Volatility, Noise, and Incentives

George Baker
Harvard University and NBER

Bjorn Jorgensen
Columbia University

March 27, 2003

Preliminary : Comments Welcome

Abstract

We develop a simple agency model that helps explain the ambiguous logical and empirical connection between environmental uncertainty and incentive strength. Our model stresses that two kinds of uncertainty (which we label volatility and noise) are important in the determination of optimal incentive strength. We define noise as uncertainty to which the agent should not react by changing his actions, and volatility as uncertainty whose outcome does change the agent's optimal action choice. We show that, consistent with standard agency theory, an increase in noise reduces optimal incentive strength. Our surprising new result is that an increase in volatility actually increases optimal incentive strength in most circumstances. We also show that, when effort is contractible, the optimal linear contract in the presence of noise is first best and puts no weight on output. However, in the presence of volatility the principal will use output-based compensation in an optimal linear incentive contract. Our results shed light on, and are consistent with the "controllability principle," by which agents should not be held accountable for risks that are beyond their control, and should be held accountable for controllable risks.

We would like to thank Luis Garicano and Canice Prendergast for their thoughtful comments, and participants in seminars at Columbia, Cornell, Syracuse and USC.

Volatility, Noise, and Incentives

Preliminary: Comments Welcome

In his paper, "The Tenuous Tradeoff between Risk and Incentives," Prendergast (2002) argues persuasively that both casual and formal empiricism fail to confirm one of the central predictions of standard agency theory: that there should be a negative tradeoff between the amount of uncertainty in the environment and the strength of the incentives that an optimal incentive contract delivers. The intuition behind the standard model is simple and compelling: since employees are risk averse and firms are risk neutral, more volatile environments should lead firms to offer more "insurance" against risk to their employees, and thus reduce the slope of any incentive contract offered.

This simple logic has not convinced all observers, however. There is an alternative logic that has also been discussed, although not modeled, in the literature. This logic is articulated in an early paper by Demsetz and Lehn:

> "In less predictable environments, however, managerial behavior … figures more prominently in a firm's fortunes…. Hence, noisier environments should give rise to more concentrated ownership structures." Demsetz and Lehn (1985) p. 1159.

This argument would suggest that in more volatile environments, incentives should be stronger, not weaker. These conflicting predictions are both borne out (!) in empirical findings: as Prendergast shows, no convincing empirical relationship has been established between the uncertainty of the environment and the strength of incentives.

In this paper, we argue that this logical and empirical ambiguity results from a failure to distinguish between different types of uncertainty, which we label "noise" and "volatility." Noise is uncertainty whose realization does not affect the agent's optimal action choice, either because the agent cannot react to it, or because it does not change his optimal actions. Examples include such random factors as measurement error in output, acts of God which shift the magnitude of measured output without changing how the agent's actions affect output, or events to which the agent cannot respond. Volatility, in contrast, is uncertainty that does affect the agent's optimal action, and which the agent is able to react to. Examples of volatility include events that change the way that output is produced, changing the value of the agent's actions.

The distinction between noise and volatility is a product of two characteristics, one affecting the information structure and the other the production function. The first characteristic is whether or not the uncertainty affects the agent's marginal product of effort; the second is whether the agent observes the realization of the uncertainty before choosing his effort level. The figure below shows how we distinguish between noise and volatility.

		Agent receives signal before acting?	
		YES	NO
Uncertainty affects marginal product of effort?	YES	**Volatility**	**Noise**
	NO	**Noise**	**Noise**

We develop two models with exactly the same production functions and costs, but with different timing. In the first model, a risk neutral principal and a risk averse agent agree to an incentive contract, and then the agent chooses his effort level before the state of the world is revealed. In the second model, the agent has pre-decision information (PDI): the state of the world is revealed to the agent before choosing his effort level. In both models, the state of the world is characterized by two random variables, one of which (θ) affects the agent's marginal product of effort, and the other (ε) does not. We show that, in the no-PDI model, increases in the variance of θ and ε reduce the slope of the optimal linear contract. In the PDI model, we show that an increase in the variance of the random variable that does not affect his marginal product (ε) leads to a reduction in the slope of the optimal linear incentive contract, while an increase in the variance of random variable that does affect his marginal product (θ) actually leads to an increase in the optimal incentive strength under most circumstances. This last result is our main finding.

The result that an increase in volatility can increase optimal incentive strength depends crucially on the fact that the uncertainty affects the marginal product, *and* that the agent receives this information before choosing his effort level. Our interpretation of this result shows why both of these elements are necessary. In a model with pre-decision information that affects the agent's optimal action, the randomness in the problem should not be interpreted as noise. Rather, it is valuable information about what the agent ought to do. The fact that the agent possesses this information before choosing his action gives the principal an opportunity to make use of this information through her design of the incentive contract. An increase in the variance of θ now not only increases the amount of risk that must be borne by the agent, but also increases the amount of information that the agent has. The principal wants to increase the strength of the incentive contract in order to get the agent to "pay more attention" to his private information.

This same intuition applies to our second result. In a model with observable and contractible effort, the distinction between noise and volatility again becomes salient. In the model without PDI, contractible effort leads to first best outcomes, with the principal simply using a contract that gets the agent to set effort at the optimal level. The principal has no use for output-based pay, which would inefficiently impose risk on the risk averse agent. Similarly, in a model with PDI, if the uncertainty is of the type that does not affect the marginal product, then the principal again uses a contract that sets effort at the optimal level, and uses no output-based pay. However, if the agent has pre-decision information about his marginal product, then the optimal linear contract will be based on both effort and output. The principal will put weight on output, thereby imposing some risk on the agent, in order to get him to use his private information in his effort choice decision.

The intuition behind these two results highlights one other important assumption of our model: communication between the agent and the principal after the information is revealed to the agent is not possible. This might be because such communication is infeasible, for example when a machine operator senses that her machine is malfunctioning and must react immediately, or when a salesman gets information from a customer during a sales call and must respond on the spot. Alternatively, it might be that the sort of information that is revealed is "soft," in the sense of Stein (2002), and thus cannot be communicated efficiently to the principal. In either case, we have in mind situations in which it is not possible for the principal to design a mechanism which ties the agent's payoff to the report of some signal sent after the state of the world is revealed, but before his action is chosen.

Both the assumptions and results of this model are reminiscent of Hayek's (1954) analysis of the market economy, and Jensen and Meckling's (1992) analysis of "specific knowledge," and its role in organization design. Hayek argues that the "miracle" of the price system is not so much that it efficiently allocates scarce resources, as that it is a system that makes efficient use of highly dispersed information. He argues that the utilization of what he calls "on-the-spot" knowledge, which cannot be efficiently aggregated or communicated to a central planner, is critical to economic efficiency. For Hayek, the price system is not so much a mechanism for getting people to take efficient actions, as a system that gets them to use their individual, idiosyncratic information efficiently.

Similarly, Jensen and Meckling argue that the existence of specific knowledge—information available to the agent which is costly to transmit to the principal—is the main reason for decentralization in organizations. They argue that the utilization of specific knowledge is enhanced when organizations decentralize, but that an attendant cost is the increased need for incentives. They also argue that organizations in more volatile environments are likely to be characterized by greater amounts of dispersed specific knowledge, and are thus likely to be characterized by higher-powered incentives and more output-based pay.

Finally, our definition of volatility—uncertainty that can be and should be acted on by the agent—brings theoretical clarity to a standard piece of managerial prescription. The so-called “controllability principle” states that managers should be held accountable for events that they control, and not held accountable for events that are beyond their control. (Dekin and Maher 1991) An adequate understanding of this simple prescription has eluded theorists for decades. (See Antle an Demski 1988) Indeed, even an understanding of the seemingly contradictory notion of controllable risk was lacking. However, our model provides a plausible definition of a “controllable risk,” and delivers results that are consistent with the controllability principle. Events whose outcomes either have no effect on managers’ optimal actions, or to which managers cannot react, are “uncontrollable.” Increases in these sorts of risks should be met with a decrease in the strength of incentives. However, risks whose outcome the manager can and optimally should react to can be thought of as controllable, and managers should face strong incentives when dealing with these types of risks.

The paper proceeds as follows. In Section 2, we specify the model, and show how it compares to other models in the agency literature. We discuss the role of “post-contractual, pre-decision information” (PCPDI, or simply PDI) in agency models, and show how this type of information affects the intuition and the results in such models. We also discuss why solutions to the sort of PDI models that we construct have not been tackled in the past.

In section 3, we derive the conditions for the optimal incentive contract with and without PDI, and with and without observable agent actions. While we are unable to derive a closed-form solution for the optimal slope of the incentive contract in many of these models, we are able to partially characterize the optimal contract, and to derive the results above. In Section 4 we discuss applications and possible empirical tests of this model. Section 5 concludes.

II. The Basic Model

Output is:

$$V = \theta a + \varepsilon$$

where a is the agent's effort, and $\theta \sim N\left(\mu_\theta, \sigma_\theta^2\right)$ and $\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right)$ are two independent random variables that characterize the state of the world. The principal and the agent share common knowledge about the distributions of θ and ε. Output should be interpreted as the benefits to the owners of the firm before the agent has been paid. To ensure tractability, and to make interpretation of the results simple, we restrict the agent's compensation to be linear. The agent's compensation (without contractible effort) is:

$$Pay = S + b_V V$$

where S is the fixed component of the agent's compensation and b_V is the "piece rate" or sensitivity of compensation to output. To avoid wealth effects on the manager's risk aversion, we assume that the agent has constant absolute risk aversion, $\beta > 0$. Finally, the agent's disutility of effort measured in pecuniary terms is quadratic in effort, that is, $c(a) = ka^2/2$ where $k > 0$. Therefore, the agent's expected utility is given by:

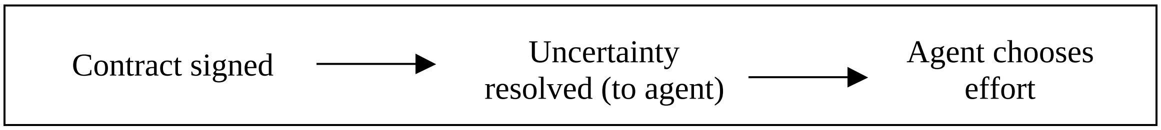
$$U(\bullet) = -\exp\left[-\beta\left(S + b_V V - \frac{k\,a^2}{2}\right)\right].$$

The principal knows that the agent will accepts any employment contract that gives the agent his reservation level of certainty equivalent, W. We make two different assumptions about the timing of effort choice and information revelation. In our first model, the timing is as follows:

Contract signed	→	Agent chooses effort	→	Uncertainty resolved

In this version of the model, the principal and the agent agree on a contract, the agent chooses effort, and then the uncertainty about θ and ε is revealed. Both θ and ε are "noise" parameters, in that they garble the link between the agent's action and output. As we will show in section 3, an increase in "noise" unambiguously decreases the slope of an optimal incentive contract. Furthermore, since the principal and the agent have the same information about how actions affect outcomes, the principal knows what she wants the agent to do. If effort is observable, then the (risk neutral) principal will set effort at the optimal level, and perfectly insure the (risk averse) agent against risk.

With pre-decision information, the timing is as follows:



As discussed in the introduction, we assume that it is not possible for the agent to communicate to the principal his private information before choosing his effort level.

Under these assumptions, it is no longer appropriate to think about θ as noise. θ gives the agent private information about his marginal product. Indeed, in this model σ_θ^2 is a measure of how much private information the agent possesses. The agent's effort depends on θ, and, as will be shown below, expected output ($E[V]$) is an increasing function of σ_θ^2.

Finding solutions to models with a multiplicative random term, especially with post-contractual pre-decision information, is very difficult, at least under the assumption that the agent is risk averse.[1] With PDI, when the contract is written, effort is uncertain for both the principal and the agent. Thus, the agent's participation constraint, and the principal's maximization problem, are evaluated treating θ, ε, and a(θ) as random variables. However, the agent's incentive compatibility constraint (the agent's effort choice decision) is evaluated after θ and ε are realized. Because of the complexity of the first-order conditions in this model, we are not able to get closed-form solutions for the optimal contract. However, we are able to derive comparative statics about this optimal contract.

[1] Baker (1992) solves models of this sort for risk neutral agents.

Our view on the importance of pre-decision information is not new. Holmstrom (1979) notes near the end of this seminal paper:

> In many respects the model we have analyzed is very primitive. One unrealistic feature is the assumption that the agent chooses his action having the same information as the principal, that is, before anything about θ [which characterizes the state of the world] is revealed. Commonly this will not be the case. After the sharing rule is fixed, the agent will often learn something new about the difficulty of his task, or the environment in which it is to be performed.

Holmstrom goes on to show that, even in the presence of PDI, his informativeness result continues to hold. However, he does not (nor does he try to) show how an optimal contract under PDI varies with increasing risk.

Harris and Raviv (1979) also build a model (Model 2 in their paper) in which the agent has pre-decision information. They show that, when the state-of-the-world is unobservable to the principal and the agent is risk averse, both the principal and the agent prefer a contract that includes monitoring of the agent (i.e. observing effort) to one that does not. They do not show, however, that the optimal contract will still depend on output even when effort is perfectly observable by the principal.

Prior models with multiplicative uncertainty when the agent does not receive private information include Sung (1995) and Feltham and Wu (2001). Sung (1995) establishes that linear contracts are optimal in a continuous-time setting where an agent's effort controls the variance of firm value. Feltham and Wu (2001) consider the granting of options when the agent's preferences exhibit mean-variance separation.[2] Christensen (1981) demonstrates that the principal can be strictly worse off when an agent privately informed. In a closely related paper, Bushman, Indjejikian, and Penno (2000) study the benefits from delegation. While they focus on the case of risk-neutral agent, they do derive the first order conditions for model in which firm value is the product of effort and a random variable.

[2] In our setting, where the agent receives private information, mean-variance representation does not arise.

III. Solving the model with and without pre-decision information

A. Model without PDI

In this section, we initially assume that the agent chooses his action before θ and ε are revealed. We begin by assuming that effort is not contractible. The principal maximizes the expected value of output minus payments to the agent, subject to the agent's participation constraint and his incentive compatibility constraint. The agent has expected utility equal to:

$$-E\left[e^{-\beta\left(S+b_V V-c(a)\right)}\right]=-e^{-\beta\left(S-c(a)\right)}E\left[e^{-\beta b_V V}\right]$$
$$=-e^{-\beta\left(S-c(a)\right)}e^{-\beta b_V a\mu_\theta+\frac{\left(\beta b_V\right)^2}{2}\left(a^2\sigma_\theta^2+\sigma_\varepsilon^2\right)}.$$

The certainty equivalent is:

(1) $$CE\left(S+b_V V,a\right)=S+b_V a\mu_\theta-\beta\frac{b_V{}^2a^2}{2}\sigma_\theta^2-\beta\frac{b_V{}^2}{2}\sigma_\varepsilon^2-c(a)$$

We assume that the agent has an outside option that yields utility W. The participation constraint is thus:

(2) $$S+b_V a\mu_\theta-\beta\frac{b_V{}^2a^2}{2}\sigma_\theta^2-\beta\frac{b_V{}^2}{2}\sigma_\varepsilon^2-c(a)\geq W$$

The agent chooses effort before observing θ or ε. Thus, the agent maximizes his certainty equivalent:

(3) $$\underset{a}{Max}\left\{S+b_V a\mu_\theta-\beta\frac{b_V{}^2a^2}{2}\sigma_\theta^2-\beta\frac{b_V{}^2}{2}\sigma_\varepsilon^2-c(a)\right\}.$$

The solution to this maximization problem yields the agent's optimal action choice given that he does not possess pre-decision information.

(4) $$a^* = \frac{b_v \mu_\theta}{k + \beta b_v^2 \sigma_\theta^2}$$

Note that, if σ_θ^2 or β are zero, this reduces to the standard result in the literature (that is, in a model without a multiplicative error.) However, the second term in the denominator arises because the agent's optimal action choice is reduced by his anticipation that higher levels of effort will lead to higher risk; since he is risk averse, he reduces his level of effort by an amount that depends on his risk aversion parameter and the variance of the multiplicative random term.

Given this optimal effort, the agent's certainty equivalent is

$$CE\left(S + b_v\left(a^*\theta + \varepsilon\right), a^* = \frac{b_v \mu_\theta}{k + \beta b_v{}^2 \sigma_\theta^2}\right)$$

$$= S + ab_v\mu_\theta - a^2\left[\beta \frac{b_v{}^2}{2}\sigma_\theta^2 + \frac{k}{2}\right] - \beta \frac{b_v{}^2}{2}\sigma_\varepsilon^2$$

$$= S + \frac{\left(b_v\mu_\theta\right)^2}{2\left(k + \beta b_v{}^2\sigma_\theta^2\right)} - \beta\frac{b_v{}^2}{2}\sigma_\varepsilon^2$$

Since the agent's participation constraint

$$CE\left(S + b\left(a^*\theta + \varepsilon\right) a^* = \frac{b\mu_\theta}{k + \beta b^2 \sigma_\theta^2}\right) \geq W$$

is binding, the principal chooses the fixed compensation such that

$$S = W - \frac{\left(b_V\mu_\theta\right)^2}{2\left(k + \beta b_V^2\sigma_\theta^2\right)} + \beta\frac{b_V^2}{2}\sigma_\varepsilon^2$$

The principal chooses the pay-to-performance to maximize

$$E\left[V - S - b_v V\right] = \left(1 - b_v\right)E\left[V\right] - S = \left(1 - b_v\right)E\left[\theta a^* + \varepsilon\right] - S$$

$$= \left(1 - b_v\right)\mu_\theta \left\{ \frac{b_v \mu_\theta}{k + \beta b_v{}^2 \sigma_\theta^2} \right\} - W + \frac{\left(b_v \mu_\theta\right)^2}{2\left(k + \beta b_v{}^2 \sigma_\theta^2\right)} - \beta \frac{b_v{}^2}{2} \sigma_\varepsilon^2$$

The principal's maximization problem is:

$$(5) \qquad \underset{b_v}{Max} \; \frac{\mu_\theta^2}{\left(k + \beta b_v{}^2 \sigma_\theta^2\right)} \left\{ b_v - \frac{b_v{}^2}{2} \right\} - W - \beta \frac{b_v{}^2}{2} \sigma_\varepsilon^2$$

Before solving for the optimal contract, two preliminary results are easily established: the principal is made unambiguously worse off by an increase in either σ^2_ε or σ_θ^2.[3] The first of these results is obvious from examination of equation 5. The second is equally obvious once we establish that b_v* is less than one, which we do below.

The first order condition is:

$$0 = \frac{\partial}{\partial b_v} E\left[V - S - b_v V\right]$$

$$= \frac{\mu_\theta^2}{\left(k + \beta b_v{}^2 \sigma_\theta^2\right)^2} \left\{ \left(1 - b_v\right)\left(k + \beta b_v{}^2 \sigma_\theta^2\right) - \left(b_v - \frac{b^2}{2}\right)\left(2\beta b_v \sigma_\theta^2\right) \right\} - \beta b_v \sigma_\varepsilon^2$$

$$= \frac{\mu_\theta^2}{\left(k + \beta b_v{}^2 \sigma_\theta^2\right)^2} \left\{ \left(1 - b_v\right)k + \left[\left(b_v{}^2 - b_v{}^3\right) - \left(2b_v{}^2 - b_v{}^3\right)\right]\beta \sigma_\theta^2 \right\} - \beta b_v \sigma_\varepsilon^2$$

$$(6) \qquad = \frac{\mu_\theta^2}{k + \beta b_v{}^2 \sigma_\theta^2} - \frac{k b_v \mu_\theta^2}{\left(k + \beta b_v{}^2 \sigma_\theta^2\right)^2} - \beta b_v \sigma_\varepsilon^2 - \frac{2 b_v{}^2 \beta \sigma_\theta^2 \mu_\theta^2}{\left(k + \beta b_v{}^2 \sigma_\theta^2\right)^2} \qquad \text{(FOC)}$$

Several terms in this expression are intuitive. The first is the marginal effect of an increase in b_v on output, without considering the effect of the agent's risk aversion on

[3] This result is consistent with Kim (1995) and Kim and Suh (1991). In their models, the agent's effort affects only the mean, but not the variance, of firm value.

effort choice: it is of course positive. All of the other terms are negative if b_V^* is positive. The second term is the marginal disutility of effort brought about by an increase in b_V, again ignoring the effect of the agent's risk aversion on effort choice. The third term, which depends on σ_ε^2, is due to the effect of ε (the additive random variable) on the agent's risk aversion. The fourth term combines all of the risk effects that come from the multiplicative random marginal product of effort. Note that, if β is zero (that is, the agent is risk neutral), then the third and fourth terms are zero, and the optimal b_v is one.

We are now able to derive several results from the "no PDI" model. In order to do so, we rewrite equation (6) and define a function L:

$$L\left(b_V \mid k, \beta, \sigma_\theta^2, \sigma_\varepsilon^2\right) = \frac{\mu_\theta^2}{\left(k + \beta b_V{}^2 \sigma_\theta^2\right)^2}\left\{\left(1 - b_V\right)k - b_V{}^2 \beta \sigma_\theta^2\right\} - \beta b_V \sigma_\varepsilon^2,$$

which equals the first order condition. We show in Appendix A1 that the second order condition for this problem is satisfied, and that the optimal b_v is between 0 and 1. Thus:

$$\frac{\partial}{\partial b_V} L\left(b_V \mid k, \beta, \sigma_\theta^2, \sigma_\varepsilon^2\right) < 0.$$

Differentiating L with respect to σ_ε^2 and σ_θ^2 allows us, by the implicit function theorem, to derive two results:

Results 1&2: In a model with non-contractible effort and without PDI:

(R1) $$\frac{db_V^*}{d\sigma_\varepsilon^2} < 0$$

(R2) $$\frac{db_V^*}{d\sigma_\theta^2} < 0$$

Proofs: See Appendix A2.

These results are to be expected. Since both ε and θ only add noise to the relationship between the agent's actions and output, an increase in their variances reduces the slope of an optimal contract for a risk-averse agent.

Contractible Effort

We now assume that the principal can observe effort perfectly, and pays for it with a linear piece rate equal to b_a. She also pays for output, V.

$Pay = S + b_a a + b_V V$.

With contractible effort, the principal can write a contract that induces the agent to choose the optimal effort level (given the information available) while permitting the agent to bear no risk.

Results 3&4: In a model with contractible effort and without PDI:

(R3) $b_a^* = \mu_\theta$

(R4) $b_V^* = 0$.

Proofs: See Appendix A3.

In this model, the principal sets the piece rate on effort equal to the (expected) marginal product of effort, and puts no weight on output in the incentive contract. Note that the principal could also have used a forcing contract, specifying that the agent choose effort equal to μ_θ/k. Either of these contracts achieve optimal effort levels with no risk to the agent.

An important feature of this model without pre-decision information is that the principal knows, as well as the agent, what the optimal action is. Thus the agency problem is one of inducing the agent to take the action that the principal wants, and the principal's problem is to infer (from a potentially noisy signal) what the agent did. The agent has no specific knowledge in this model. Thus, if the principal has a noiseless signal about that the agent did, she can write a first-best incentive contract. This is very different from the model with pre-decision information, where the principal does not know what she would like the agent to do, and so uses incentive contracting to get the agent to utilize his specific knowledge.

B. Model with PDI

In our PDI model, we again begin under the assumption that the principal cannot contract on effort. She pays:

$$S + b_V V \cdot$$

Now, however, the agent's effort choice decision is taken after θ and ε have been revealed to him. This implies that his maximization problem is:

$$\underset{a}{Max}\left\{-e^{-\beta(S+bV-c(a))}\right\} = \underset{a}{Max}\left\{-e^{-\beta\left(S+b_V\theta a-\frac{ka^2}{2}\right)}\right\}$$

The solution to this maximization yields:

(7) $$a^*(\theta) = \frac{b_V}{k}\theta .$$

Note that, under PDI, the agent's optimal action is a function of θ, the (privately revealed) marginal product of effort.

With the optimal effort shown in equation (7), the agent's expected utility at the time of the signing of the contract (prior to observing θ) can be evaluated as

$$E\left[U\left(S + b_V\left(\theta a^*(\theta) + \varepsilon\right) - \frac{k}{2}\left(a^*(\theta)\right)^2\right)\right]$$

$$= \int \exp\left\{-\beta CE\left[\left(S + b_V\left(\theta a^*(\theta) + \varepsilon\right) - \frac{k}{2}\left(a^*(\theta)\right)^2\right)|\theta\right]\right\} f(\theta)d\theta$$

$$= \int \exp\left\{-\beta\left(S - \frac{\beta}{2} b_V^2\sigma_\varepsilon^2 + \frac{(b_V\theta)^2}{2k}\right)\right\} f(\theta)d\theta$$

$$= \exp\left\{-\beta\left(S - \frac{\beta}{2} b_V^2\sigma_\varepsilon^2\right)\right\}\int \exp\left\{-\beta\frac{(b_V\theta)^2}{2k}\right\} f(\theta)d\theta$$

This expression can be converted into the certainty equivalent below. (See Appendix B1 for derivation.)

(8) $$CE\left(S, b_V, a^*(\theta)\right) = S - \frac{\beta}{2} b_V^2 \sigma_\varepsilon^2 + \frac{b_V^2 \mu_\theta^2}{2\left(k + \beta b_V^2 \sigma_\theta^2\right)} + \frac{\beta^{-1}}{2} \ln\left(1 + \beta \frac{b_V^2 \sigma_\theta^2}{k}\right).$$

Since the agent's participation constraint

$$CE\left(S, b_V, a^*(\theta)\right) \geq W$$

is binding:

$$S - \frac{\beta}{2} b_V^2 \sigma_\varepsilon^2 + \frac{\{b_V \mu_\theta\}^2}{2\left(k + \beta \left[b_V \sigma_\theta\right]^2\right)} + \frac{\beta^{-1}}{2} \ln\left(1 + \beta \frac{\left[b_V \sigma_\theta\right]^2}{k}\right) = W$$

or

(9) $$-S = -\frac{\beta}{2} b_V^2 \sigma_\varepsilon^2 + \frac{\{b_V \mu_\theta\}^2}{2\left(k + \beta \left[b_V \sigma_\theta\right]^2\right)} + \frac{\beta^{-1}}{2} \ln\left(1 + \beta \frac{\left[b_V \sigma_\theta\right]^2}{k}\right) - W$$

Expected output given the optimal effort choice shown in Eq. 7 is:

$$E[V] = E\left[\theta a^*(\theta)\right] = b_V E\left[\theta^2\right]/k = b_V\left(VAR[\theta] + E[\theta]\right)/k = b_V\left(\sigma_\theta^2 + \mu_\theta^2\right)/k$$

The risk-neutral principal has (*ex ante*) expected pay-off of:

(10) $$E\left[V - \left(S + b_V V\right)\right] = \left(1 - b_V\right) E[V] - S = \left(1 - b_V\right) b_V \frac{\left(\sigma_\theta^2 + \mu_\theta^2\right)}{k} - S$$

Combining equations 9 and 10 yields the principal's maximization problem:

(11) $$\underset{b_V}{Max}\left\{E\left[V - \left(S + b_V V\right)\right]\right\} = \underset{b_V}{Max}\left\{\left(1 - b_V\right) b_V \frac{\left(\sigma_\theta^2 + \mu_\theta^2\right)}{k} - \frac{\beta}{2} b_V^2 \sigma_\varepsilon^2\right.$$

$$+\frac{\{b_V\mu_\theta\}^2}{2\left(k+\beta[b_V\sigma_\theta]^2\right)}+\frac{\beta^{-1}}{2}\ln\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)-W\left.\vphantom{\frac{1}{1}}\right\}$$

Differentiating (11) with respect to b_V yields the first order condition for an optimal linear contract. Once again, define this first order condition to be a function J as follows:

(12) $$J\left(b_V \mid \sigma_\varepsilon^2, \sigma_\theta^2, \mu_\theta, k, \beta\right) = 0 =$$

$$-\beta b_V\sigma_\varepsilon^2+\mu_\theta^2 b_V\frac{k}{\left(k+\beta b_V^2\sigma_\theta^2\right)^2}+\frac{b_V\sigma_\theta^2}{\left(k+\beta b_V^2\sigma_\theta^2\right)}+\frac{(1-2b_V)\left[\sigma_\theta^2+\mu_\theta^2\right]}{k} \quad (FOC)$$

There are several things to notice about this first order condition. Note first that if $\sigma_\theta^2=0$, then the optimal b_v is the same as in Holmstrom-Milgrom (1987):

$$b_v^* = \frac{\mu_\theta^2}{\mu_\theta^2+\beta k\sigma_\varepsilon^2}.$$

This is because, with $\sigma_\theta^2=0$, the agent receives no action-relevant pre-decision information. He still receives pre-decision information (about ε), but this information has no effect on his action choice, nor on the optimal linear contract.[4] Notice also that, as σ^2_ε goes to infinity, b_v goes to zero.

As in the no-PDI model, we will not derive a closed form solution to b_V^*. However, we can characterize b_V^* in several important ways. First, we show that an optimal b_V^* exists, and is between 0 and 1 for all possible parameter values. (See Appendix B2 for proof.) We can then use the implicit function theorem to sign the derivatives of b_V^* with respect

[4] This result only holds with linear contracts and exponential utility functions. If the realization of ε affected either the slope of the incentive contract, or the marginal utility of effort, then his pre-decision observation of ε would affect his action choice.

to β , σ_ε^2 and σ_θ^2 . In Appendix B2, we show that the second order condition for the principal's problem is satisfied for all values of b_V . Therefore we know that

(13) $$\frac{dJ}{db_V^*} < 0.$$

Differentiating J with respect to β , σ_ε^2 and σ_θ^2 yields the following three expressions.

(14) $$\frac{\partial}{\partial\beta} J\left(b_V \mid \sigma_\theta^2, \sigma_\varepsilon^2, \mu_\theta, k, \beta\right) = -b_V\sigma_\varepsilon^2 - 2\mu_\theta^2 \frac{k\beta b_V^3\sigma_\theta^2}{\left(k + \beta b_V^2\sigma_\theta^2\right)^3} - \sigma_\theta^2 \frac{b_V^3\sigma_\theta^2}{\left(k + \beta b_V^2\sigma_\theta^2\right)^2} < 0$$

(15) $$\frac{\partial}{\partial\sigma_\varepsilon^2} J\left(b_V \mid \sigma_\varepsilon^2, \sigma_\theta^2, \mu_\theta, k, \beta\right) = -\beta b_V < 0$$

(16) $$\frac{\partial}{\partial\sigma_\theta^2} J\left(b_V \mid \sigma_\varepsilon^2, \sigma_\theta^2, \mu_\theta, k, \beta\right) = \frac{2k\beta b_V^3\mu_\theta^2}{\left(k + \beta b_V^2\sigma_\theta^2\right)^3} + \frac{kb_V}{\left(k + \beta b_V^2\sigma_\theta^2\right)^2} + \frac{\left(1 - 2b_V\right)}{k}$$

This set of derivations allows us to compute comparative statics on b_V^* .

Results 5: With pre-decision information and non-contractible effort:

(R5) $$\frac{db_V^*}{d\beta} < 0$$

Proof: By the implicit function theorem, since

$$\frac{dJ}{db_V^*} < 0 \text{ and } \frac{dJ}{d\beta} < 0.$$

Results 6: With pre-decision information and non-contractible effort:

(R5) $\frac{db_V^*}{d\sigma_\varepsilon^2} < 0$

Proof: By the implicit function theorem, since

$\frac{dJ}{db_V^*} < 0$ and. $\frac{dJ}{d\sigma_\varepsilon^2} < 0$

These two results confirm that this model of pre-decision information does not overturn any of the standard results in agency theory. If the agent is less risk averse, the optimal slope of the incentive contract is higher. Also, if the variance of the additive noise component of the production function (ε) increases, the optimal slope decreases.

Result 6: With pre-decision information and non-contractible effort, there exist parameter values where:

(R6) $\frac{db_V^*}{d\sigma_\theta^2} > 0.$

The fact that $\frac{db_V^*}{d\sigma_\theta^2}$ may be positive is the main result of this paper. Intuition about this result can be developed by rewriting Equation 12 (the first order condition):

$$0 = \frac{(\sigma_\theta^2 + \mu_\theta^2)}{k} - \frac{b_V(\sigma_\theta^2 + \mu_\theta^2)}{k} - \beta b_V \sigma_\varepsilon^2 - \frac{\beta\, b_V^3 \sigma_\theta^2 \left[k\sigma_\theta^2 + 2k\mu_\theta^2 + \beta\, b_V^2 \sigma_\theta^2 (\sigma_\theta^2 + \mu_\theta^2)\right]}{k(k + \beta\, b_V^2 \sigma_\theta^2)^2} \tag{17}$$

The first term represents the marginal effect of an increase in b_V on output. Note that it is increasing in σ_θ^2. The second term is the marginal effect of b_V on the disutility of effort. The third term is the effect of b_V on the agent's disutility for risk that is coming from the *additive* noise (that which comes from ε) in the production function. Finally, the fourth term is the effect of an increase of b_V on the agent's risk aversion coming from the multiplicative noise.

This expression makes it clear how an increase in σ_θ^2 can increase b_V^*. Since an increase in σ_θ^2 raises the marginal value of b_V (the first term in Eq. 17), and all of the remaining terms become more negative as b_V increases (by the second order condition), then if σ_θ^2 makes the second and fourth terms more negative, it drives b_V^* up.

The conditions under which $\frac{db_V^*}{d\sigma_\theta^2} > 0$ can be seen by looking at Equation 16. It is clear that $\frac{db_V^*}{d\sigma_\theta^2}$ will be positive when b_V^* is small. As discussed above, b_V^* can be driven arbitrarily small by allowing σ_ε^2 to get large (by inspection of Equation 17). Thus, we can assure that $\frac{db_V^*}{d\sigma_\theta^2}$ is positive by choosing σ_ε^2 to be sufficiently large. However, we have also performed simulations on $\frac{db_V^*}{d\sigma_\theta^2}$, and have found that it is positive over most of parameter space. Only when σ_ε^2 is very small, or μ_θ is very large is $\frac{db_V^*}{d\sigma_\theta^2}$ negative.

Contractible Effort

We now model an agent who gets pre-decision information, but whose effort level is observable and contractible. Once again, we choose a functional form for pay in which both effort and output are rewarded linearly by the principal.[5]

---

[5] There is a non-linear contract which achieves outcomes arbitrarily close to first-best with contractible effort. This contract involves paying the agent a bonus equal exactly to his disutility of effort ($ka^2/2$). This makes him indifferent about his effort level. Then add an arbitrarily small bonus paid on V-c(a), to get him to choose the optimal effort level.. While this contract works in theory, it is highly vulnerable to small errors in the measurement of output, since the agent is paid for whatever he does.

$$Pay = S + b_a a + b_V V \ .$$

In a model with pre-decision information, the principal no longer wants to set the agent's effort level at μ_θ/k, because the agent has private information that would allow him to adjust his effort with differing states of the world. Thus, a forcing contract will not be optimal.

The expected utility of the agent is

$$-E\left[e^{-\beta\left(S + b_a a + b_V V - c(a)\right)} \mid \theta,\varepsilon\right] = -e^{-\beta\left(S + b_a a - c(a)\right)} E\left[e^{-\beta b_V V} \mid \theta,\varepsilon\right] =$$
$$-e^{-\beta\left(S + b_a a - c(a)\right)} e^{-\beta b_V E[V \mid \theta,\varepsilon]}$$

and the agent's certainty equivalent is

$$CE\left(S + b_a a + b_V V, a \mid \theta,\varepsilon\right) = S + b_a a + b_V a\theta + b_V \varepsilon - c(a).$$

The first order condition for the agent's optimal action choice is:

$$0 = b_a + b_V\theta - ka$$

or

$$a^*(\theta) = \frac{b_a + b_V\theta}{k}$$

and the agent's certainty equivalent for any given realization of the productivity parameter, θ, is easily evaluated as

$$CE\left(S + b_a a + b_V V, a(\theta) = \frac{b_a + b_V\theta}{k} \mid \theta,\varepsilon\right) = S + \frac{\left(b_a + b_V\theta\right)^2}{2k} + b_V\varepsilon$$

The derivation of the agent's expected utility at the time of contracting is shown in Appendix B3. Combining this with the agent's (binding) participation constraint allows us to derive the principal's profit net of compensation to the agent:

$$E[V]-\left\{S+E\left[b_a a^*(\theta)+b_V V\right]\right\}=(1-b_V)E[V]-b_a E\left[a^*(\theta)\right]-S$$
$$=(1-b_V)k^{-1}\left\{b_a\mu_\theta+b_V\left(\sigma_\theta^2+\mu_\theta^2\right)\right\}-b_a\left(\frac{b_a+b_V\mu_\theta}{k}\right)-S$$

(See Appendix B3.)

Differentiating the principal's profit net of compensation with respect to b_a and b_V yields the two-equation system of non-linear equations shown below.

$$0=(1-2b_V)\mu_\theta-2b_a+k\frac{\{\mu_\theta b_V+b_a\}}{\left(k+\beta b_V^2\sigma_\theta^2\right)}$$

and

$$0=(1-2b_V)\left(\sigma_\theta^2+\mu_\theta^2\right)-2b_a\mu_\theta+k\left\{\frac{b_V\sigma_\theta^2-\mu_\theta\{\mu_\theta b_V+b_a\}}{\left(k+\beta b_V^2\sigma_\theta^2\right)}-\frac{\{\mu_\theta b_V+b_a\}^2\left(\beta b_V\sigma_\theta^2\right)}{\left(k+\beta b_V^2\sigma_\theta^2\right)^2}-\beta b_V\sigma_\varepsilon^2\right\}$$

We are unable to solve this system of equations, or even to derive a set of comparative statics comparable to those calculated above. However, it is possible to prove that setting either b_a or b_V equal to zero leads to a contradiction. (See Appendix B4.) Thus we can establish the following results:

Results 7&8: In a model with pre-decision information and contractible effort, the optimal contract will put weight on both effort and output.

(R7) $b_a^*\neq 0$

(R8) $b_V^*\neq 0$.

Proof: See Appendix B4.

This result establishes the fact that, even when effort can be costlessly monitored and contracted upon, the principal will still choose to put weight on output in an optimal linear contract when the agent has pre-decision information.

IV. Implications and Empirical Tests

We show that, as the uncertainty of the environment increases, the optimal slope of an incentive contract may increase. The intuition driving this result is simple, and derives from the difference between what we have labeled "volatility" (variability that affects the manager's optimal action choice) and the "noise" which is usually modeled in agency theory. θ in this pre-decision information model is not noise, but a valuable signal that the agent possesses allowing him to choose effort in a state-contingent way. This distinction between volatility and noise is what drives the new results in this model, and ultimately must be at the heart of any attempt to confirm this model empirically.

Before we discuss empirical tests, however, it is worth differentiating our results from those of Prendergast's model. Although both models are motivated by the positive, negative, or ambiguous empirical relationship between risk and incentive strength, there are other predictions of the two models that differ. The most important distinction is that our model makes no prediction about (indeed has no role for) delegation. While Prendergast's model derives a prediction of a positive relation between risk and incentives by assuming that increased volatility leads to increased delegation, we make no such prediction. Thus we would not argue that a positive empirical relationship between risk and incentives is the result of "omitted variable bias," caused by the statistician's inability to observe decision rights. Rather, this relationship results from the fact that, when the agent has more pre-decision information, it is more valuable to give him higher-powered incentives. Thus, in contrast with Prendergast, we would expect to find instances in which, *holding delegation constant*, there could be higher incentives in riskier environments.

The fact that our model has no role for delegation may make it a better one for studying the relationship between risk and CEO incentives. As Prendergast points out, his model is not a particularly good one for understanding CEO pay, since there is likely to be little variation in the amount of delegation to the CEO from the board. Virtually all boards delegate virtually all decision rights to their CEOs. Thus, our model may be the more appropriate one to explain any positive relationships found between CEO pay and environmental uncertainty.

One other distinction between our model and that of Prendergast is that our model predicts the simultaneous use of both input-based and output-based pay. In the Prendergast model, there is some level of risk at which the firm switches from monitoring inputs to monitoring outputs. In our model, the firm uses both types of incentive at all times, choosing the optimal weight to put on each.

Our model helps resolve the ambiguous findings on the relationship between risk and incentive strength, and suggests an empirical strategy for clarifying this ambiguity. The novelty in our model is that we predict that *certain types of risk* should increase optimal incentive strength, while other types should decrease incentive strength. Risk of the sort that affects managers' optimal action choices (which we have called volatility) will increase the slope of optimal incentive packages. Risk that is "uncontrollable," in the sense that managers ought not or cannot change their actions in response to the state of the world, should reduce the slope of the incentive contract. How might one test this prediction from our model?

One strategy would be to look for indicators that the risk that a firm faces is more or less likely to be controllable. Thus, one might expect that events that affect an entire industry are more likely to be controllable by managers than events that are idiosyncratic to particular firms. This suggests an empirical test in which the total variance in performance of a firm is decomposed into the industry variance and the firm-specific variance. If industry-wide turbulence is more controllable, then the effect of industry variance on incentive strength should be less negative than the effect of firm variance, and perhaps even positive.

On the other hand, it is surely true that there are certain types of events (such as macro-economic fluctuations) whose consequences managers have little ability to mitigate. To the extent that the variability in a firm's value is driven mainly by such uncontrollables, we would predict that it would have a lower-powered incentive scheme than that for a firm whose variability is driven by controllable risks. Thus our model would predict that, holding the total variability of firm value constant, high-beta firms should have lower-powered incentives than low-beta firms.

Another possible empirical approach might be to look for firms whose performance variability is driven by changes in its output prices or exchange rates, versus firms whose performance variability is driven by changes in input markets. If one assumes that variability in input markets provides managers with more opportunities to adjust and react, then this sort of variability might increase incentive strength, or at least not decrease it as much as variability that derives from instability in output market prices.

V. Conclusion

We develop a very simple model of an agent with valuable specific knowledge, and show that increases in turbulence in such a model can lead to increases, rather than decreases, in incentive strength. While the solution to this model is quite messy, the intuition is very clear. An agent with post-contractual pre-decision information has knowledge about his optimal action choice that the principal lacks. This knowledge can be measured by the variance of the pre-decision signal received by the agent. When this variance increases, the value of getting the agent to "pay attention" to the signal goes up, and so may outweigh the countervailing benefit of shifting risk away from the risk-averse agent.

We believe that the approach that we have developed is superior in many respects to the standard one in the literature, in which the principal is as well informed as the agent about optimal actions. It incorporates much of the intuition of the multi-tasking models of Holmstrom and Milgrom (1991), and provides a formal but intuitive way to model concepts like specific knowledge and controllability. Unfortunately, the model is very hard to solve, and requires particular assumptions about the structure of the random terms and the agent's utility function. Nonetheless, we feel that the benefits from working with this model outweigh the costs.

References

Baker, George P. 1992. Incentive Contracts and Performance Measurement. *Journal of Political Economy* 100 (June): 598-614.

Bushman, Robert, Raffi Indjejikian, and Mark Penno. 2000. Private Pre-decision Information, Performance Measure Congruity and the Value of Delegation. *Contemporary Accounting Research* 17 (Winter): 561-587.

Christensen, John. 1981. Communication in Agencies. *The Bell Journal of Economics* 12 (Autumn): 661-674.

Dekin, Edward and Michael Maher. 1991 *Cost Accounting.* 3rd edition. Irwin, Homewood, IL.

Demsetz, Harold, and Kenneth Lehn. 1985. The Structure of Corporate Ownership: Causes and Consequences. *Journal of Political Economy* 93 (December): 1155-1177.

Feltham, Gerald A., and Martin G. Wu. 2001. Incentive efficiency of Stock versus Options. *Review of Accounting Studies* 6 (March): 7-28.

Harris, Milton and Artur Raviv. 1979. Optimal Incentive Contracts with Imperfect Information. *Journal of Economic Theory* 20: 231-259.

Holmstrom, Bengt. 1979. Moral Hazard and Observability. *Bell Journal of Economics* 10 (Spring): 74-91.

Holmstrom, Bengt, and Paul Milgrom. 1987. Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica* 55 (March): 303-328.

Holmstrom, Bengt, and Paul Milgrom. 1991. Multi-task Principal-agent Analysis: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organizations* 7: 24-52.

Jensen, Michael and William Meckling. 1992. Specific and General Knowledge and Organizational Structure. *Contract Economics*, Lars Werin and Hans Wijkander, eds. (Oxford: Blackwell)

Kim, Son Ku. 1995. Efficiency of an Information System in an Agency Model. *Econometrica* 36 (January): 89-102.

Kim, Son Ku, and Yoon S. Suh. 1991. Ranking of Information Systems for Management Control. *Journal of Accounting Research* 29 (Autumn): 386-396.

Prendergast, Canice. 2002. The Tenuous Trade-off between Risk and Incentives. *Journal of Political Economy* 110 (October): 1071-1102.

Stein, Jeremy. 2002. Information Production and Capital Allocation: Decentralized versus Hierarchical Firms. *Journal of Finance* LVII (October).

Sung, Jaeyoung. 1995. Linearity with Project Selection and Controllable Diffusion Rate in Continuous-time Principal-agent Problems. *The Rand Journal of Economics* 26 (Winter): 720-744.

Appendix A: No Pre-Decision Information

Non-contractible effort case.

A0-1: Proof that principal prefers non-negative b_v.

Refer to the principal's objective function in (5):

$$l\left(b_V \mid \sigma_\theta^2,\sigma_\varepsilon^2\right) \equiv E\left[V - S - b_V V\right] = \frac{\mu_\theta^2}{\left(k+\beta b_V^2\sigma_\theta^2\right)}\left\{b_V - \frac{b_V^2}{2}\right\} - W - \beta\frac{b_V^2}{2}\sigma_\varepsilon^2$$

Note that for any two contracts with opposite sign on the slope, the principal prefers the positive slope, that is,

$$l\left(b_V \mid \sigma_\theta^2,\sigma_\varepsilon^2\right) \geq 0 \geq l\left(-b_V \mid \sigma_\theta^2,\sigma_\varepsilon^2\right).$$

Hence, the optimal slope is non-negative, $b_V^* \geq 0$.

Further note that $\left(b_V = 0, a = 0, S = -W\right)$ is a feasible solution which yields expected payoff of $l\left(0 \mid \sigma_\theta^2,\sigma_\varepsilon^2\right) = -W$ to the principal. Consequently, since $b_V^\# > 2$:
$l\left(b_V^\# \mid \sigma_\theta^2,\sigma_\varepsilon^2\right) < l\left(2 \mid \sigma_\theta^2,\sigma_\varepsilon^2\right) < -W$ it follows that $b_V^* < 2$.

A1: Proof that second order condition is negative for positive b_v.

The first-order condition for the principal's choice of b_V is:

$$L(b_V \mid k, \beta, \sigma_\theta^2, \sigma_\varepsilon^2) = \frac{\mu_\theta^2}{(k + \beta b_V^2 \sigma_\theta^2)^2}\{(1 - b_V)k - \beta b_V^2 \sigma_\theta^2\} - \beta b_V \sigma_\varepsilon^2$$

By differentiation

$$\frac{\partial}{\partial b_V} L(b_V \mid k, \beta, \sigma_\theta^2, \sigma_\varepsilon^2)$$

$$= \mu_\theta^2 \frac{\{[-k - 2b_V \beta \sigma_\theta^2](k + \beta b_V^2 \sigma_\theta^2)^2 - [(1 - b_V)k - \beta b_V^2 \sigma_\theta^2] 2(k + \beta b_V^2 \sigma_\theta^2)\beta\, 2 b_V \sigma_\theta^2\}}{(k + \beta b_V^2 \sigma_\theta^2)^4} - \beta \sigma_\varepsilon^2$$

$$= -\mu_\theta^2 \frac{\{[k + 2b_V \beta \sigma_\theta^2](k + \beta b_V^2 \sigma_\theta^2) + [(1 - b_V)k - \beta b_V^2 \sigma_\theta^2] 4\beta b_V \sigma_\theta^2\}}{(k + \beta b_V^2 \sigma_\theta^2)^3} - \beta \sigma_\varepsilon^2$$

$$= -\mu_\theta^2 \frac{[k + 2b_V \beta \sigma_\theta^2]}{(k + \beta b_V^2 \sigma_\theta^2)^2} - \left\{ \frac{\mu_\theta^2 [(1 - b_V)k - \beta b_V^2 \sigma_\theta^2]}{(k + \beta b_V^2 \sigma_\theta^2)^2} \right\} \frac{4\beta b_V \sigma_\theta^2}{(k + \beta b_V^2 \sigma_\theta^2)} - \beta \sigma_\varepsilon^2$$

Substituting in the first order condition in the second term

$$\frac{\partial}{\partial b_V} L(b_V = b_V^* \mid k, \beta, \sigma_\theta^2, \sigma_\varepsilon^2)$$

$$= -\mu_\theta^2 \frac{[k b_V + 2 b_V^2 \beta \sigma_\theta^2]}{b_V (k + \beta b_V^2 \sigma_\theta^2)^2} - \{\beta b_V \sigma_\varepsilon^2\} \frac{2(k + \beta b_V^2 \sigma_\theta^2)\beta\, 2 b_V \sigma_\theta^2}{(k + \beta b_V^2 \sigma_\theta^2)} - \beta \sigma_\varepsilon^2$$

$$= -\mu_\theta^2 \frac{[k + 2b_V \beta \sigma_\theta^2]}{(k + \beta b_V^2 \sigma_\theta^2)^2} - \left\{ \frac{4\beta\, 2 b_V^2 \sigma_\theta^2}{(k + \beta b_V^2 \sigma_\theta^2)} + 1 \right\} \beta \sigma_\varepsilon^2 < 0$$

Since this expression is negative for every positive values of b_V^*, every positive solution to the first order condition is a local maximum. We also know that the function, $\frac{\partial}{\partial b_V} L(b_V)$ is continuous and differentiable everywhere which means that every positive interior local maximum is also a global interior maximum, since (from A0-1) the optimal

b_V^* is non-negative. As is standard, we rule out the only remaining possible optimum, $b_V^* = 0$, by assuming that it is worthwhile for the principal to hire the agent.

Finally since L is a continuous function and

$$L\left(0 \mid k, \beta, \sigma_\theta^2, \sigma_\varepsilon^2\right) = \mu_\theta^2 > 0$$

$$L\left(1 \mid k, \beta, \sigma_\theta^2, \sigma_\varepsilon^2\right) = \frac{\mu_\theta^2}{\left(k + \beta\sigma_\theta^2\right)^2}\left\{-\beta\sigma_\theta^2\right\} - \beta\sigma_\varepsilon^2 < 0$$

the unique interior optimum, $b_V^* \quad [0,1]$.

A2: Proofs that $\frac{db_V^*}{d\sigma_\varepsilon^2} < 0$ and $\frac{db_V^*}{d\sigma_\theta^2} < 0$.

The first result follows immediately from A1 and implicit function theorem since

$\frac{\partial}{\partial\sigma_\varepsilon^2} L\left(b_V \mid k, \beta, \sigma_\theta^2, \sigma_\varepsilon^2\right) = -\beta b_V < 0$ for all $b_V > 0$.

To prove the second result note that

$$\begin{aligned}
&\frac{\partial}{\partial\sigma_\theta^2} L\left(b_V \mid k, \beta, \sigma_\theta^2, \sigma_\varepsilon^2\right) \\
&= \mu_\theta^2 \frac{\left\{\left[-b_V^2\beta\right]\left(k + \beta b_V^2\sigma_\theta^2\right)^2 - \left[(1 - b_V)k - \beta b_V^2\sigma_\theta^2\right] 2\left(k + \beta b_V^2\sigma_\theta^2\right)\beta b_V^2\right\}}{\left(k + \beta b_V^2\sigma_\theta^2\right)^4} \\
&= -\mu_\theta^2 b_V^2 \beta \frac{\left\{3k - 2b_V k - \beta b_V^2\sigma_\theta^2\right\}}{\left(k + \beta b_V^2\sigma_\theta^2\right)^3}
\end{aligned}$$

For b_V $\left[0, \min\{1, b_{\max}\}\right]$: $\left\{3k - 2b_V k - b_V^2\beta\sigma_\theta^2\right\} \geq \left\{k - b_V^2\beta\sigma_\theta^2\right\} \geq 0$ such that

$\frac{\partial}{\partial\sigma_\theta^2} L\left(b_V \mid k, \beta, \sigma_\theta^2, \sigma_\varepsilon^2\right) < 0$.

A3: Proof that $b_a^* = \mu_\theta$ and $b_v^*=0$

When both effort and firm value are available for contracting, the agent receives the contract $S + b_a a + b_V V$, then his expected utility is

$$-E\left[e^{-\beta\left(S+b_a a+b_V V-c(a)\right)}\right] = -e^{-\beta\left(S+b_a a-c(a)\right)} E\left[e^{-\beta b_V V}\right] = -e^{-\beta\left(S+b_a a-c(a)\right)} e^{-\beta b_V a\mu_\theta + \frac{(\beta b_V)^2}{2}\left(a^2\sigma_\theta^2+\sigma_\varepsilon^2\right)}$$

and the certainty equivalent is

$$CE\left(S + b_a a + b_V V, a\right) = S + b_a a + b_V a\mu_\theta - \beta\frac{b_V^2 a^2}{2}\sigma_\theta^2 - \beta\frac{b_V^2}{2}\sigma_\varepsilon^2 - c(a)$$

The first order condition for an interior optimum is

$$0 = b_a + b_V\mu_\theta - \beta b_V^2 a\sigma_\theta^2 - ka$$

or

$$a^* = \frac{b_a + b_V\mu_\theta}{k + \beta b_V^2\sigma_\theta^2}$$

Note that $\frac{\partial a^*}{\partial b_a} = \frac{1}{k + \beta b_V^2\sigma_\theta^2} > 0$ so any effort level can be implemented even though $\mathbf{b}_V = 0$. Given this optimal effort, the agent's certainty equivalent is

$$CE\left(S + b_a a^* + b_V\left(a^*\theta + \varepsilon\right) a^* = \frac{b_a + b_V\mu_\theta}{k + \beta b_V^2\sigma_\theta^2}\right)$$

$$= S + a\left(b_a + b_V\mu_\theta\right) - a^2\left[\beta\frac{b_V^2}{2}\sigma_\theta^2 + \frac{k}{2}\right] - \beta\frac{b_V^2}{2}\sigma_\varepsilon^2$$

$$= S + \frac{\left(b_a + b_V\mu_\theta\right)^2}{2\left(k + \beta b_V^2\sigma_\theta^2\right)} - \beta\frac{b_V^2}{2}\sigma_\varepsilon^2$$

and the participation constraint reduces to

$$-S = \frac{(b_a + b_V\mu_\theta)^2}{2(k+\beta b_V^2\sigma_\theta^2)} - \beta\frac{b_V^2}{2}\sigma_\varepsilon^2 - W$$

Clearly the optimal pay-to-performance, b, is zero (and independent of the variance of the performance measure used.

Define the principal's objective function similar to (5):

$$m(b_a, b_V \mid \sigma_\theta^2, \sigma_\varepsilon^2) \equiv E[V - S - b_a a - b_V V] = (1-b_V)a\mu_\theta - b_a a - S$$
$$= (1-b_V)\left[\frac{b_a + b_V\mu_\theta}{k+\beta b_V^2\sigma_\theta^2}\right]\mu_\theta - b_a\left[\frac{b_a + b_V\mu_\theta}{k+\beta b_V^2\sigma_\theta^2}\right] + \left\{\frac{(b_a + b_V\mu_\theta)^2}{2(k+\beta b_V^2\sigma_\theta^2)} - \beta\frac{b_V^2}{2}\sigma_\varepsilon^2 - W\right\}$$
$$= \left[\frac{b_a + b_V\mu_\theta}{k+\beta b_V^2\sigma_\theta^2}\right]\mu_\theta - \frac{(b_a + b_V\mu_\theta)^2}{2(k+\beta b_V^2\sigma_\theta^2)} - \beta\frac{b_V^2}{2}\sigma_\varepsilon^2 - W$$

The first order condition with respect to $\mathbf{b}_a$ is

$$0 = \frac{\partial}{\partial b_a} m(b_a, b_V \mid \sigma_\theta^2, \sigma_\varepsilon^2) = \left[\frac{1}{k+\beta b_V^2\sigma_\theta^2}\right]\mu_\theta - \frac{(b_a + b_V\mu_\theta)}{(k+\beta b_V^2\sigma_\theta^2)}$$

or $b_a = (1-b_V)\mu_\theta$. Further

$$\frac{\partial}{\partial b_V} m(b_a, b_V \mid \sigma_\theta^2, \sigma_\varepsilon^2)$$
$$= \left[\frac{\mu_\theta(k+\beta b_V^2\sigma_\theta^2) - (b_a + b_V\mu_\theta)\beta\, 2b_V\sigma_\theta^2}{(k+\beta b_V^2\sigma_\theta^2)^2}\right]\mu_\theta - \frac{(b_a + b_V\mu_\theta)\mu_\theta(k+\beta b_V^2\sigma_\theta^2) - (b_a + b_V\mu_\theta)^2\beta b_V\sigma_\theta^2}{(k+\beta b_V^2\sigma_\theta^2)^2} - \beta b_V\sigma_\varepsilon^2$$
$$= \frac{\mu_\theta - (b_a + b_V\mu_\theta)}{(k+\beta b_V^2\sigma_\theta^2)}\mu_\theta - (b_a + b_V\mu_\theta)\frac{2\mu_\theta - (b_a + b_V\mu_\theta)}{(k+\beta b_V^2\sigma_\theta^2)^2}\beta b_V\sigma_\theta^2 - \beta b_V\sigma_\varepsilon^2$$

By substition of $b_a = (1-b_V)\mu_\theta$ then

$$\frac{\partial}{\partial b_V} m(b_a, b_V \mid \sigma_\theta^2, \sigma_\varepsilon^2) = 0 - (b_a + b_V\mu_\theta)\frac{\mu_\theta}{(k+\beta b_V^2\sigma_\theta^2)^2}\beta b_V\sigma_\theta^2 - \beta b_V\sigma_\varepsilon^2 < 0$$ for all $b_V \geq 0$,

from which it follows that $b_V^* = 0$, $b_a^* = \mu_\theta$, and $a^* = \frac{\mu_\theta}{k}$.

Appendix B: Pre-Decision Information

Appendix B1: Derivation of agent's ex ante certainty equivalent with PDI and non-contractible effort.

The agent's expected utility is:

$$E\left[U\left(S + b_V\left(\theta a^*(\theta) + \varepsilon\right) - \frac{k}{2}\left(a^*(\theta)\right)^2\right)\right]$$

$$= \int \exp\left\{-\beta CE\left[\left(S + b_V\left(\theta a^*(\theta) + \varepsilon\right) - \frac{k}{2}\left(a^*(\theta)\right)^2\right) | \theta\right]\right\} f(\theta)d\theta$$

$$= \int \exp\left\{-\beta\left(S - \frac{\beta}{2} b_V^2 \sigma_\varepsilon^2 + \frac{(b_V\theta)^2}{2k}\right)\right\} f(\theta)d\theta$$

$$= \exp\left\{-\beta\left(S - \frac{\beta}{2} b_V^2 \sigma_\varepsilon^2\right)\right\} \int \exp\left\{-\beta \frac{(b_V\theta)^2}{2k}\right\} f(\theta)d\theta$$

To evaluate the integral

$$\int \exp\left\{-\beta \frac{(b_V\theta)^2}{2k}\right\} f(\theta)d\theta$$

$$= \int \exp\left\{-\beta \frac{(b_V\theta)^2}{2k}\right\} \frac{1}{\sigma_\theta\sqrt{2\pi}} \exp\left\{\frac{-1}{2}\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2\right\} d\theta$$

$$= \sigma_\theta^{-1} \int \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2 + \beta\frac{(\theta b_V)^2}{k}\right]\right\} d\theta$$

we first analyze the term in the square bracket in the exponent

$$\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2+\beta\frac{(\theta b_V)^2}{k}$$

$$=\theta^2\sigma_\theta^{-2}\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)-2\theta\sigma_\theta^{-2}\mu_\theta+\sigma_\theta^{-2}\mu_\theta^2$$

$$=\sigma_\theta^{-2}\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)\left[\theta^2-2\theta\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)^{-1}\mu_\theta\right]+\sigma_\theta^{-2}\mu_\theta^2$$

$$=\sigma_\theta^{-2}\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)\left[\theta^2-2\theta\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)^{-1}\mu_\theta+\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)^{-2}\mu_\theta^2\right]$$

$$-\sigma_\theta^{-2}\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)^{-2}\mu_\theta^2+\sigma_\theta^{-2}\mu_\theta^2$$

$$=\sigma_\theta^{-2}\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)\left[\theta-\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)^{-1}\mu_\theta\right]^2-\sigma_\theta^{-2}\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)^{-1}\mu_\theta^2+\sigma_\theta^{-2}\mu_\theta^2$$

To simplify the derivation define the three constants, $\{\Sigma_\theta^{-2}, M_\theta, K_\theta\}$, as follows

$$\Sigma_\theta^{-2}=\sigma_\theta^{-2}\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)>0$$

$$M_\theta=\Sigma_\theta^2\sigma_\theta^{-2}\mu_\theta=\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)^{-1}\mu_\theta$$

$$2\beta K_\theta=-M_\theta^2\Sigma_\theta^{-2}+\sigma_\theta^{-2}\mu_\theta^2=-\sigma_\theta^{-2}\left(1+\beta\frac{[b_V\sigma_\theta]^2}{k}\right)^{-1}\mu_\theta^2+\sigma_\theta^{-2}\mu_\theta^2$$

then

$$\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2+\beta\frac{(\theta b_V)^2}{k}$$

$$=\theta^2\Sigma_\theta^{-2}-2\theta M_\theta\Sigma_\theta^{-2}+\sigma_\theta^{-2}\mu_\theta^2$$

$$=\theta^2\Sigma_\theta^{-2}-2\theta M_\theta\Sigma_\theta^{-2}+M_\theta^2\Sigma_\theta^{-2}-M_\theta^2\Sigma_\theta^{-2}+\sigma_\theta^{-2}\mu_\theta^2$$

$$=\frac{(\theta-M_\theta)^2}{\Sigma_\theta^2}+2\beta K_\theta$$

Hence

$$
\begin{aligned}
&\int \exp\left\{-\beta\,\frac{\left(b_V\theta\right)^2}{2k}\right\} f\left(\theta\right) d\theta \\
&= \sigma_\theta^{-1}\int \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left[\frac{\left(\theta - M_\theta\right)^2}{\Sigma_\theta^2} + 2\beta K_\theta\right]\right\} d\theta \\
&= \sigma_\theta^{-1}\Sigma_\theta \exp\left\{-\beta K_\theta\right\}\int \frac{1}{\Sigma_\theta\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{\left(\theta - M_\theta\right)^2}{\Sigma_\theta^2}\right\} d\theta \\
&= \sigma_\theta^{-1}\Sigma_\theta \exp\left\{-\beta K_\theta\right\}
\end{aligned}
$$

and the agent's expected utility is

$$
\begin{aligned}
&E\left[U\left(S + b_V\left(\theta a^*\left(\theta\right) + \varepsilon\right) - \frac{k}{2}\left(a^*\left(\theta\right)\right)^2\right)\right] \\
&= \exp\left\{-\beta\left(S - \frac{\beta}{2} b_V^2\sigma_\varepsilon^2\right)\right\}\sigma_\theta^{-1}\Sigma_\theta \exp\left\{-\beta K_\theta\right\} \\
&= \exp\left\{-\beta\left(S - \frac{\beta}{2} b_V^2\sigma_\varepsilon^2 + K_\theta - \beta^{-1}\ln\left(\sigma_\theta^{-2}\Sigma_\theta^2\right)^{1/2}\right)\right\}
\end{aligned}
$$

Next, note that since

$$
\begin{aligned}
2\beta K_\theta &= -\sigma_\theta^{-2}\left(1 + \beta\,\frac{\left[b_V\sigma_\theta\right]^2}{k}\right)^{-1}\mu_\theta^2 + \sigma_\theta^{-2}\mu_\theta^2 \\
&= \sigma_\theta^{-2}\mu_\theta^2\left\{-\frac{1}{\left(1 + \beta\,\frac{\left[b_V\sigma_\theta\right]^2}{k}\right)} + 1\right\} = \sigma_\theta^{-2}\mu_\theta^2\left\{\frac{\beta\,\frac{\left[b_V\sigma_\theta\right]^2}{k}}{\left(1 + \beta\,\frac{\left[b_V\sigma_\theta\right]^2}{k}\right)}\right\} = \frac{\beta b_V^2\mu_\theta^2}{\left(k + \beta\left[b_V\sigma_\theta\right]^2\right)}
\end{aligned}
$$

it follows that

$$
K_\theta = \frac{\left\{b_V\mu_\theta\right\}^2}{2\left(k + \beta\left[b_V\sigma_\theta\right]^2\right)}
$$

Thus, the agent's certainty equivalent is

$$CE\left(S, b_V, a^*(\theta)\right) = \left(S - \frac{\beta}{2} b_V^2 \sigma_\varepsilon^2 + K_\theta\right) - \beta^{-1} \ln\left(\sigma_\theta^2 \Sigma_\theta^{-2}\right)^{-1/2}$$

$$= S - \frac{\beta}{2} b_V^2 \sigma_\varepsilon^2 + \frac{\{b_V \mu_\theta\}^2}{2\left(k + \beta [b_V \sigma_\theta]^2\right)} - \beta^{-1} \ln\left(1 + \beta \frac{[b_V \sigma_\theta]^2}{k}\right)^{-1/2}$$

$$= S - \frac{\beta}{2} b_V^2 \sigma_\varepsilon^2 + \frac{\{b_V \mu_\theta\}^2}{2\left(k + \beta [b_V \sigma_\theta]^2\right)} + \frac{\beta^{-1}}{2} \ln\left(1 + \beta \frac{[b_V \sigma_\theta]^2}{k}\right)$$

Appendix B2: Proof that $0<b_v^*<1$ with PDI and non-contractible effort.

First, show that the second order condition for a maximum is satisfied.

$$\frac{\partial J}{\partial b_V}\left(b_V \mid \sigma_\theta^2, \sigma_\varepsilon^2, \mu_\theta, k, \beta\right)$$

$$= -2\frac{\left(\sigma_\theta^2+\mu_\theta^2\right)}{k} - \beta\sigma_\varepsilon^2 + \mu_\theta^2 k \frac{\left(k+\beta b_V^2\sigma_\theta^2\right)^2 - b_V 2\left(k+\beta b_V^2\sigma_\theta^2\right)\beta 2 b_V \sigma_\theta^2}{\left(k+\beta b_V^2\sigma_\theta^2\right)^4} + \sigma_\theta^2 \frac{\left(k+\beta b_V^2\sigma_\theta^2\right) - b_V \beta 2 b_V \sigma_\theta^2}{\left(k+\beta b_V^2\sigma_\theta^2\right)^2}$$

$$= -2\frac{\left(\sigma_\theta^2+\mu_\theta^2\right)}{k} - \beta\sigma_\varepsilon^2 + \mu_\theta^2 k \frac{\left(k+\beta b_V^2\sigma_\theta^2\right) - 4\beta b_V^2\sigma_\theta^2}{\left(k+\beta b_V^2\sigma_\theta^2\right)^3} + \sigma_\theta^2 \frac{\left(k-\beta b_V^2\sigma_\theta^2\right)}{\left(k+\beta b_V^2\sigma_\theta^2\right)^2}$$

$$= -2\frac{\left(\sigma_\theta^2+\mu_\theta^2\right)}{k} - \beta\sigma_\varepsilon^2 + \mu_\theta^2 k \frac{\left(k-3\beta b_V^2\sigma_\theta^2\right)}{\left(k+\beta b_V^2\sigma_\theta^2\right)^3} + \sigma_\theta^2 \frac{\left(k-\beta b_V^2\sigma_\theta^2\right)}{\left(k+\beta b_V^2\sigma_\theta^2\right)^2}$$

$$= \frac{1}{k\left(k+\beta b_V^2\sigma_\theta^2\right)^3}\left[-2\left(\sigma_\theta^2+\mu_\theta^2\right)\left(k+\beta b_V^2\sigma_\theta^2\right)^3 + \mu_\theta^2 k^2\left(k-3\beta b_V^2\sigma_\theta^2\right) + k\sigma_\theta^2\left(k^2-\beta^2 b_V^4\sigma_\theta^4\right)\right] - \beta\sigma_\varepsilon^2$$

$$= \frac{1}{k\left(k+\beta b_V^2\sigma_\theta^2\right)^3}\left[-2\left(\sigma_\theta^2+\mu_\theta^2\right)\left(k+\beta b_V^2\sigma_\theta^2\right)^3 + \mu_\theta^2 k^3 - 3\beta b_V^2\sigma_\theta^2 + k^3\sigma_\theta^2 - \beta^2 b_V^4\sigma_\theta^4\right] - \beta\sigma_\varepsilon^2 < 0$$

So $\frac{\partial J}{\partial b_V}\left(b_V \mid \sigma_\theta^2, \sigma_\varepsilon^2, \mu_\theta, k, \beta\right)$ <0 for all values of b_V. Thus the principal's problem is globally convex.

Now show that $J\left(b_V = 0 \mid \sigma_\theta^2, \sigma_\varepsilon^2, \mu_\theta, k, \beta\right) > 0$:

$$J\left(b_V \mid \sigma_\varepsilon^2, \sigma_\theta^2, \mu_\theta, k, \beta\right) =$$

$$-\beta b_V \sigma_\varepsilon^2 + \mu_\theta^2 b_V \frac{k}{\left(k+\beta b_V^2\sigma_\theta^2\right)^2} + \frac{b_V\sigma_\theta^2}{\left(k+\beta b_V^2\sigma_\theta^2\right)} + \frac{\left(1-2b_V\right)\left[\sigma_\theta^2+\mu_\theta^2\right]}{k}$$

when $b_V = 0$, J = $\dfrac{\left[\sigma_\theta^2 + \mu_\theta^2\right]}{k}$ >0.

When $b_V = 1$:

$$J = -\beta\sigma_\varepsilon^2 \left(k\left(k + \beta\sigma_\theta^2\right)^2 \right) + \mu_\theta^2 k^2 + k\sigma_\theta^2\left(k + \beta\sigma_\theta^2\right) - \left(\sigma_\theta^2 + \mu_\theta^2\right)\left(k + \beta\sigma_\theta^2\right)^2$$

$$= -\beta\sigma_\varepsilon^2 \left(k\left(k + \beta\sigma_\theta^2\right)^2 \right) + \mu_\theta^2 k^2 + k^2\sigma_\theta^2 + k\beta\sigma_\theta^4$$

$$- k^2\sigma_\theta^2 - k^2\mu_\theta^2 - 2k\beta\sigma_\theta^4 - 2k\beta\sigma_\theta^2\mu_\theta^2 - \beta^2\sigma_\theta^4\left(\sigma_\theta^2 + \mu_\theta^2\right)$$

$$= -\beta\sigma_\varepsilon^2 \left(k\left(k + \beta\sigma_\theta^2\right)^2 \right) - k\beta\sigma_\theta^4 - 2k\beta\sigma_\theta^2\mu_\theta^2 - \beta^2\sigma_\theta^4\left(\sigma_\theta^2 + \mu_\theta^2\right) < 0$$

Since $\dfrac{\partial J}{\partial b_V}$ is everywhere negative, and J is positive at $b_V = 0$, negative at $b_V = 1$, it must be that $0 < b_V^* < 1$.

Appendix B3: Derivation of agent's ex ante certainty equivalent and principal's net profit with PDI and contractible effort.

The agent's expected utility is:

$$E\left[-e^{-\beta\left(S+b_a a+b_V V-c(a)\right)}\right]=-\int_{-\infty}^{\infty}E\left[e^{-\beta\left(S+b_a a+b_V V-c(a)\right)}\,|\,\theta\right]f(\theta)\,d\theta$$

$$=-e^{-\beta S}\left(\int_{-\infty}^{\infty}E\left[e^{-\beta\left(\frac{(b_a+b_V\theta)^2}{2k}\right)}\,|\,\theta\right]f(\theta)\,d\theta\right)e^{\frac{(\beta b_V)^2}{2}\sigma_\varepsilon^2}$$

where effort is anticipated to be chosen optimally once the productivity parameter has been observed, that is, $a^*(\theta)=\frac{b_a+b_V\theta}{k}$. To evaluate the integral term in the parenthesis

$$\int_{-\infty}^{\infty}E\left[e^{-\beta\left(\frac{(b_a+b_V\theta)^2}{2k}\right)}\,|\,\theta\right]f(\theta)\,d\theta=\int_{-\infty}^{\infty}\left\{e^{-\beta\left(\frac{(b_a+b_V\theta)^2}{2k}\right)}\frac{1}{\sigma_\theta\sqrt{2\pi}}e^{-\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2}\right\}d\theta$$

we first focus on the terms in the exponents only:

$$-\beta\left(\frac{(b_a+b_V\theta)^2}{2k}\right)-\frac{1}{2}\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2$$

$$=\frac{-1}{2}\left[k^{-1}\beta\left(b_a^2+2b_ab_V\theta+b_V^2\theta^2\right)+\sigma_\theta^{-2}\left(\theta^2-2\theta\mu_\theta+\mu_\theta^2\right)\right]$$

$$=\frac{-1}{2}\left[\theta^2\left(k^{-1}\beta b_V^2+\sigma_\theta^{-2}\right)-2\theta\left(\sigma_\theta^{-2}\mu_\theta-k^{-1}\beta b_ab_V\right)+\left(k^{-1}\beta b_a^2+\sigma_\theta^{-2}\mu_\theta^2\right)\right]$$

$$=\frac{-1}{2}\left(\frac{1}{\left(k^{-1}\beta b_V^2+\sigma_\theta^{-2}\right)^{-1}}\left\{\theta^2-2\theta\frac{\left(\sigma_\theta^{-2}\mu_\theta-k^{-1}\beta b_ab_V\right)}{\left(k^{-1}\beta b_V^2+\sigma_\theta^{-2}\right)}\right\}+\left[k^{-1}\beta b_a^2+\sigma_\theta^{-2}\mu_\theta^2\right]\right)$$

$$=\frac{-1}{2}\left(\frac{1}{\Sigma^2}\left\{\theta^2-2\theta M\right\}+\left[k^{-1}\beta b_a^2+\sigma_\theta^{-2}\mu_\theta^2\right]\right)$$

where

$$
\begin{aligned}
M &= \left(\sigma_\theta^{-2}\mu_\theta - k^{-1}\beta b_a b_V\right)\Sigma^2 \\
\Sigma^2 &= \left(k^{-1}\beta b_V^2 + \sigma_\theta^{-2}\right)^{-1} = \sigma_\theta^2\left(1 + \beta\frac{\left[b_V\sigma_\theta\right]^2}{k}\right)^{-1} > 0
\end{aligned}
$$

Adding and subtracting the same M^2 –term we get

$$
\begin{aligned}
&-\beta\left(\frac{\left(b_a + b_V\theta\right)^2}{2k}\right) - \frac{1}{2}\left(\frac{\theta - \mu_\theta}{\sigma_\theta}\right)^2 \\
&= \frac{-1}{2}\left(\frac{1}{\Sigma^2}\left\{\theta^2 - 2\theta M\right\} + \left[k^{-1}\beta b_a^2 + \sigma_\theta^{-2}\mu_\theta^2\right]\right) \\
&= \frac{-1}{2}\left(\frac{1}{\Sigma^2}\left\{\theta^2 - 2\theta M + M^2 - M^2\right\} + \left[k^{-1}\beta b_a^2 + \sigma_\theta^{-2}\mu_\theta^2\right]\right) \\
&= \frac{-1}{2}\left(\frac{\left\{\theta - M\right\}^2}{\Sigma^2} - \Sigma^{-2}M^2 + \left[k^{-1}\beta b_a^2 + \sigma_\theta^{-2}\mu_\theta^2\right]\right) \\
&= \frac{-1}{2}\left(\frac{\left\{\theta - M\right\}^2}{\Sigma^2} + 2\beta\kappa\right)
\end{aligned}
$$

where

$$
\begin{aligned}
2\beta\kappa &= -\Sigma^{-2}M^2 + \left[k^{-1}\beta b_a^2 + \sigma_\theta^{-2}\mu_\theta^2\right] \\
&= -\Sigma^{-2}\left\{\left(\sigma_\theta^{-2}\mu_\theta - k^{-1}\beta b_a b_V\right)\Sigma^2\right\}^2 + \left[k^{-1}\beta b_a^2 + \sigma_\theta^{-2}\mu_\theta^2\right] \\
&= -\Sigma^2\left(\sigma_\theta^{-2}\mu_\theta - k^{-1}\beta b_a b_V\right)^2 + \left[k^{-1}\beta b_a^2 + \sigma_\theta^{-2}\mu_\theta^2\right] \\
&= \Sigma^2\left\{-\left(\sigma_\theta^{-2}\mu_\theta - k^{-1}\beta b_a b_V\right)^2 + \Sigma^{-2}\left[k^{-1}\beta b_a^2 + \sigma_\theta^{-2}\mu_\theta^2\right]\right\} \\
&= \Sigma^2\left\{-\left(\sigma_\theta^{-2}\mu_\theta - k^{-1}\beta b_a b_V\right)^2 + \left(k^{-1}\beta b_V^2 + \sigma_\theta^{-2}\right)\left[k^{-1}\beta b_a^2 + \sigma_\theta^{-2}\mu_\theta^2\right]\right\}
\end{aligned}
$$

or

$$
\begin{aligned}
2\beta\kappa\Sigma^{-2} &= -\left(\sigma_\theta^{-2}\mu_\theta - k^{-1}\beta b_a b_V\right)^2 + \left(k^{-1}\beta b_V^2 + \sigma_\theta^{-2}\right)\left[k^{-1}\beta b_a^2 + \sigma_\theta^{-2}\mu_\theta^2\right] \\
&= -\left(\sigma_\theta^{-2}\mu_\theta\right)^2 + 2\left(\sigma_\theta^{-2}\mu_\theta k^{-1}\beta b_a b_V\right) - \left(k^{-1}\beta b_a b_V\right)^2 + \left(k^{-1}\beta b_V^2\right)\left[k^{-1}\beta b_a^2\right] + \left(k^{-1}\beta b_V^2\right)\left[\sigma_\theta^{-2}\mu_\theta^2\right] + \sigma_\theta^{-2}\left[k^{-1}\beta b_a^2\right] + \sigma_\theta^{-2}\left[\sigma_\theta^{-2}\mu_\theta^2\right] \\
&= +2\left(\sigma_\theta^{-2}\mu_\theta k^{-1}\beta b_a b_V\right) + \left(k^{-1}\beta b_V^2\right)\left[\sigma_\theta^{-2}\mu_\theta^2\right] + \sigma_\theta^{-2}\left[k^{-1}\beta b_a^2\right] \\
&= \sigma_\theta^{-2}k^{-1}\beta\left\{2\mu_\theta b_a b_V + b_V^2\mu_\theta^2 + b_a^2\right\} \\
&= \sigma_\theta^{-2}k^{-1}\beta\left\{\mu_\theta b_V + b_a\right\}^2
\end{aligned}
$$

Note that

$$\kappa = \sigma_\theta^{-2} k^{-1}\{\mu_\theta b_V + b_a\}^2 2^{-1}\Sigma^2 = \frac{\sigma_\theta^{-2} k^{-1}\{\mu_\theta b_V + b_a\}^2}{2\left(k^{-1}\beta b_V^2 + \sigma_\theta^{-2}\right)} = \frac{\{\mu_\theta b_V + b_a\}^2}{2\left(k + \beta b_V^2 \sigma_\theta^2\right)}$$

is a constant that does not depend on θ .

The integral in the agent's expected utility can now be written as

$$\int_{-\infty}^{\infty} e^{-\beta\left(\frac{(b_a + b_V\theta)^2}{2k}\right)} \left\{\frac{1}{\sigma_\theta\sqrt{2\pi}} e^{-\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2}\right\} d\theta$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma_\theta\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\{\theta-M\}^2}{\Sigma^2}+2\beta\kappa\right)} d\theta = e^{-\beta\kappa}\int_{-\infty}^{\infty} \frac{1}{\sigma_\theta\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\{\theta-M\}^2}{\Sigma^2}} d\theta$$

$$= \frac{\Sigma}{\sigma_\theta} e^{-\beta\kappa}\int_{-\infty}^{\infty} \frac{1}{\Sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\{\theta-M\}^2}{\Sigma^2}} d\theta = \sigma_\theta^{-1}\Sigma e^{-\beta\kappa}$$

where the last equality follows from recognizing that the integral of any (normal) probability distribution function is 1. At the time of contracting, the agent's expected utility is therefore

$$E\left[-e^{-\beta(S + b_a a + b_V V - c(a))}\right] = -e^{-\beta S}\left(\sigma_\theta^{-1}\Sigma e^{-\beta\kappa}\right) e^{\frac{(\beta b_V)^2}{2}\sigma_\varepsilon^2}$$

Consequently, the agent's certainty equivalent at the time of contracting is

$$
\begin{aligned}
&CE\left(S + b_a a + b_V V, a(\theta) = \frac{b_a + b_V\theta}{k}\right) \\
&= S - \beta^{-1}\log\left(\sigma_\theta^{-1}\Sigma\right) + \beta\kappa - \beta\,\frac{b_V^2}{2}\sigma_\varepsilon^2 \\
&= S - \beta^{-1}\left\{\log\left(\sigma_\theta^{-1}\Sigma\right)\right\} + \kappa - \beta\,\frac{b_V^2}{2}\sigma_\varepsilon^2 \\
&= S - \beta^{-1}\left\{\log\left(\left\{\sigma_\theta^2\Sigma^{-2}\right\}^{-1/2}\right)\right\} + \kappa - \beta\,\frac{b_V^2}{2}\sigma_\varepsilon^2 \\
&= S + \frac{\beta^{-1}}{2}\log\left(\sigma_\theta^2\Sigma^{-2}\right) + \kappa - \beta\,\frac{b_V^2}{2}\sigma_\varepsilon^2
\end{aligned}
$$

Since the agent's participation constraint,

$$
W \le CE\left(S + b_a a + b_V V, a(\theta) = \frac{b_a + b_V\theta}{k}\right)
$$

is binding it follows that

$$
\begin{aligned}
&-S = \frac{\beta^{-1}}{2}\log\left(\sigma_\theta^2\Sigma^{-2}\right) + \kappa - \beta\,\frac{b_V^2}{2}\sigma_\varepsilon^2 - W \\
&= \frac{\beta^{-1}}{2}\log\left(1 + \beta\,\frac{\left[b_V\sigma_\theta\right]^2}{k}\right) + \frac{\left\{\mu_\theta b_V + b_a\right\}^2}{2\left(k + \beta b_V^2\sigma_\theta^2\right)} - \beta\,\frac{b_V^2}{2}\sigma_\varepsilon^2 - W
\end{aligned}
$$

Since

$$
\frac{\partial}{\partial b_a}\left\{\log\left(\sigma_\theta^2\Sigma^{-2}\right)\right\} = 0
$$

$$
\frac{\partial}{\partial b_V}\left\{\log\left(\sigma_\theta^2\Sigma^{-2}\right)\right\} = \frac{\frac{\partial}{\partial b_V}\left(\sigma_\theta^2\Sigma^{-2}\right)}{\left(\sigma_\theta^2\Sigma^{-2}\right)} = \frac{k^{-1}\beta\, 2b_V\sigma_\theta^2}{\left(k^{-1}\beta b_V^2\sigma_\theta^2 + 1\right)} = \frac{\beta\, 2b_V\sigma_\theta^2}{\left(k + \beta b_V^2\sigma_\theta^2\right)}
$$

$$
\frac{\partial}{\partial\sigma_\theta^2}\left\{\log\left(\sigma_\theta^2\Sigma^{-2}\right)\right\} = \frac{\beta b_V^2}{\left(k + \beta b_V^2\sigma_\theta^2\right)}
$$

and

$$\frac{\partial\kappa}{\partial b_a} = \frac{\{\mu_\theta b_V + b_a\}}{\left(k + \beta b_V^2 \sigma_\theta^2\right)}$$

$$\frac{\partial\kappa}{\partial b_V} = \frac{\mu_\theta \{\mu_\theta b_V + b_a\}\left(k + \beta b_V^2 \sigma_\theta^2\right) - \{\mu_\theta b_V + b_a\}^2 \left(\beta b_V \sigma_\theta^2\right)}{\left(k + \beta b_V^2 \sigma_\theta^2\right)^2}$$

$$\frac{\partial\kappa}{\partial \sigma_\theta^2} = -\frac{\{\mu_\theta b_V + b_a\}^2}{2\left(k + \beta b_V^2 \sigma_\theta^2\right)^2} \beta b_V^2$$

it follows that

$$-\frac{\partial S}{\partial b_a} = \frac{\{\mu_\theta b_V + b_a\}}{\left(k + \beta b_V^2 \sigma_\theta^2\right)}$$

and

$$-\frac{\partial S}{\partial b_V} = \frac{\beta^{-1}}{2} \frac{\beta\, 2 b_V \sigma_\theta^2}{\left(k + \beta b_V^2 \sigma_\theta^2\right)} + \frac{\mu_\theta \{\mu_\theta b_V + b_a\}\left(k + \beta b_V^2 \sigma_\theta^2\right) - \{\mu_\theta b_V + b_a\}^2 \left(\beta b_V \sigma_\theta^2\right)}{\left(k + \beta b_V^2 \sigma_\theta^2\right)^2} - \beta b_V \sigma_\varepsilon^2$$

$$= \frac{b_V \sigma_\theta^2 + \mu_\theta \{\mu_\theta b_V + b_a\}}{\left(k + \beta b_V^2 \sigma_\theta^2\right)} - \frac{\{\mu_\theta b_V + b_a\}^2 \left(\beta b_V \sigma_\theta^2\right)}{\left(k + \beta b_V^2 \sigma_\theta^2\right)^2} - \beta b_V \sigma_\varepsilon^2$$

and

$$-\frac{\partial S}{\partial \sigma_\theta^2} = \frac{\beta^{-1}}{2} \frac{\beta b_V^2}{\left(k + \beta b_V^2 \sigma_\theta^2\right)} - \frac{\{\mu_\theta b_V + b_a\}^2}{2\left(k + \beta b_V^2 \sigma_\theta^2\right)^2} \beta b_V^2$$

$$= \frac{\beta^{-1}}{2} \log\left(1 + \beta \frac{\left[b_V \sigma_\theta\right]^2}{k}\right) + \frac{\{\mu_\theta b_V + b_a\}^2}{2\left(k + \beta b_V^2 \sigma_\theta^2\right)} - \beta \frac{b_V^2}{2} \sigma_\varepsilon^2 - W$$

= = =

To evaluate the objective function of the principal, apply the derived decision rule of the agent to rewrite the principal's gross profits as

$$E[V] = E[\theta a^*(\theta) + \varepsilon] = E\left[\theta\left\{\frac{b_a + b_V\theta}{k}\right\}\right] + E[\varepsilon]$$
$$= \frac{b_a}{k}E[\theta] + \frac{b_V}{k}E[\theta^2] = \frac{b_a}{k}\mu_\theta + \frac{b_V}{k}\left(VAR[\theta] + (E[\theta])^2\right)$$
$$= k^{-1}\left\{b_a\mu_\theta + b_V\left(\sigma_\theta^2 + \mu_\theta^2\right)\right\}$$

The principal's profit net of compensation to the agent is

$$E[V] - \left\{S + E\left[b_a a^*(\theta) + b_V V\right]\right\} = (1 - b_V)E[V] - b_a E\left[a^*(\theta)\right] - S$$
$$= (1 - b_V)k^{-1}\left\{b_a\mu_\theta + b_V\left(\sigma_\theta^2 + \mu_\theta^2\right)\right\} - b_a\left(\frac{b_a + b_V\mu_\theta}{k}\right) - S$$

If the optimal solution is interior then it is determined by the two first order conditions for choice of b_a:

$$0 = (1 - b_V)k^{-1}\mu_\theta - \left(\frac{2b_a + b_V\mu_\theta}{k}\right) - \frac{\partial S}{\partial b_a}$$

And choice of b_V:

$$0 = k^{-1}\left\{-b_a\mu_\theta + (1 - 2b_V)\left(\sigma_\theta^2 + \mu_\theta^2\right)\right\} - b_a\left(\frac{\mu_\theta}{k}\right) - \frac{\partial S}{\partial b_V}$$

These two equations can be written as

$$0 = (1 - 2b_V)\mu_\theta - 2b_a - k\frac{\partial S}{\partial b_a}$$

and

$$0 = (1 - 2b_V)\left(\sigma_\theta^2 + \mu_\theta^2\right) - 2b_a\mu_\theta - k\frac{\partial S}{\partial b_V}$$

By substitution from above, we find that

$$0 = (1 - 2b_V)\mu_\theta - 2b_a + k\frac{\left\{\mu_\theta b_V + b_a\right\}}{\left(k + \beta b_V^2\sigma_\theta^2\right)}$$

and

$$0 = \left(1 - 2b_V\right)\left(\sigma_\theta^2 + \mu_\theta^2\right) - 2b_a\mu_\theta + k\left\{\frac{b_V\sigma_\theta^2 + \mu_\theta\left\{\mu_\theta b_V + b_a\right\}}{\left(k + \beta b_V^2\sigma_\theta^2\right)} - \frac{\left\{\mu_\theta b_V + b_a\right\}^2\left(\beta b_V\sigma_\theta^2\right)}{\left(k + \beta b_V^2\sigma_\theta^2\right)^2} - \beta b_V\sigma_\varepsilon^2\right\}$$

Appendix B4: Proofs that neither b nor b_1 can be zero.

THEOREM B4.1:

Assume that effort is observable and contractible and $\sigma_\theta^2 > 0$. Then the principal will use firm value for contracting, that is, $b_V \neq 0$.

PROOF OF THEOREM B4.1:

Proof by contradiction. Assume that $(b_V, b_a) = (0, b_a)$. Then the optimal effort would be $a(\theta) = b_a / k$. The first order conditions would reduce to

$$0 = \mu_\theta - 2b_a + k\frac{b_a}{k} \qquad [1.1]$$

and

$$0 = (\sigma_\theta^2 + \mu_\theta^2) - 2b_a\mu_\theta + k\frac{\mu_\theta b_a}{k} \qquad [2.1]$$

Since [1] implies that $b_a = \mu_\theta$ while [2] implies that

$$b_a = \frac{(\sigma_\theta^2 + \mu_\theta^2)}{\mu_\theta} = \frac{\sigma_\theta^2}{\mu_\theta} + \mu_\theta$$

the contradiction arises whenever $\sigma_\theta^2 > 0$. Note that additive risk, σ_ε^2, plays no role for this result. Hence, the optimal contract does put some weight will be put on firm value.

THEOREM B4.2:

Assume that effort is observable and contractible and that $\mu_\theta = 0$ then $b_a = 0$

PROOF OF THEOREM B4.3:

When $\mu_\theta = 0$, the first order conditions reduce to

$$0 = -2b_a + k\frac{b_a}{\left(k + \beta b_V^2 \sigma_\theta^2\right)} \qquad [1.2]$$

and

$$0 = \left(1 - 2b_V\right)\sigma_\theta^2 + k\left\{\frac{b_V\sigma_\theta^2}{\left(k + \beta b_V^2\sigma_\theta^2\right)} - \frac{\{b_a\}^2\left(\beta b_V\sigma_\theta^2\right)}{\left(k + \beta b_V^2\sigma_\theta^2\right)^2} - \beta b_V \sigma_\varepsilon^2\right\} \qquad [1.3]$$

It follows immediately from [1.2] that $b_a = 0$. This implies that [1.3] reduces to

$$0 = \left(1 - 2b_V\right)\sigma_\theta^2 + k\frac{b_V\sigma_\theta^2}{\left(k + \beta b_V^2\sigma_\theta^2\right)} - k\beta b_V\sigma_\varepsilon^2$$

which is a third order polynomial

$$0 = b_V^3\left[k\beta^2\sigma_\varepsilon^2\sigma_\theta^2 - 2\beta\sigma_\theta^4\right] + b_V^2\left[\beta\sigma_\theta^4\right] + b_V\left[-3k\sigma_\theta^2 - k^2\beta\sigma_\varepsilon^2\right] + \left[k\sigma_\theta^2\right]$$

that identifies (at least) one solution to b_V.

THEOREM B4.3:

Assume that $\mu_\theta \neq 0$. When effort is observable and contractible, the principal uses firm value for contracting, that is, $b_a \neq 0$, almost surely (i.e., except on a set of parameters of measure zero).

PROOF OF THEOREM B4.3:

Proof by contradiction. Assume that $(b_V, b_a) = (b, 0)$. Then the optimal effort would be $a(\theta) = b_V \theta / k$. The first order conditions would reduce to

$$0 = (1 - 2b)\mu_\theta + k \frac{\mu_\theta b}{(\beta b^2 \sigma_\theta^2 + k)} \qquad [1]$$

and

$$0 = (1 - 2b)(\sigma_\theta^2 + \mu_\theta^2) + k \left\{ \frac{b\sigma_\theta^2 + \mu_\theta \{\mu_\theta b\}}{(\beta b^2 \sigma_\theta^2 + k)} - \frac{\{\mu_\theta b\}^2 (\beta b \sigma_\theta^2)}{(\beta b^2 \sigma_\theta^2 + k)^2} - \beta b \sigma_\varepsilon^2 \right\} \qquad [2]$$

These equations define the roots of two polynomials in b: the first is a third order polynomial and the second is a fifth order polynomial where only the latter depends on σ_ε^2. Hence the polynomials only coincide “rarely”. To make this statement more precise, note that [1] implies that

$$(2b - 1)\mu_\theta = k \frac{\mu_\theta b}{(\beta b^2 \sigma_\theta^2 + k)}$$

or $b > .5$ and hence

$$\frac{\{(2b - 1)\mu_\theta\}^2}{k} = k \frac{\{\mu_\theta b\}^2}{(\beta b^2 \sigma_\theta^2 + k)^2}$$

Substituting these two equations into [2], we find that [2] reduces to the third order polynomial:

$$0 = (1 - 2b)(\sigma_\theta^2 + \mu_\theta^2) + (2b - 1)\sigma_\theta^2 + (2b - 1)\mu_\theta - \frac{\{(2b - 1)\mu_\theta\}^2}{k} (\beta b \sigma_\theta^2) - k\beta b \sigma_\varepsilon^2$$

These two reduced first order conditions are non-trivial and do not coincide since they have different slope almost everywhere (since $b \neq 0$ according to Theorem B4.1).