the optimal share $\alpha^*(\mathbf{t})$ for $\mathbf{t} \in \mathbf{T}_C$. ■

Proof of Proposition 2. Let $P(\mathbf{T}_C) \equiv \Pr(\mathbf{t} \in \mathbf{T}_C)$ and let $P(\mathbf{T}_I) \equiv \Pr(\mathbf{t} \in \mathbf{T}_I)$.

$$COV(\tilde{\alpha}^*, \tilde{\sigma}^*) = \mathbb{E}[\tilde{\alpha}^*\tilde{\sigma}^*] - \mathbb{E}[\tilde{\alpha}^*]\,\mathbb{E}[\tilde{\sigma}^*].$$

We can rewrite this as

$$\begin{aligned} COV(\tilde{\alpha}^*, \tilde{\sigma}^*) &= \mathbb{E}[\tilde{\alpha}^*\tilde{\sigma}^* |\, \mathbf{T}_C]\, P(\mathbf{T}_C) + \mathbb{E}[\tilde{\alpha}^*\tilde{\sigma}^* |\, \mathbf{T}_I]\, P(\mathbf{T}_I) \\ &\quad - \left(\mathbb{E}[\tilde{\alpha}^* |\, \mathbf{T}_C]\, P(\mathbf{T}_C) + \mathbb{E}[\tilde{\alpha}^* |\, \mathbf{T}_I]\, P(\mathbf{T}_I)\right)\left(\mathbb{E}[\tilde{\sigma}^* |\, \mathbf{T}_C]\, P(\mathbf{T}_C) + \mathbb{E}[\tilde{\sigma}^* |\, \mathbf{T}_I]\, P(\mathbf{T}_I)\right). \end{aligned}$$

Simplifying, we obtain

$$\begin{aligned} COV(\tilde{\alpha}^*, \tilde{\sigma}^*) &= \overline{\sigma}\mathbb{E}[\tilde{\alpha}^* |\, \mathbf{T}_C]\, P(\mathbf{T}_C) + \underline{\alpha}\mathbb{E}[\tilde{\sigma}^* |\, \mathbf{T}_I]\, P(\mathbf{T}_I) \\ &\quad - \left(\mathbb{E}[\tilde{\alpha}^* |\, \mathbf{T}_C]\, P(\mathbf{T}_C) + \underline{\alpha} P(\mathbf{T}_I)\right)\left(\overline{\sigma} P(\mathbf{T}_C) + \mathbb{E}[\tilde{\sigma}^* |\, \mathbf{T}_I]\, P(\mathbf{T}_I)\right). \end{aligned}$$

Multiplying out and rearranging yields

$$\begin{gathered} COV(\tilde{\alpha}^*, \tilde{\sigma}^*) = \overline{\sigma}\mathbb{E}[\tilde{\alpha}^* |\, \mathbf{T}_C]\, P(\mathbf{T}_C)(1 - P(\mathbf{T}_C)) + \underline{\alpha}\mathbb{E}[\tilde{\sigma}^* |\, \mathbf{T}_I]\, P(\mathbf{T}_I)(1 - P(\mathbf{T}_I)) \\ -\underline{\alpha}\overline{\sigma} P(\mathbf{T}_I) P(\mathbf{T}_C) - \mathbb{E}[\tilde{\sigma}^* |\, \mathbf{T}_I]\,\mathbb{E}[\tilde{\alpha}^* |\, \mathbf{T}_C]\, P(\mathbf{T}_I) P(\mathbf{T}_C). \end{gathered}$$

Using the fact that $P(\mathbf{T}_I) = 1 - P(\mathbf{T}_C)$ we can rewrite this into

$$COV(\tilde{\alpha}^*, \tilde{\sigma}^*) = \left(\overline{\sigma} - \mathbb{E}[\tilde{\sigma}^* |\, \mathbf{T}_I]\right) P(\mathbf{T}_I) \left(\mathbb{E}[\tilde{\alpha}^* |\, \mathbf{T}_C] - \underline{\alpha}\right) P(\mathbf{T}_C) > 0,$$

where the conclusion follows from the facts that $\overline{\sigma} > \mathbb{E}[\tilde{\sigma}^* |\, \mathbf{T}_I])$ and $\mathbb{E}[\tilde{\alpha}^* |\, \mathbf{T}_C] > \underline{\alpha}$.

Part ii) is a trivial consequence of the fact that the functions $\alpha^*(\mathbf{t})$, $\mu^*(\mathbf{t})$, and $\sigma^*(\mathbf{t})$ are comonotone.

Part iii): Since $\tilde{\mu}^*$ is strictly decreasing in each variable a and c, fixing one of these variables, one can write $\tilde{\alpha}^*$ as a (weakly increasing) function of $\tilde{\mu}^*$. The function $\tilde{\alpha}^*(\tilde{\mu}^*)$ is nondecreasing in $\tilde{\mu}^*$, because both $\tilde{\alpha}^*(a,c)$ and $\tilde{\mu}^*(a,c)$ are decreasing functions of a and c. We have

$$COV\left(\tilde{\alpha}^*(\tilde{\mu}^*), \tilde{\mu}^*\right) = \mathbb{E}\left[(\tilde{\mu}^* - \mathbb{E}\left[\tilde{\mu}^*\right])(\tilde{\alpha}^*(\tilde{\mu}^*) - \mathbb{E}\left[\tilde{\alpha}^*(\tilde{\mu}^*)\right])\right].$$

Expanding terms, we have

$$\begin{aligned} COV\left(\tilde{\alpha}^*(\tilde{\mu}^*), \tilde{\mu}^*\right) &= \mathbb{E}\left[\left(\tilde{\mu}^* - \mathbb{E}\left[\tilde{\mu}^*\right]\right)\left(\tilde{\alpha}^*\left(\tilde{\mu}^*\right) - \tilde{\alpha}^*\left(\mathbb{E}\left[\tilde{\mu}^*\right]\right)\right)\right] \\ &\quad + \mathbb{E}\left[\left(\tilde{\mu}^* - \mathbb{E}\left[\tilde{\mu}^*\right]\right)\left(\tilde{\alpha}^*\left(\mathbb{E}\left[\tilde{\mu}^*\right]\right) - \mathbb{E}\left[\tilde{\alpha}^*(\tilde{\mu}^*)\right]\right)\right]. \end{aligned}$$

The term on the second line is zero because $\tilde{\alpha}^*\left(\mathbb{E}\left[\tilde{\mu}^*\right]\right) - \mathbb{E}\left[\tilde{\alpha}^*(\tilde{\mu}^*)\right]$ is non-stochastic and $\mathbb{E}\left[\tilde{\mu}^* - \mathbb{E}\left[\tilde{\mu}^*\right]\right] = 0$. Hence, we have

$$COV\left(\tilde{\alpha}^*(\tilde{\mu}^*), \tilde{\mu}^*\right) = \mathbb{E}\left[\left(\tilde{\mu}^* - \mathbb{E}\left[\tilde{\mu}^*\right]\right)\left(\tilde{\alpha}^*\left(\tilde{\mu}^*\right) - \tilde{\alpha}^*\left(\mathbb{E}\left[\tilde{\mu}^*\right]\right)\right)\right] > 0.$$

The conclusion follows from the fact that $\tilde{\alpha}^*\left(\tilde{\mu}^*\right)$ is non-decreasing, so $\tilde{\alpha}^*\left(\tilde{\mu}^*\right) - \tilde{\alpha}^*\left(\mathbb{E}\left[\tilde{\mu}^*\right]\right) \gtreqless 0$ if $\tilde{\mu}^* - \mathbb{E}\left[\tilde{\mu}^*\right] \gtreqless 0$. Since the function $\tilde{\alpha}^*(\tilde{\mu}^*)$ is strictly increasing on a set of positive measure, the strict inequality holds.

The proof for the covariance of $\tilde{\mu}^*$ and $\tilde{\sigma}^*$ is identical. In particular, because $\mu^*(a,c)$ and $\sigma^*(a,c)$ are comonotone, we can write $\tilde{\sigma}^*$ as a monotonic function of $\tilde{\mu}^*$, $\tilde{\sigma}^*\left(\tilde{\mu}^*\right)$. The remainder of the argument is then exactly as stated above. ■

Proof of Lemma 1. A simple, and incentive compatible way to exclude types $\theta < \theta_m$ is to offer the null contract $\beta(\theta) = \alpha(\theta) = 0$ for all types $\theta < \theta_m$. Assume thus that the principal offers such a scheme.

Consider now incentive compatibility. A type θ should not have an incentive to mimic any type $\hat{\theta}$, so

$$\beta(\theta) + (\Delta - \Gamma)\theta\alpha(\theta)^\eta \geq \beta\left(\hat{\theta}\right) + (\Delta - \Gamma)\theta\alpha\left(\hat{\theta}\right)^\eta.$$

Likewise, a type $\hat{\theta}$ should not have an incentive to mimic any type θ, so

$$\beta\left(\hat{\theta}\right) + (\Delta - \Gamma)\hat{\theta}\alpha\left(\hat{\theta}\right)^\eta \geq \beta(\theta) + (\Delta - \Gamma)\hat{\theta}\alpha(\theta)^\eta.$$

Adding the two constraints, and rearranging, we get

$$(\Delta - \Gamma)\left(\theta - \hat{\theta}\right)\left(\alpha(\theta)^\eta - \alpha\left(\hat{\theta}\right)^\eta\right) \geq 0.$$

Hence $\alpha(\theta)$ must be nondecreasing in θ.

Let $\omega(\theta) \equiv \max_{\hat{\theta}} \left\{ \beta\left(\hat{\theta}\right) + (\Delta - \Gamma)\,\theta\alpha\left(\hat{\theta}\right)^{\eta} \right\}$. Given truthtelling at the optimum, it follows that $\omega_{\theta}(\theta) = (\Delta - \Gamma)\,\alpha(\theta)^{\eta}$ almost everywhere. Imposing the participation constraint at θ_m, we get

$$\omega(\theta) = \int_{\underline{\theta}}^{\theta} (\Delta - \Gamma)\,\alpha(z)^{\eta}\,dz = \int_{\theta_m}^{\theta} (\Delta - \Gamma)\,\alpha(z)^{\eta}\,dz,$$

where the second equality follows from the fact that $\alpha(\theta) = 0$ for all $\theta < \theta_m$. Since

$$\omega(\theta) = \beta(\theta) + (\Delta - \Gamma)\,\theta\alpha(\theta)^{\eta},$$

we have

$$\beta(\theta) = \int_{\theta_m}^{\theta} (\Delta - \Gamma)\,\alpha(z)^{\eta}\,dz - (\Delta - \Gamma)\,\theta\alpha(\theta)^{\eta} \text{ for all } \theta \geq \theta_m.$$

The proof that these conditions are sufficient is standard and thus omitted.

For completeness, observe that no type $\theta < \theta_m$ has any incentive to mimic any type $\hat{\theta} \geq \theta_m$ by the standard reasoning. In particular, the utility such a type can get this way is

$$\begin{aligned} & \beta\left(\hat{\theta}\right) + (\Delta - \Gamma)\left(\theta - \hat{\theta}\right)\alpha\left(\hat{\theta}\right)^{\eta} + (\Delta - \Gamma)\,\hat{\theta}\alpha\left(\hat{\theta}\right)^{\eta} \\ = \; & \omega\left(\hat{\theta}\right) + (\Delta - \Gamma)\left(\theta - \hat{\theta}\right)\alpha\left(\hat{\theta}\right)^{\eta} \end{aligned}$$

However, since $\omega_{\theta}(\theta) = (\Delta - \Gamma)\,\alpha(\theta)^{\eta}$ and $\omega(\theta_m) = \omega$, we have

$$\begin{aligned} & \omega\left(\hat{\theta}\right) + (\Delta - \Gamma)\left(\theta - \hat{\theta}\right)\alpha\left(\hat{\theta}\right)^{\eta} \\ = \; & \omega + \int_{\theta_m}^{\hat{\theta}} (\Delta - \Gamma)\,\alpha(z)^{\eta}\,dz + (\Delta - \Gamma)\left(\theta - \hat{\theta}\right)\alpha\left(\hat{\theta}\right)^{\eta} \\ = \; & \omega + \int_{\theta_m}^{\hat{\theta}} (\Delta - \Gamma)\left(\alpha(z)^{\eta} - \alpha\left(\hat{\theta}\right)^{\eta}\right)dz + (\Delta - \Gamma)\,(\theta - \theta_m)\,\alpha\left(\hat{\theta}\right)^{\eta} < \omega \end{aligned}$$

where the inequality follows from the monotonicity of $\alpha(\theta)$ and the fact that $\theta < \theta_m$. ■

Proof of Lemma 2. We demonstrate the following three facts:

i) for $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$,

$$F(\theta) = \int_{\underline{r}}^{\min\left\{\overline{r}, \frac{\theta}{\underline{s}}\right\}} G\left(\frac{\theta}{r}\middle| r\right) q(r)\,dr,$$

and

$$f(\theta) = \int_{\underline{r}}^{\min\left\{\overline{r}, \frac{\theta}{\underline{s}}\right\}} \frac{1}{r} g\left(\frac{\theta}{r}\middle| r\right) q(r)\,dr,$$

where $G(s|r)$ $(g(s|r))$ is the cdf (pdf) of the conditional distribution of s given r and $q(r)$ is the density of the marginal distribution of r;

ii) the distribution satisfies $f(\underline{\theta}) = 0$ and $f(\theta) > 0$ for $\theta > \underline{\theta}$;

iii) the distribution satisfies, for $\theta > \underline{\theta}$,

$$\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{\theta f(\theta)} \leq 0$$

if $g(s|r)$ satisfies

$$\frac{g_s(s|r)}{\frac{g(s|r)}{s}} \geq -1.$$

i) Consider the random variable

$$\tilde{\theta} = rs.$$

With a slight abuse of notation, let θ denote the level that the rv $\tilde{\theta}$ takes. Let $h(r,s)$ denote the joint density of r and s. Hence, for $\underline{rs} \leq \theta \leq \overline{r}\underline{s}$,

$$\Pr\left[\tilde{\theta} \leq \theta\right] = \Pr\left[rs \leq \theta\right] = \int_{\underline{r}}^{\frac{\theta}{\underline{s}}}\int_{\underline{s}}^{\frac{\theta}{r}} h(r,s)\,dsdr, \tag{26}$$

while for $\overline{r}\underline{s} \leq \theta \leq \overline{rs}$,

$$\Pr\left[\tilde{\theta} \leq \theta\right] = \Pr\left[rs \leq \theta\right] = \int_{\underline{r}}^{\overline{r}}\int_{\underline{s}}^{\frac{\theta}{r}} h(r,s)\,dsdr.$$

We treat these two cases in sequence now beginning with the former.

We can rewrite (26), for $\frac{\theta}{\underline{s}} < \overline{r}$,

$$\int_{\underline{r}}^{\frac{\theta}{\underline{s}}}\int_{\underline{s}}^{\frac{\theta}{r}} h(r,s)\,dsdr = \int_{\underline{r}}^{\frac{\theta}{\underline{s}}}\int_{\underline{s}}^{\frac{\theta}{r}} g(s|r)\,dsq(r)\,dr = \int_{\underline{r}}^{\frac{\theta}{\underline{s}}} G\left(\frac{\theta}{r}\middle| r\right) q(r)\,dr. \tag{27}$$

The derivative of this expression with respect to θ is

$$\frac{\partial}{\partial\theta}\left[\int_{\underline{r}}^{\frac{\theta}{\underline{s}}} G\left(\frac{\theta}{r}\middle| r\right) q(r)\,dr\right] = \int_{\underline{r}}^{\frac{\theta}{\underline{s}}}\frac{1}{r} g\left(\frac{\theta}{r}\middle| r\right) q(r)\,dr + G\left(\frac{\theta}{\frac{\theta}{\underline{s}}}\middle| r = \frac{\theta}{\underline{s}}\right) q\left(\frac{\theta}{\underline{s}}\right).$$

Since $G\left(\frac{\theta}{\frac{\theta}{\underline{s}}}\middle| r = \frac{\theta}{\underline{s}}\right) = 0$, this simplifies to

$$\frac{\partial}{\partial\theta}\left[\int_{\underline{r}}^{\frac{\theta}{\underline{s}}} G\left(\frac{\theta}{r}\middle| r\right) q(r)\,dr\right] = \int_{\underline{r}}^{\frac{\theta}{\underline{s}}}\frac{1}{r} g\left(\frac{\theta}{r}\middle| r\right) q(r)\,dr. \tag{28}$$

For $\frac{\theta}{\underline{s}} > \overline{r}$, we can write

$$\Pr\left[\tilde{\theta} \leq \theta\right] = \int_{\underline{r}}^{\overline{r}}\int_{\underline{s}}^{\frac{\theta}{r}} h(r,s)\,dsdr = \int_{\underline{r}}^{\overline{r}}\int_{\underline{s}}^{\frac{\theta}{r}} g(s|r)\,dsq(r)\,dr = \int_{\underline{r}}^{\overline{r}} G\left(\frac{\theta}{r}\middle| r\right) q(r)\,dr. \tag{29}$$

The derivative of this expression with respect to θ is

$$\int_{\underline{r}}^{\overline{r}} \frac{1}{r} g\left(\frac{\theta}{r}\middle| r\right) q(r)\, dr \tag{30}$$

It follows that we can write the density for all θ as

$$f(\theta) = \int_{\underline{r}}^{\min\left\{\overline{r}, \frac{\theta}{\underline{s}}\right\}} \frac{1}{r} g\left(\frac{\theta}{r}\middle| r\right) q(r)\, dr,$$

and the cdf for all θ as

$$F(\theta) = \int_{\underline{r}}^{\min\left\{\overline{r}, \frac{\theta}{\underline{s}}\right\}} G\left(\frac{\theta}{r}\middle| r\right) q(r)\, dr.$$

ii) Evaluating the density at

$$\underline{\theta} = \underline{rs},$$

we obtain

$$\int_{\underline{r}}^{\underline{r}} \frac{1}{r} g\left(\frac{\underline{rs}}{r}\middle| r\right) q(r)\, dr = 0.$$

That is, the density goes to zero at the low end. It is easy to see that the density is strictly positive for $\theta > \underline{\theta}$.

iii)

$$\frac{1-F(\theta)}{\theta f(\theta)} = \frac{1 - \int_{\underline{r}}^{\min\left\{\overline{r}, \frac{\theta}{\underline{s}}\right\}} G\left(\frac{\theta}{r}\middle| r\right) q(r)\, dr}{\int_{\underline{r}}^{\min\left\{\overline{r}, \frac{\theta}{\underline{s}}\right\}} \frac{\theta}{r} g\left(\frac{\theta}{r}\middle| r\right) q(r)\, dr}.$$

Differentiating for $\theta > \overline{r}\underline{s}$, we find that $\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{\theta f(\theta)}$ is proportional to

$$-\theta\left(\int_{\underline{r}}^{\overline{r}} \frac{1}{r} g\left(\frac{\theta}{r}\middle| r\right) q(r)\, dr\right)^2 - \left(\int_{\underline{r}}^{\overline{r}} \left(g\left(\frac{\theta}{r}\middle| r\right) + \frac{\theta}{r} g_s\left(\frac{\theta}{r}\middle| r\right)\right) \frac{q(r)}{r} dr\right)\left(1 - \int_{\underline{r}}^{\overline{r}} G\left(\frac{\theta}{r}\middle| r\right) q(r)\, dr\right).$$

The expression is negative if $g(s|\, r) + s g_s(s|\, r) \geq 0$, which is satisfied if the elasticity of the density is larger than minus unity,

$$\frac{g_s(s|\, r)}{\frac{g(s|r)}{s}} \geq -1.$$

Differentiating for $\theta < \overline{r}\underline{s}$, we find that $\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{\theta f(\theta)}$ is proportional to

$$-\left(\int_{\underline{r}}^{\frac{\theta}{\underline{s}}}\frac{1}{r}g\left(\frac{\theta}{r}\middle| r\right)q(r)\,dr + G\left(\underline{s}\middle|\frac{\theta}{\underline{s}}\right)q\left(\frac{\theta}{\underline{s}}\right)\frac{1}{\underline{s}}\right)\int_{\underline{r}}^{\frac{\theta}{\underline{s}}}\frac{\theta}{r}g\left(\frac{\theta}{r}\middle| r\right)q(r)\,dr$$

$$-\left(\int_{\underline{r}}^{\frac{\theta}{\underline{s}}}\frac{1}{r}g\left(\frac{\theta}{r}\middle| r\right)q(r)\,dr + \int_{\underline{r}}^{\frac{\theta}{\underline{s}}}\frac{\theta}{r^2}g_s\left(\frac{\theta}{r}\middle| r\right)q(r)\,dr + \underline{s}g\left(\underline{s}\middle|\frac{\theta}{\underline{s}}\right)q\left(\frac{\theta}{\underline{s}}\right)\frac{1}{\underline{s}}\right)\left(1-\int_{\underline{r}}^{\frac{\theta}{\underline{s}}}G\left(\frac{\theta}{r}\middle| r\right)q(r)\,dr\right).$$

Since $G\left(\underline{s}\middle|\frac{\theta}{\underline{s}}\right)=0$, the same condition on the conditional density of r implies the result. ■

Proof of Proposition 4. Using Leibniz' rule for differentiation of integrals, we find

$$V'(\theta_m) = -\left\{\left(\theta_m\Delta\alpha(\theta_m)^{\eta-1} - \Gamma\theta_m\alpha(\theta_m)^{\eta}\right)f(\theta_m) - (\Delta-\Gamma)\alpha(\theta_m)^{\eta}(1-F(\theta_m))\right\} + \omega f(\theta_m).$$

Recall the first-order condition for the optimal α :

$$\left((\eta-1)\theta\Delta\alpha(\theta)^{\eta-2} - \Gamma\theta\eta\alpha(\theta)^{\eta-1}\right)f(\theta) - (\Delta-\Gamma)\eta\alpha(\theta)^{\eta-1}(1-F(\theta)) = 0.$$

Multiplying by $\alpha(\theta)$ and rearranging, we find that

$$\left(\theta\Delta\alpha(\theta)^{\eta-1} - \Gamma\theta\alpha(\theta)^{\eta}\right)f(\theta) - (\Delta-\Gamma)\alpha(\theta)^{\eta}(1-F(\theta)) = \theta\frac{\Delta}{\eta}\alpha(\theta)^{\eta-1}f(\theta).$$

Noting that $\mu(\theta) = \theta\Delta\alpha(\theta)^{\eta-1}$, the result follows. ■

Proof of proposition 5. Part i): Note that $\alpha^*(\theta)$ and $\mu^*(\theta) = \Delta\theta\alpha(\theta)^{\eta-1}$ are nondecreasing functions of θ. Applying the same argument as in the proof of proposition 2 the result follows immediately.

Part ii): Since $\alpha^*(\theta)$ is nondecreasing in θ, the proof of proposition 2 can be extended to the present case if σ^* is nondecreasing in its arguments r and s and if these random variables are associated.[14]

We now give conditions for the monotonicity properties of $\sigma(\alpha,\mathbf{t})^*$. Recall that $r \equiv a^{\frac{-\delta}{2(1-\lambda)-\delta}}$ and $s \equiv c^{\frac{-2\lambda}{2(1-\lambda)-\delta}}$ and that

$$\begin{aligned}\sigma(\alpha,\mathbf{t})^* &= \Lambda a^{\frac{\lambda-1}{2(1-\lambda)-\delta}}c^{\frac{-\lambda}{2(1-\lambda)-\delta}}\alpha(\theta)^{\frac{2\lambda-1}{2(1-\lambda)-\delta}}\\ &= \Lambda r^{\frac{1-\lambda}{\delta}}s^{\frac{1}{2}}\alpha(rs)^{\frac{2\lambda-1}{2(1-\lambda)-\delta}}.\end{aligned}$$

We have

$$\frac{\partial\sigma(\alpha,\mathbf{t})^*}{\partial r} = \frac{1-\lambda}{\delta}\frac{\sigma(\alpha,\mathbf{t})^*}{r} + \frac{2\lambda-1}{2(1-\lambda)-\delta}\frac{\sigma(\alpha,\mathbf{t})^*}{\alpha}\frac{\partial\alpha(\theta)}{\partial\theta}s,$$

[14] In principle, one could apply the same logic to the case where σ is decreasing in its arguments to conclude that the covariance between $\tilde{\alpha}^*$ and $-\tilde{\sigma}^*$ becomes positive (and hence the covariance between $\tilde{\alpha}^*$ and $\tilde{\sigma}^*$ negative); however, we have not been able to find a meaningful sufficient condition on the distribution that ensure this.

so $\frac{\partial\sigma(\alpha,\mathbf{t})^*}{\partial r} \geq 0 (\leq 0)$ if and only if

$$\alpha(\theta) \geq (\leq) \frac{\delta(1-2\lambda)}{(1-\lambda)(2(1-\lambda)-\delta)} \frac{\partial\alpha(\theta)}{\partial\theta}\theta.$$

Similarly, we have

$$\frac{\partial\sigma(\alpha,\mathbf{t})^*}{\partial s} = \frac{1}{2}\frac{\sigma(\alpha,\mathbf{t})^*}{s} + \frac{2\lambda-1}{2(1-\lambda)-\delta}\frac{\sigma(\alpha,\mathbf{t})^*}{\alpha(\theta)}\frac{\partial\alpha(\theta)}{\partial\theta}r,$$

so $\frac{\partial\sigma(\alpha,\mathbf{t})^*}{\partial s} \geq 0 (\leq 0)$ if and only if

$$\alpha(\theta) \geq (\leq) \frac{2(1-2\lambda)}{2(1-\lambda)-\delta}\frac{\partial\alpha(\theta)}{\partial\theta}\theta.$$

Noting that $2 \geq \frac{\delta}{1-\lambda}$ by assumption, $\sigma(\alpha,\mathbf{t})$ is increasing in both arguments iff

$$\alpha(\theta) \geq \frac{2(1-2\lambda)}{2(1-\lambda)-\delta}\frac{\partial\alpha(\theta)}{\partial\theta}\theta.$$

Likewise, $\sigma(\alpha,\mathbf{t})$ is decreasing in both arguments iff

$$\alpha(\theta) \leq \frac{\delta(1-2\lambda)}{(1-\lambda)(2(1-\lambda)-\delta)}\frac{\partial\alpha(\theta)}{\partial\theta}\theta.$$

Differentiating the optimal α with respect to θ, we obtain

$$\begin{aligned}
\alpha_\theta(\theta) &= -\frac{\eta-1}{\eta}\Delta\left(\Gamma+(\Delta-\Gamma)\frac{1-F(\theta)}{\theta f(\theta)}\right)^{-2}(\Delta-\Gamma)\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{\theta f(\theta)} \\
&= -\alpha(\theta)\frac{\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{\theta f(\theta)}}{\frac{\Gamma}{\Delta-\Gamma}+\frac{1-F(\theta)}{\theta f(\theta)}} \\
&= -\alpha(\theta)\frac{\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{\theta f(\theta)}}{\frac{\delta+2\lambda}{2(1-\lambda)-\delta}+\frac{1-F(\theta)}{\theta f(\theta)}}
\end{aligned}$$

Therefore, σ is increasing in both arguments iff

$$\alpha(\theta) \geq -\alpha(\theta)\frac{2(1-2\lambda)}{2(1-\lambda)-\delta}\frac{\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{\theta f(\theta)}}{\frac{\delta+2\lambda}{2(1-\lambda)-\delta}+\frac{1-F(\theta)}{\theta f(\theta)}}\theta$$

Since $\delta \leq 2\lambda$ implies that $\frac{2(1-2\lambda)}{2(1-\lambda)-\delta} \leq 1$, a sufficient condition is

$$-\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{\theta f(\theta)}\theta \leq \frac{\delta+2\lambda}{2(1-\lambda)-\delta}+\frac{1-F(\theta)}{\theta f(\theta)}$$

which is equivalent to

$$-\frac{\partial}{\partial\theta}\frac{1-F(\theta)}{f(\theta)} \leq \frac{\delta+2\lambda}{2(1-\lambda)-\delta},$$

the condition given in the proposition. ■

9 References

- Ackerberg, D., and Botticini, M. (2002) "Endogenous Matching and the Empirical Determinants of Contract Form", Journal of Political Economy, 110(3), 564-591
- Aggarwal, R. and Samwick, A. (2002) "The Other Side of the Trade-off: The Impact of Risk on Executive Compensation - A Reply",
- Allen, D. and Lueck, D. (1992) "Contract Choice in Modern Agriculture: Cash Rent versus Crop Share", Journal of Law and Economics, 35, 397-426
- Allen, D. and Lueck, D. (1995) "Risk, Uncertainty and Contracts", Manuscript New Haven, Yale Law School
- Araujo et al. (2007) "The Trade-Off Between Incentives and Endogenous Risk", Revista de Econometria, 27, 193-198
- Armstrong, M. (1996) "Multiproduct Nonlinear Pricing" Econometrica, 64, 51-75
- Bandiera et al. (2009) "Matching Firms, Managers, and Incentives", CEPR Discussion paper 7207
- Berle, A. and Means, G. (1932) "The Modern Corporation and Private Property", New York: Macmillan
- Biais et al. (2000) "Competing Mechanisms in a Common Value Environment", Econometrica, 68(4), 799-838
- Budde, J. and Kräkel, M. (2011) "Limited Liability and the Risk-Incentive Relationship", Discussion paper 08-03 in Economics and Management
- Bushmann et al. (1996) "CEO compensation: The role of individual performance evaluation", Journal of Accounting and Economics, 21, 161-193
- Core, J. and Guay, W. (1999) "The use of equity grants to manage optimal equity incentive levels", Journal of Accounting and Economics, 28, 151-184
- Demski, J. and Dye, R. (1999) "Risk, Return, and Moral Hazard", Journal of Accounting Research, 37(1), 27-55

- Esary, J. and Proschan, F. (1972) “Relationship among some Concepts of Bivariate Dependence”, The Annals of Mathematical Statistics, 43(2), 651-655
- Esary et al. (1967) “Association of Random Variables with Applications”, The Annals of Mathematical Statistics, 38(5), 1466-1474
- Garcia, D. (2013)“Optimal Contracts with Privately Informed Agents and Active Principals”, mimeo, UNC at Chapel Hill
- Garen, J. (1994) “Executive Compensation and Principal-Agent Theory”, The Journal of Political Economy, 102(6), 1175-1199
- Greene, William H. (1993), "Econometric Analysis", Prentice Hall International
- Hellwig, M. (2009), “A Reconsideration of the Jensen-Meckling Model of Outside Finance”J. Finan. Intermediation 18, 495–525
- 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
- Holmström, B. and Milgrom, P. (1994) “The Firm as an Incentive System”, American Economic Association, 84(4), 972-991
- Inderst, R. and Müller, H. (2010) “CEO Replacement under Private Information”, Rev. Financ. Stud. 23 (8): 2935-2969.
- Ittner et al. (1997) “The Choice of Performance Measures in Annual Bonus Contracts”, The Accounting Review, 72(2), 231-255
- Lafontaine, F. (1992) “Agency Theory and Franchising: Some Empirical Results”, The RAND Journal of Economics, 23(2), 263-283
- Lafontaine, F. and Slade, M. (1998) “Incentive contracting and the franchise decision”, NBER working paper 6544
- Lafontaine, F. and Slade, M. (2007) “Vertical Integration and Firm Boundaries: The Evidence”, Journal of Economic Literature, 45, 629-685

- Lambert, R. and Larcker, D. (1987) “An Analysis of the Use of Accounting and Market Measures of Performance in Executive Compensation Contracts”, Journal of Accounting Research, 25, 85-125
- Lazear, E. (1986) “Salaries and Piece Rates”, Journal of Business, 59(3), 405-431
- Oyer, P. and Schaefer, S. (2004) “Why do some firms give stock options to all employees?: An empirical examination of alternative theories”, NBER Working Paper 10222
- Prendergast, C. (1999) “The Provision of Incentives in Firms”, Journal of Economic Literature, 37(1), 7-63
- Prendergast, C. (2002) “The Tenuous Trade-off between Risk and Incentives”, Journal of Political Economy, 110(5), 1071-1102
- Raith, M. (2003) “Competition, Risk and Managerial Incentives”, American Economic Review, 93, 1425-1436
- Rao, C. (1971) “Uncertainty, Entrepreneurship, and Sharecropping in India”, Journal of Political Economy, 79(3), 578-595
- Rochet, J.-C. and Choné, P. (1998) “Ironing, Sweeping, and Multidimensional Screening” Econometrica , 66(4), 783-826
- Serfes, K. (2005) “Risk sharing vs. incentives: Contract design under two-sided heterogeneity”, Economics Letters, 88(3), 343-349
- Shaked, M. and Shantikumar, J. (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, 18(3), 1021-1073
- Weinschenk, P. (2014) “Sharecropping and Performance; A Reexamination of the Marshallian Hypothesis” mimeo MPI Bonn
- Wright, D. (2004) “The Risk and Incentive Trade-off in the Presence of Heterogeneous Managers”, Journal of Economics, 83(3), 209-233
- Yermack (1994) “Do Corporations award CEO stock options effectively”, NYU working paper