Moral Hazard in Innovation: The Relationship between Risk Aversion and Performance Pay*

Sanxi Li†

Jianye Yan‡

Xundong Yin§

Abstract

We study a model when there exists both moral hazard in developing an innovative production method and moral hazard within a given production method. Unlike the classic result of Milgrom and Holmstrom (1987), we find that the relationships between the power of incentive, risk, and the agent's risk attitude are not monotone. Our results are consistent with the empirical findings on these relationships.

Key words: *Risk aversion*; *The power of incentive*; *The production effort*; *The innovation effort*

JEL classification: D80; J33

*We thank David Martimort and Wilfried Sand-Zantman for their useful comments and advice.

†Toulouse School of Economics. Email: lisanxi@gmail.com.

‡Research Center for Applied Finance, School of Banking and Finance, University of International Business and Economics. Email: yanjianye@gmail.com.

§Toulouse School of Economics. Email: yinxundong@gmail.com.

1 Introduction

Standard principal-agent models (e.g. Holmstrom and Milgrom 1987) predict both a negative relationship between risk and performance pay and a negative relationship between agent's risk aversion attitude and performance pay. Yet empirical studies suggest that the two predictions are not true. In this paper, we develop a simple agency model to illustrate that the relationships, in general, are not monotonic, which is more consistant with the empirical results.

Motivating innovation is important in many incentive problems (Manso 2011). For example, shareholders may want to motivate CEO to adopt more innovative business strategies and CEO may also want to induce their employees to explore new method of production. The key assumption in our paper is that the agent can exert both innovation effort and production effort at the same time. It is this innovation effort that distinguishes our paper from Milgrom and Holmstrom's. More precisely, for a given production method, the agent can exert a production effort to increase the expected output. However, it does not change the risk itself involved in the production process. On the other hand, the agent can also exert an innovation effort to develop an innovative production method. Adopting the new method changes not only the expected output, but also the risk in the production process.

Innovation may either increase or decrease the risk in the production process. It could be the case that the old production method is not risky, because it has been tested and improved for a long time. The innovative method, however, could be very risky because the new method could bring much uncertainty in the production process. The relationship between innovation and risk can also be negative. An innovative approach may increase expected returns by rationalizing operations and streamlining processes. The risk in the production process then decreases rather than increases. In either case, we have little reason to assume that the risk remains unchanged after the innovation.

When the agent can exert the production effort but not the innovation effort, the standard trade-off between incentive and insurance applies. A higher power of incentive, although it induces more effort, it exposes the agent to a higher risk and therefore leads to a higher risk premium cost to the principal. As the risk in the production process and/or the agent's risk averse attitude increases, the risk-premium cost increases. The principal responds by reducing the power of incentive. This logic leads to the standard prediction of *negative* relationship between the power of incentive, risk, and the agent's risk-averse attitude.

Now consider that the agent can exert the innovation effort but not the production effort. The agent can either adopt the old production method or develop an innovative method by exerting an innovation effort. Assume it is beneficial for the principal to induce an innovation effort. The observed risk is then the one associated with the innovative method. As the innovative method becomes riskier, a risk-averse agent is more reluctant to do innovation. In order to induce the innovation effort, the principal must provide the agent with a higher power of incentive. As a result, there is a *positive* relationship between the power of incentive and risk. Moreover, when the innovative method is riskier than the old one, a more risk-averse agent will be more reluctant to do innovation. The principal, therefore, has to provide him a higher power of incentive to induce innovation effort. Therefore, there is a *positive* relationship between the power of incentive and the agent's risk-averse attitude. Similar logic leads to a *negative* relationship between the power of incentive and the agent's risk-averse attitude if the innovative method is less risky than the old one.

In the hybrid model in which the agent can exert both a production effort and innovation effort, the relationship between the power of incentive and risk,[1] the agent's risk-averse attitude is no longer monotone. We start the analysis by assuming the riskness of the new method and the old one are

[1]Here we refer to the risk under new technology because we always assume it is profitable to adopt innovative method.

the same. Now from this point, we increase the riskiness of the new method. The power of incentive will first decrease and then increase with risk. To see this, assume that the principal ignores the incentive problem of innovation and naively believes that the agent already has the innovative production method. To induce the production effort, the principal chooses an optimal power of incentive to balance the trade-off between incentive and insurance. The principal provides some incentive to induce production effort. But this power of incentive is enough to induce the agent to also do innovation if the innovative method is not very risky. Therefore, the principal can indeed ignore the incentive problem of innovation. In this case, the relationship between the power of incentive and risk is negative. However, when the innovative method becomes very risky, the power of incentive to induce the production effort is not enough to induce the innovation effort. The principal must take into account the incentive problem of innovation. In this case, the power of the incentive that induces innovation also induces a very high production effort. Therefore, the incentive problem in the production process can be ignored. Because it is the incentive problem of innovation that matters, the relationship between the power of incentive and risk is positive.

Similar logic also suggests that the power of incentive first decreases and then increases in the agent's risk-averse attitude if the innovative method is riskier than the old one. However, when the innovative method is less risky than the old one, the relationship between the power of incentive and the agent's risk-averse attitude is unambiguously negative. Indeed, whether it is the incentive of production or the incentive of innovation that matters, the power of incentive decreases with the agent's risk-averse attitude.

Throughout this paper, we assume whether the agent adopts new technology or not, and exert production effort or not, are not observable.

Related Literature. Prendergast (2002) argues that empirical studies provide weak support for the central prediction of the negative relationship between risks and the power of the contract. To explain this puzzle, he builds a model in which the principal can also monitor the agent's effort, on top of using an output-based incentive contract. In a stable environment, the principal has a good idea of what the agent should do, and so he therefore identifies the agent's task and monitors his effort. In an uncertain environment, although the principal can monitor the agent's effort, he has no idea on what the agent should focus his time. Therefore, an output-based contract should be used to induce the agent to exert his effort on the right things.

Since Prendergast's seminal work, other explanations have arisen. Baker and Jorgensen (2003) find that a positive relationship between uncertainty and incentive is possible, when the uncertainty is the one whose outcome changes the agent's optimal effort. The actual realization of uncertainty is the agent's valuable private information. The benefit of providing the agent a strong incentive is that it induces the agent to respond optimally to his private information. As the uncertainty increases, this benefit increases and may even outweigh the corresponding increased risk premium cost. The principal responds optimally by increasing the incentive strength.

Wright (2004) and Serfes (2005) argue that competition matches a less risk-averse agent with a principal of more risk. Therefore, the final outcome may be a positive relationship between risk and incentives. Ackerberg and Botticini (2002) address this endogenous matching problem empirically. They find that after controlling for biases resulting from endogenous matching, the insurance and incentive trade-off is an important determinant of contract choice.

Guo and Ou-Yang (2004) argue that the puzzle can be explained if the risk is endogenous. In their model, the agent can exert effort to control the risk of output. When the cost of controlling risk is relatively high compared to that of improving the mean, the agent may choose a high level of risk to reduce costs, and choose a high level of mean, given a high performance pay. As a result, the relationship between performance pay and risk is positive.

Finally, Budde and Krakel (2008) find that the puzzle can be explained by combing the risk aversion and limited liability in a standard principal-agent model. In the presence of limited liability, the principal needs to provide a higher power of incentive to induce a given level of effort if the performance signal is less informative about the agent's effort. Thus, a positive relationship between risk and performance pay may arise.

Unlike in the above literature, by introducing a new dimension of effort, i.e., "innovation", our models provide a simple and direct explanation of the conventional puzzle between risk aversion, risk and incentives without resorting to limited liability or the endogenous mathcing between agents and principals. Depending on the riskiness of the new method, the power of incentive could either decrease or increase with the new method's risk and the agent's risk-averseness. In some range of the parameters, the power of incentive could be irresponsive to the riskiness of the new method and the risk attitude of the agent, which in turn can fit the empirical findings well.

2 The Holmstrom-Milgrom Model: Moral Hazard With a Given Production Method

A risk neutral principal hires a risk-averse agent to produce output. The agent has CARA utility function with absolute risk averse coefficient r: $u(x) = \frac{1-\exp(-rx)}{r}$. In this section, we dispense with the problem of developing innovative production methods. With a given production method, the agent can exert costly effort to increase output. The output is:

$$y = \mu + e + \varepsilon,$$

where $\varepsilon \sim N(0, \sigma^2)$. μ and σ^2 represent the exogenous expected pay-off and the riskiness of the given production method respectively, while e is effort exerted by the agent. The effort cost is $c(e) = \frac{ke^2}{2}$.

Contract. The effort is not observable. The only observable and contractible variable is the output y. Assume that the principal uses a linear contract:

$$w = w_0 + \alpha y,$$

where w_0 is the fixed salary and α is the power of incentive.

Given the contract, the agent can choose e to maximize his certainty equivalent:

$$CE = w_0 + \alpha\mu + \alpha e - \frac{r\alpha^2\sigma^2}{2} - \frac{ke^2}{2}.$$

The first order condition with respect to e gives the following incentive compatible condition:

$$IC_1 : e = \alpha/k.$$

The participation constraint requires

$$IR_1 : CE \geq 0.$$

The participation constraint should be binding. Otherwise, the principal may be better off by reducing the fixed salary w_0. By adjusting the fixed salary, the principal extracts all the surplus. His problem becomes choosing the power of incentive α in order to maximize the total surplus:

$$\max_{\alpha} \mu + e - \frac{ke^2}{2} - \frac{r\alpha^2\sigma^2}{2}$$
$$s.t. IC_1.$$

Solving this problem, one can get the optimal power of incentive:

$$\alpha_1^* = \frac{1}{1 + kr\sigma^2}.$$

Proposition 1 *(Holmstrom and Milgrom 1987) There is both a negative relationship between risk and performance pay and a negative relationship between the agent's risk aversion attitude and performance pay.*

For the principal, there is a trade off between inducing effort and providing insurance. The benefit of a higher power of incentive is that it induces a higher effort. The cost is that the principal needs to pay the agent a higher risk premium. As the risk or the agent's risk averse attitude increases, a higher power of incentive becomes more costly. Accordingly, the principal will reduce the power of incentive.

3 Moral Hazard in Innovation

In this section, we dispense with the moral hazard problem within a given production method. Instead, we focus on the incentive problem of developing an innovative method, in which the agent has two choices. The first choice is to adopt the standard production method, which generates an output:

$$y = \mu_1 + \varepsilon_1,$$

where $\varepsilon_1 \sim N(0, \sigma_1^2)$. μ_1 and σ_1^2 represent the expected pay-off and the riskiness of the standard method respectively. Notice here that we assume the agent can do nothing to increase the output once the production method is given.

Alternatively, by exerting an effort which costs him ψ, the agent will be able to develop an innovative method. Adopting the innovative method generates an output:

$$y = \mu_2 + \varepsilon_2,$$

where $\varepsilon_2 \sim N(0, \sigma_2^2)$. μ_2 and σ_2^2 represent the expected pay-off and the riskiness of the innovative method respectively.

Assumption:

1) $\mu_2 - \mu_1 > \psi$;

2) $\mu_2 - \mu_1 - r(\sigma_2^2 - \sigma_1^2) > 0$;

3) $\mu_2 - \mu_1 - \frac{r(\sigma_2^2 - \sigma_1^2)}{2} > \psi$.

The first assumption ensures that the innovative method is desirable from the point view of the risk-neutral principal. The second assumption ensures that the agent is more willing to exert effort if he is given a stronger incentive, as we will see later. The third assumption says that it is worthwhile exerting an innovative effort from the point view of the risk-averse agent.

Contract. Neither the effort nor the production method is observable. The only observable and contractible variable is the output y. Again, assume that the principal uses the linear contract $w = w_0 + \alpha y$.

If the agent exerts no effort, he will only be able to use the standard production method. Thus, he receives a certainty equivalent:

$$CE_1 = w_0 + \alpha\mu_1 - \frac{r\alpha^2\sigma_1^2}{2}.$$

Alternatively, by exerting effort ψ, he will be able to obtain an innovative method. Adopting the innovative method gives him a certainty equivalent:

$$CE_2 = w_0 + \alpha\mu_2 - \frac{r\alpha^2\sigma_2^2}{2}.$$

The incentive compatible constraint requires $CE_2 - CE_1 \geq \psi$. That is,

$$IC_2 : \alpha\left(\mu_2 - \mu_1\right) - \frac{r\alpha^2\left(\sigma_2^2 - \sigma_1^2\right)}{2} \geq \psi. \tag{1}$$

The individual rationality condition requires $CE_2 - \psi \geq 0:$

$$IR_2 : w_0 + \alpha\mu_2 - \frac{\alpha^2\sigma_2^2}{2} - \psi \geq 0. \tag{2}$$

Again, IR_2 is binding, therefore, the principal gets all the surplus. The principal's problem is then:

$$\max_{\alpha} \mu_2 - \frac{r\alpha^2\sigma_2^2}{2} - \psi$$
$$s.t. IC_2.$$

Obviously, IC_2 should also be binding. Otherwise, the principal will do better by reducing the power of incentive. While the fixed salary, w_0, is determined by the binding IR_2 (condition (2)), the power of incentive α is totally determined by the binding IC_2 (condition (1)).

It is worthwhile noticing that $\frac{\partial}{\partial\alpha}\left(CE_2 - CE_1\right) = \mu_2 - \mu_1 - \alpha r\left(\sigma_2^2 - \sigma_1^2\right)$, which is positive on interval $[0, 1]$ by assumption 2). Hence, our assumption ensures that the agent is more willing to exert effort if he is given a higher power of incentive. Furthermore, $CE_2 - CE_1|_{\alpha=0} = 0 < \psi$, $CE_2 - CE_1|_{\alpha=1} = \left(\mu_2 - \mu_1\right) - \frac{r\left(\sigma_2^2 - \sigma_1^2\right)}{2} > \psi$ (by assumption 3)), hence there is a unique solution $\alpha_2^* \in (0, 1)$ such that $CE_2 - CE_1|_{\alpha=\alpha_2^*} = \psi$. Solving it, we obtain

$$\alpha_2^* = \frac{\left(\mu_2 - \mu_1\right) - \sqrt{\left(\mu_2 - \mu_1\right)^2 - 2r\left(\sigma_2^2 - \sigma_1^2\right)\psi}}{r\left(\sigma_2^2 - \sigma_1^2\right)}. \tag{3}$$

By applying the implicit function theorem to the binding IC_2 condition, one can easily obtain the relationship between the power of incentive, risks, and the agent's risk attitude.

Proposition 2 *1)* $\frac{\partial \alpha_2^*}{\partial \sigma_1^2} < 0$; $\frac{\partial \alpha_2^*}{\partial \sigma_2^2} > 0$.
2) $\frac{\partial \alpha_2^*}{\partial r} > 0$ *iff* $\sigma_2^2 > \sigma_1^2$.

Although the power of incentive decreases with the risk of the standard production method, it increases with the risk of the innovative production method. Since the agent is risk-averse, a production method becomes less attractive if it becomes riskier. In order to induce effort, the principal now only needs to provide a lower power of incentive if the standard method becomes less attractive. On the contrary, she must provide a higher power of incentive if the innovative method becomes less attractive. Given the fact that the principal induces the adoption of the innovative method, our model suggests a positive relationship between the observed risk and incentive.

If the innovative method is riskier than the standard one, the power of incentive increases with the agent's risk-averse attitude. As the agent becomes more risk-averse, both methods become less attractive. However, as the innovative method is riskier, it becomes relatively less favorable. As a result, the principal must provide a higher incentive to encourage the agent to adopt the innovative method.

In contrast to the standard principal-agent models, this model predicts both a *positive* relationship between risk and performance pay and a *positive* relationship between the agent's risk aversion attitude and performance pay.

Now we derive the condition under which the principal wants to induce innovative effort. By inducing no innovative effort, the principal can simply offer the agent a fixed wage, which is equal to zero. The principal's expected pay-off is μ_1. By inducing innovative effort, the principal gets the expected pay-off $\mu_2 - \psi - \frac{r\alpha_2^{*2}\sigma_2^2}{2}$. She gets all the expected revenue from the innovative method net of the agent's effort cost and the risk premium. The principal induces innovative effort if and only if $\mu_2 - \psi - \frac{r\alpha_2^{*2}\sigma_2^2}{2} > \mu_1$. A sufficient condition occurs if the difference between the expected payoff of the two methods is large enough:

$$\mu_2 - \mu_1 \geq \psi + \frac{r\sigma_2^2}{2}. \tag{4}$$

4 The Hybrid Model

In this section, we assume that there are both a moral hazard in developing an innovative method and a moral hazard within a given production method. The agent can choose whether to exert the innovation effort ψ to develop an innovative production method or not. Moreover, he can also exert effort to improve the output when he adopts method i:

$$y = \mu_i + e + \varepsilon_i.$$

Given the contract $w = w_0 + \alpha y$, if the agent chooses to adopt the standard method, he will choose production effort e to maximize his certainty equivalent,

$$\max_e CE_1 = w_0 + \alpha\mu_1 + \alpha e - \frac{r\alpha^2\sigma_1^2}{2} - \frac{ke^2}{2},$$

which gives $e = \alpha/k$.

Now suppose the agent chooses to exert innovation effort cost ψ. After the innovative method has been developed, the agent also chooses production effort e to maximize his certainty equivalent:

$$\max_e CE_2 = w_0 + \alpha\mu_2 + \alpha e - \frac{r\alpha^2\sigma_2^2}{2} - \frac{ke^2}{2},$$

which gives $e = \alpha/k$. The incentive compatible condition for the agent to exert innovation cost ψ is $CE_2 - CE_1 \geq \psi$. Notice that for a given power of incentive α, the production effort e remains the same no matter whether the agent exerts an innovation effort or not. Hence, the incentive compatible condition for innovation is exactly the same as that in previous section, and is given by inequality IC_2.

Again, the principal extracts all the surplus by adjusting the fixed salary. His problem is then:

$$\max \mu_2 + e - \frac{r\alpha^2\sigma_2^2}{2} - \frac{ke^2}{2} - \psi$$
$$s.t. IC_1, IC_2.$$

Lemma 1 *The optimal power of incentive is given by:*

$$\alpha^* = \begin{cases} \alpha_1^*, \text{ if } \alpha_1^* \geq \alpha_2^* \\ \alpha_2^*, \text{ if } \alpha_2^* \geq \alpha_1^* \end{cases},$$

where $\alpha_1^* = \frac{1}{1+kr\sigma_2^2}$, α_2^* *is the same as (3).*

Proof. First, ignore the constraint IC_2. Then the optimal power of incentive is α_1^*. If $\alpha_1^* \geq \alpha_2^*$, then IC_2 is slack. Hence $\alpha^* = \alpha_1^*$. However, when $\alpha_1^* < \alpha_2^*$, if the principal still sets $\alpha = \alpha_1^*$, then the agent will not exert the innovation effort. To induce the innovation effort, IC_2 must be satisfied. Notice that the principal's payoff decreases in α when $\alpha > \alpha_1^*$. Hence, IC_2 is binding and determines $\alpha^* = \alpha_2^*$. ■

Proposition 3 *1) The power of incentive first decreases and then increases in* σ_2^2.

2) If $\sigma_2^2 > \sigma_1^2$*, then the power of incentive first decreases and then increases in* r.

3) If $\sigma_2^2 < \sigma_1^2$*, then the power of incentive is monotonically decreasing in* r.

Proof. Since α_1^* is decreasing in σ_2 and α_2^* is increasing in σ_2, there exists a threshold value $\widehat{\sigma}_2$, such that $\alpha_1^* < \alpha_2^*$ iff $\sigma_2 > \widehat{\sigma}_2$. Hence,

$$\alpha^* = \begin{cases} \alpha_1^*, \text{ if } \sigma_2 < \widehat{\sigma}_2 \\ \alpha_2^*, \text{ if } \sigma_2 > \widehat{\sigma}_2 \end{cases}.$$

As a result, α^* first decreases and then increases in σ_2.

The proof of 2) is similar. The proof of 3) is obvious by noticing both α_1^* and α_2^* are decreasing in r. ■

The intuition is as follows: suppose the principal naively takes the innovative method as given and ignores the incentive problem in innovation. Because there is also a moral hazard in the production process, the principal must provide incentives to induce the production effort. When the innovative method is not risky, the principal provides a high power of incentive. However, this incentive itself is enough to encourage the agent to do innovation; hence, the incentive problem in innovation can be ignored. In this case, the standard trade-off between incentive and insurance applies as there is negative relationship between incitation and risk. As the innovative method becomes very risky, the incentive of inducing the production effort is not enough to induce the agent to do innovation. Hence, the incentive compatible constraint of innovation becomes relevant and determines the power of incentive. In this case, our intuition from the previous section applies and there is a positive relationship between the observed risk σ_2^2 and the power of incentive. The explanation of the relationship between incentive and risk attitude is similar if $\sigma_2^2 > \sigma_1^2$. If $\sigma_2^2 < \sigma_1^2$, the the power of incentive is decreasing in r no matter whether the incentive compatible constraint of innovation is binding or not.

Again, we need to derive conditions under which the principal induces innovative effort. For $\sigma_2 \leq \widehat{\sigma}_2$, inducing innovative effort has no cost for the principal; therefore he will do so. However, for large $\sigma_2 > \widehat{\sigma}_2$, inducing innovative effort is costly for the principal and she may not want to do so. The principal will induce innovation effort if and only if

$$\mu_1 + \alpha_1^*/k - \frac{\alpha_1^{*2}}{2k} - \frac{r\alpha_1^*\sigma_1^2}{2} \leq \mu_2 + \alpha_2^*/k - \frac{\alpha_2^{*2}}{2k} - \psi - \frac{r\alpha_2^*\sigma_2^2}{2},$$

where the LHS is the principal's expected pay-off by inducing no innovation effort, and the RHS is the principal's expected pay-off by inducing innovative effort. The above inequality will hold if the innovative method significantly increases the exogenous expected pay-off, i.e., if $\mu_2 - \mu_1$ is large enough. One can easily check that condition (4) is sufficient to ensure the principal to indeed induce innovation effort.

5 Conclusion

When the risks are associated with the particular production method adopted by the agent, distinguishing the production effort and innovation effort is extremely important to understand the relationship between the power of incentive, risk, and the agent's risk attitude. The classic paper by Holmstrom and Milgrom (1987) predicts negative relationships between them based on the assumption that the agent cannot adopt new technology. However, their predictions are not consistant with empirical findings (Prendergast, 2002). Besides, there exist many stylized facts indicating that motivating agent to exert the innovation effort is relevant. Our paper develops a simple and crisp model to illustrate that introducing innovation effort could well explain the puzzle of the relationships between the power of incentives, risk and risk-averseness of the agent. Our results show that in general, if it is more profitable for the principal to induce the agent to take the innovative method, then the observed relationships between the power of incentives, the risk (associated with the innovative method), and risk-averseness of the agent can be negative *or positive* depending on the specific parameter values.

References

[1] Ackerberg, Daniel A., and Botticini, Maristella, 2002, "Endogenous matching and the empirical determinants of contract form", *The Journal of Political Economy* 110, 564–592.

[2] Baker, G., and Jorgensen, B., 2003, "Volatility, Noise, and Incentives", Harvard University and NBER Working Paper.

[3] Budde, J., and Krakel, M., 2011, "Limited Liability and the Risk-Incentive Relationship", *Journal of Economics* 102, 97-110.

[4] Guo, M., and Ou-Yang, H, 2006, "Incentives and performance in the presence of wealth effects and endogenous risk", *Journal of Economic Theory* 129, 150-191.

[5] Holmstrom, Bengt, Milgrom, Paul, 1987. "Aggregation and linearity in the provision of intertemporal incentives", *Econometrica* 55, 303–328.

[6] Manso, G., 2011, "Motivating innovation", *The Journal of Finance* 66, 1823–186.

[7] Prendergast, C., 2002, "The tenuous trade-off between risk and incentives", *Journal of Political Economy* 110, 1071–1102.

[8] Serfes, K., 2005, "Risk sharing vs. incentives: contract design under two sided heterogeneity", *Economics Letters* 88, 343–349.

[9] Wright, D.J., 2004, "The risk and incentives trade-off in the presence of heterogeneous managers", *Journal of Economics* 83, 209–223.