# Optimal Contracts Under Adverse Selection and Moral Hazard: A Continuous-Time Approach

**Jaeyoung Sung**
University of Illinois at Chicago

This article presents a continuous-time agency model in the presence of adverse selection and moral hazard with a risk-averse agent and a risk-neutral principal. Under the model setup, we show that the optimal controls are constant over time, and thus the optimal menu consists of contracts that are linear in the final outcome. We also show that when a moral hazard problem adds to an adverse selection problem, the monotonicity condition well known in the pure adverse selection literature needs to be modified to ensure the incentive compatibility for information revelation. The model is applied to a few managerial compensation problems involving managerial project selection and capital budgeting decisions. We argue that in the third-best world, the relationship between the volatility of the outcome and the sensitivity of the contract depends on interactions between the managerial cost and the firm's production functions. Contrary to conventional wisdom, sometimes the higher the volatility, the higher the sensitivity of the contract. The firm receiving good news sometimes chooses safer projects or invests less than it does with bad news. We also examine the effects of the observability of the volatility on corporate investment decisions.

In modern financial economics, managerial moral hazard and adverse selection are widely accepted as impediments to effective contractual arrangements between the manager (the agent) and investors (the principal), resulting in inefficiencies of overall corporate management. Adverse selection problems arise from information asymmetry between the manager and investors before and/or after contracting, and moral hazard problems stem from the unobservability of managerial effort after contracting. Although in real-life problems, moral hazard and adverse selection are often inseparable, in the literature each is often investigated in isolation.[1]

This article develops a continuous-time agency model to examine characteristics of optimal managerial decisions and compensation schemes in

I am grateful to Michael Fishman (the editor), Mark Loewenstein, Heinz Schättler, and the two anonymous referees for many helpful comments and suggestions. I am particularly grateful to the editor and the referees who provided numerous insightful comments which led to a significant improvement of the paper. Of course, all remaining errors are mine. Earlier versions of this paper were presented at the UIC, Boston University, the 2001 Com2Mac Mathematical Finance Workshop, Korea, the 2001 Far Eastern Econometric Society, Japan, the 2001 EFA, Spain, and the 2002 KAFA/KFA, Korea. Address correspondence to Jaeyoung Sung: Department of Finance (M/C 168), University of Illinois at Chicago, 601 South Morgan Street, Chicago, IL 60607-7124, or email: jsung@uic.edu.

[1] Specific examples include contracting problems between an investment banker and an initial public offering (IPO) issuer, between a venture capitalist and an entrepreneur, and between investors and a portfolio manager. See Baron (1982) and Baron and Holmstrom (1980) for IPO contracting and Dybvig, Farnsworth and Carpenter (2001) for portfolio management contracting.

© The Author 2005. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oupjournals.org
doi:10.1093/rfs/hhi020 Advance Access publication May 25, 2005

the presence of both adverse selection and moral hazard problems. The main focus is on optimal managerial risk-taking behaviors in project selection and capital budgeting.

It is well recognized that, for general applicability to financial economic problems, an agency model should be rich enough to allow the manager (1) to control both the mean and variance of the outcome and (2) to be risk-averse. However, no existing agency model for optimal contracting appears to have enough flexibility to accommodate these two features under adverse selection and moral hazard.

Obviously, the controllability of the mean and variance is important because most financial management decisions involve trade-offs between risk and return.[2] Managerial risk aversion is also critical in managerial-contracting problems. Without risk aversion, managerial behavior in corporate risk management becomes trivial because the manager only considers expected payoffs, completely ignoring risks.[3]

In general, however, it is extremely difficult to incorporate the above two important features into a discrete-time contracting model, partly because of the well-known serious technical obstacles in using the first-order approach and partly because of the unmanageably complex algebraic structures of solutions, if any.[4] These obstacles can be overcome with a continuous-time agency framework.

In this article, the outcome is driven by a Brownian motion with drift. During a contracting period, the risk-averse agent continuously exerts effort to control both the drift and diffusion (volatility) rates of the outcome. The risk-neutral principal has an imperfect knowledge of the agent's ability to control the outcome. She observes the final outcome only, and she may or may not observe the agent's volatility controls.

This model extends Holmstrom and Milgrom (1987), Schättler and Sung (1993), and Sung (1995) by incorporating adverse selection problems into their pure moral hazard models.[5] It can also be viewed as an extension of Laffont and Tirole (1986) in that it introduces both uncertainty to the outcome of the agent's effort and risk aversion to the agent's preference. Conceptually, this article is most closely related to McAfee and McMillan

---

[2] Early literature also recognizes the importance but lack of this decision-making feature in agency models. See Kaplan (1984), Holmstrom and Ricart i Costa (1986), and Lambert (1986).

[3] The managerial risk aversion is also important from the financial economics perspective. The agency problem between the manager and investors typically arises from the idiosyncratic (firm specific) part of the overall profit. See, for example, Darrough and Stoughton (1986). Since the compensation scheme necessarily depends on the idiosyncratic risks, the manager exhibits risk aversion over the same risks, whereas well-diversified investors still remain risk-neutral.

[4] See, for example, Mirrlees (1999), Rogerson (1985), and Jewitt (1988).

[5] There are many other related continuous-time agency models. See Sung (1997) for the outcome driven by point (jump) processes; Sung (1991) and Müller (1998) for the first-best solutions; Ou-Yang (1999) for portfolio management applications; Detemple, Govindaraj, and Loewenstein (2001) for a general class of utility functions; and Schättler and Sung (1997), Hellwig and Schmidt (1998), and Müller (2000) for relationships between discrete-time and continuous-time models.

(1986) and Baron and Besanko (1987), who investigated interactions of moral hazard and adverse selection problems with a noisy outcome and a risk-averse agent. Unlike them, we allow the agent to control the variance as well as the mean of the outcome and solve the problem with optimal contracts.[6]

We show that the optimal menu consists of linear contracts. This linearity basically follows from the informational constraint imposed on the principal's problem that requires her to observe only the final outcome and to implement controls based on her initial (time - 0) information. However, the linearity can also be related to the existing literature. It is well known that contracting under adverse selection requires the principal to pay an information rent to induce the agent to reveal his private information. In our continuous-time framework, the rent can be expressed as the expected value of a stochastic process, which we call the information rent process. In fact, the information rent process at any intermediate time $t$ ($<1$) turns out to be the instantaneous rate of an extra cumulative utility gain up to time $t$, the agent would enjoy had he chosen a (marginally) wrong contract at time 0. In general, the rate is history-dependent. Thus, if allowed to observe the whole outcome process, the principal should take this history-dependent information rent process into account in designing the menu of contracts, which makes her problem no longer stationary. However, observing only the final outcome, the principal is concerned with the expected value of the rent process. Since the principal's information remains the same until time 1, the expected value does not change over time, and thus the whole system of the risk-neutral principal's problem becomes stationary as seen in Holmstrom and Milgrom (1987) and Sung (1995). Therefore, the optimal controls (effort levels) remain the same over time, and contracts become linear.

We also show that when a moral hazard problem adds to an adverse selection problem, the monotonicity condition that is well known in the pure adverse selection literature needs to be modified. Typically in the literature, Spence–Mirrlees single-crossing and monotonicity conditions are imposed to induce the agent to reveal his private information. Roughly speaking, the monotonicity condition stipulates that sensitivities of contracts in the optimal menu be nondecreasing in the agent type.[7] This monotonicity condition is derived based on the fact that once the

[6] For other pure adverse selection problems with risk-averse agents, see Matthews (1983), Maskin and Riley (1984), and Salanié (1990). Picard (1987) and Caillaud, Guesnerie, and Rey (1992) examined adverse selection and moral hazard problems with linear and quadratic contracts when both the principal and agent are risk-neutral and the outcome of the agent's effort is noisy. Faynzilberg and Kumar (2000) established a closed-form solution to a binomial outcome model with a risk-averse agent under moral hazard and adverse selection.

[7] See Garcia (2002) for a multidimensional-outcome version of the monotonicity condition.

agent chooses a contract with a particular sensitivity from the menu, he is forced to choose a particular effort level. However, under our adverse selection and moral hazard, the agent's choice of a contract from the menu cannot prevent him from opportunistically choosing effort levels differently from what the principal desires. Thus, the classical monotonicity needs to be modified to counteract such an agent's opportunism in effort choices. [See inequalities (6) and (15) in the text.]

Our model is applied to managerial-contracting problems where the manager exerts costly effort to control the drift of the outcome, but can costlessly choose the volatility. This volatility choice affects the net present value (NPV) of the firm before managerial compensation. Typically, given the drift, the higher the volatility, the higher the NPV.[8] Thus, the volatility choice can be interpreted as a project selection.[9] We call the optimal project under moral hazard the second best and the optimal project under adverse selection and moral hazard the third best.

We argue that in the third-best world, the relationship between the volatility of the outcome and the sensitivity of the contract depends on interactions between the managerial cost and firm's production functions. If the managerial effort productivity is not affected by the volatility, then the well-known negative relationship can hold: regardless of the observability of the volatility, the higher the sensitivity, the lower the volatility. However, if the effort productivity is favorably affected by high volatility, then it is sometime possible that the opposite relationship can hold: regardless of the observability of the volatility, the higher the sensitivity, the higher the volatility.

If the effort productivity increases with the volatility, an increase in volatility brings two opposite effects to the investors' expected wealth: a negative effect of increasing the managerial compensation risk premium and a positive effect of increasing the effort productivity. If the net effect is positive, investors would like to increase the sensitivity even when the volatility increases. Hence, the sensitivity can sometimes be positively related to the volatility.[10]

We further explore, in Section 5, direct effects of the observability on the volatility choices in the second- and third-best worlds. Note that given a contract in place, the manager tends to make more conservative project

[8] Demski and Dye (1999) examined a similar problem with a different approach, focusing on the agent's reporting strategies about the project quality without relying on the revelation mechanism.

[9] The volatility level can alternatively be interpreted as the amount of investment on a project, in the sense that high investment makes the dollar outcome riskier. Thus, project selection problems throughout this article can be viewed as capital-budgeting problems.

[10] In the second-best world examined in Propositions 3-II-ii and 4-II-ii, the net effect is zero. Thus, there is no relationship between the volatility and the sensitivity. However, given the above intuition, one may conjecture that the positive relationship can sometimes hold even in the second best, if the agent's cost structure changes from the one specified in Section 5. We leave the detailed investigation of this possibility for future research.

selection decisions when the decisions are unobservable than he does when they are observable. Thus, it seems reasonable to argue that both in the second- and the third-best worlds, the unobservable volatility is likely to be lower than the observable volatility. However, opposite cases can also occur: the unobservable volatility can sometimes be higher than the observable volatility, particularly when the managerial effort is sufficiently less important than the project selection in improving the total expected profit of the firm.[11] We argue that such opposite cases are more likely to happen in the third best than in the second best, because the third-best sensitivity is typically lower than the second best, and thus the third-best managerial effort is less important than the second best. The possibility of the opposite cases in the third-best world also helps to conditionally support a conjecture by DeMarzo and Duffie (1995) that inefficient managers may choose riskier projects in a hope of lucky draw. In most cases, however, their conjecture does not hold because of the managerial conservatism in the private project selection.

The remainder of this article is organized as follows: Section 1 describes the general continuous-time model of adverse selection and moral hazard for both observable and unobservable volatilities. Section 2 provides a summary of the analysis for the case of observable volatility without proofs and present comparative statics results. Section 3 outlines the analysis for the case of unobservable volatility, with detailed explanations and proofs delegated to the Appendix. This section also presents comparative statics results. Section 4 presents an application of our general results to a few managerial-contracting problems. Section 5 investigates effects of the observability of volatility choices. Finally, Section 6 offers a summary of the article and suggestions for future research.

## 1. The General Model

The formal model of this article belongs to a category of complete contracting under adverse selection and moral hazard with full commitment.[12] The principal is risk-neutral, and the agent exhibits constant absolute risk aversion (CARA) with a CARA coefficient of $r \geq 0$. The principal has an imperfect knowledge of agent's ability $\theta$, whereas the agent knows $\theta$ with certainty. The principal knows that $\theta$ is a random variable independent of all other random variables in the article and

[11] Sung (1995) also reports this possibility using a numerical example in a second-best world. We say the managerial effort is less important than the project selection if the managerial effort level is less sensitive to the sensitivity of the contract than the project NPV.

[12] It is well known that contracting with full commitment is ex ante optimal, but allowing renegotiation is ex post optimal, and that full commitment is suitable for short-term contracting. For long-term contracting, renegotiation can be a significant issue. For conditions where an optimal long-term contract can simply be a sequence of short-term contracts, see Rey and Salanié (1996). However, we leave for future research further implications of renegotiation in a continuous-time agency model with long-term contracting.

distributed on a strictly positive compact interval $\Theta := [\theta_L, \theta_H]$, with a probability density function given by $h(\theta)$, where $h(\theta) > 0$ for all $\theta \in \Theta$.

Before time zero, the principal offers a menu of contracts from which the agent is allowed to choose one. The menu can contain different contracts for agents with different abilities, $\theta$'s. Once a contract is chosen, the agent exerts effort to affect the outcome process during a continuous-time period [0,1]. At time 1, the agent is compensated according to the contract, and the principal keeps the outcome after the compensation.

The uncertainty environment to the agent is given by the probability space $(\Omega, F, P)$, where $\Omega$ is $C[0,1]$, the space of all continuous functions on the interval [0,1]. A standard Wiener process $\{B_t\}$ is defined on $(\Omega, F, P)$ and generates the filtration $\{F_t; t \in [0,1]\}$, satisfying the "usual" condition with $F_1 = F$; that is, $F_0$ contains all probability null sets and $F_t$ is right continuous. Here $\{B_t\}$ is short for $\{B_t; t \in [0,1]\}$. Throughout the article, this short notational convention for all random processes is used.

The outcome process resulting from the agent's effort during [0,1] is denoted by $\{Y_t\}$ and is driven by the following stochastic differential equation (SDE) with an initial value of $Y_0 \in R$:

$$dY_t = f(\mu_t, \sigma_t)dt + \sigma_t dB_t.$$

The drift $f$: $U \times \Sigma \rightarrow R$ is bounded and continuously differentiable in $(\mu, \sigma) \in U \times \Sigma$, where $U$ is an open interval and $\Sigma$ is a strictly positive open interval. Further, we assume that $f_\mu$ is bounded away from zero and strictly increasing in $\mu$ and that $f_{\mu\mu}, f_{\sigma\sigma} \leq 0$. Throughout the article, we use subscripts $\mu$, $\sigma$, and $\theta$ to denote partial derivatives of a function with respect to its corresponding arguments. Further, we use the dot operator to denote the total differentiation of a function with respect to $\theta$, for example, $\dot{x}(\theta) = dx(\theta)/d\theta$.

The agent exerts effort to control $\{(\mu_t, \sigma_t)\}$ and observes the whole outcome process $\{Y_t\}$ in addition to his own effort $\{(\mu_t, \sigma_t)\}$, which means he observes $\{B_t\}$ as well.[13] For effort $(\mu_t, \sigma_t)$at time $t$, the agent incurs an instantaneous cost of $c$. The cost effectiveness of his effort is also affected by his ability $\theta \in \Theta$. In particular, the cost, $c$: $U \times \Sigma \times \Theta \rightarrow R_+$, is bounded and continuously differentiate, and the cumulative cost up to time $t$ is $\int_0^t c(\mu_s, \sigma_s, \theta)ds$ without discounting. Assume that $c_\mu$ and $c_\theta$ are bounded and that $c_\mu$, $c_\theta$, $c_{\mu\mu}$, $c_{\sigma\sigma} \geq 0$.

The principal can observe $Y_0$ and $Y_1$ but not intermediate outcomes. For example, investors (the principal) receive (credibly audited/verified) financial statements such as a balance sheet and an income statement to verify the final cumulative profit at the end of period. Even without knowing intermediate profits, the principal (or the auditor) may be able

---

[13] This setup for the agent control spaces is basically the same as in Sung (1995) except that the outcome process in this article is slightly more general than that of Sung who sets $f(\mu_t, \sigma_t) = \mu_t$.

to verify the final cumulative profit by checking all entry items in those financial statements. It may also be possible to force the manager (the agent) to report intermediate profits. However, verifying those intermediate profits may be too costly to be economically justifiable.[14]

Let $\{G_t\}$be the filtration generated by the principal's observation process $\{Y_0, Y_1\}$. Thus, $G_0 \equiv F_0 \equiv G_t, \forall\, t \in [0, 1]$, and $G_1$ is the augmented sigma algebra generated by $Y_1$. Needless to say, $G_t \subset F_t, \forall\, t \in [0,1]$. In addition to $\{Y_0, Y_1\}$, the principal may or may not observe the agent's volatility controls. We consider two cases:

**Case 1.** The agent chooses the volatility only once at time zero such that $\sigma_t = \sigma_0$ for all $t \in [0,1]$, and the principal observes the volatility. We assume that the volatility is verifiable and contractible. Thus, each contract consists of a sharing rule $S(Y_1,\theta)$ and a level of volatility $\sigma(\theta) \in \Sigma$, and we write the menu of contracts as $\{(S(Y_1,\theta),\ \sigma(\theta));\ \theta \in \Theta\}$. If agent $\theta$ chooses $(S(Y_{1,}\hat{\theta}), \sigma(\hat{\theta}))$, then he should choose the volatility such that $\sigma_t = \sigma(\hat{\theta})$, for all $t \in [0, 1]$. That is, the volatility is fixed by the choice of a contract.

**Case 2.** The agent continuously controls the volatility process $\{\sigma_t\}$ over time based on information $\{F_t\}$, and the principal does not observe the volatility controls. Thus, the volatility is not contractible and we write the menu as $\{S(Y_1,\theta);\ \theta \in \Theta\}$.

In both cases, we assume that an admissible sharing rule $S$ for each $\theta \in \Theta$ is a $G_1$- or $Y_1$-measurable random variable of the following structure: $S(Y_1,\theta) = F(Y_1,\theta) + \alpha(\theta) Y_1$, where $\alpha \in \mathcal{R}$ and $F$ is bounded $Y_1$-measurable random variable. We further require that the admissible sharing rule $S$ implement (i.e., induce the agent to choose) $G_0$-measurable drift and diffusion rates over time.

If agent $\theta$ choose a contract $(S(Y_1, \hat{\theta}), \sigma(\hat{\theta}))$ in Case 1 or $S(Y_1, \hat{\theta})$ in Case 2, then he would exert effort $[(\mu_t(\hat{\theta}, \theta), \sigma_t(\hat{\theta}, \theta))]$ and the outcome would evolve as follows:

$$dY_t = f(\mu_t(\hat{\theta}, \theta), \sigma_t(\hat{\theta}, \theta))dt + \sigma_t(\hat{\theta}, \theta) dB_t, \tag{1}$$

where $\mu_t(\hat{\theta}, \theta)$ and $\sigma_t(\hat{\theta}, \theta)$ are short for $\mu_t((Y_s; 0 \le s \le t), \hat{\theta}, \theta)$ and $\sigma_t((Y_s; 0 \le s \le t), \hat{\theta}, \theta)$, respectively. In both Cases 1 and 2, the agent is allowed to choose an $F_t$-predictable drift process $\{\mu_t\}$. In Case 1, $\sigma_t(\hat{\theta}, \theta) = \sigma(\hat{\theta}) \in \Sigma$ for all $t \in [0,1]$. In Case 2, the agent is free to choose

[14] One can also think of a case where the agent voluntarily reports intermediate outcomes. However, it can be shown that if they cannot be verified, the agent's voluntary reports of a finite number of intermediate outcomes after exerting effort cannot improve the contract. The reason is that to induce the agent to truthfully report intermediate outcomes, it is necessary that the contract should be independent of the agent's reports. However, an issue may arise if the agent reports continuously, because then the reports reveal not only all intermediate outcomes but the whole volatility process. We conjecture that the optimal contract should be independent of both the agent's reports and the volatility process revealed by the reports, and thus unverifiable continuous reports are not useful either.

an $F_t$-predictable volatility process $\{\sigma_t(\hat{\theta},\theta)\}$. We assume that for each $(\hat{\theta},\theta)$, $f$ and $\sigma$ satisfy a Lipschitz condition so that SDE (1) has a unique strong solution.[15]

Finally, we define the principal's Hamiltonian as follows:

$$H(\mu,\sigma,\theta) := f(\mu,\sigma) - c(\mu,\sigma,\theta) - \frac{r}{2}\frac{c_\mu^2(\mu,\sigma,\theta)}{f_\mu^2(\mu,\sigma)}\sigma^2 - v(\theta)c_\theta(\mu,\sigma,\theta), \quad (2)$$

where

$$v(\theta) := \frac{\int_{\theta_L}^{\theta} h(\theta')d\theta'}{h(\theta)}.$$

As will be seen later, this Hamiltonian is relevant to both cases of observable and unobservable volatilities.

## 2. The Optimal Menu of Contracts with Observable Volatility

In this section, we analyze Case 1 where the volatility is observable and can be chosen only once at time 0. The principal's problem is stated as follows:

**Problem 1.** *Choose $G_1$-measurable controls $\{(\mu_t(\theta,\theta));\ \theta \in \Theta\}$ and a $G_1$-measurable menu of contracts, $\{(S(Y_1,\theta),\ \sigma(\theta));\ \theta \in \Theta\}$, to maximize $E(Y_1 - S(Y_1,\theta))$ subject to the following four constraints:*[16]

(i) $\forall\,\theta,\ dY_t = f(\mu_t(\theta,\theta),\ \sigma(\theta))dt + \sigma(\theta)dB_t,$

(ii) $\forall(\hat{\theta},\theta) \in \Theta^2, \{\mu_t(\hat{\theta},\theta)\}$

$$\in \underset{\{\mu_t' \in F_t\}}{\operatorname{argmax}} \left\{ \begin{array}{l} E^B\left(-\exp\left\{-r\left[S(Y_1,\hat{\theta}) - \int_0^1 c(\mu_t',\sigma(\hat{\theta}),\theta)dt\right]\right\}\right) \\ s.t.\ dY_t = f(\mu_t',\sigma(\hat{\theta}))dt + \sigma(\hat{\theta})dB_t \end{array} \right.,$$

[15] The SDE yields a unique strong solution, if the following Lipschitz condition is satisfied. [See Revuz and Yor (1999, p. 375f).] For $\omega,\bar{\omega} \in \Omega\{\equiv C([0,1])\}$, there exists a constant $K$ such that for each $(\hat{\theta},\theta)$,

$$|f_t(\omega,\hat{\theta},\theta) - f_t(\overline{\omega},\hat{\theta},\theta)| + |\sigma_t(\omega,\hat{\theta},\theta) - \sigma_t(\overline{\omega},\hat{\theta},\theta)| \le K \sup_{0\le s\le t}|\omega_s - \overline{\omega}_s|,$$

where $f_t(\omega,\hat{\theta},\theta)$ is short for $f(\mu_t\{Y_s(\omega);\ 0\le s\le t\},\hat{\theta},\theta), \sigma_t(\{Y_s(\omega);\ 0\le s\le t\},\hat{\theta},\theta)$ and $\sigma_t(\omega,\hat{\theta},\theta)$ is another short notation for $\sigma_t(\{Y_s(\omega);\ 0\le s\le t\},\hat{\theta},\theta)$. Note that even though we allow the agent to choose $F_t$-predictable controls, optimal controls will turn out to be $G_0$-measurable, because the principal's restrictive information structure $\{G_0,G_1\}$.

[16] Note that the principal only chooses a menu of contracts and the agent chooses effort levels on $\mu$. However, the problem statement also allows the principal to choose $\{\mu_t(\theta,\theta)\}$ from the agent's argmax set of $\{\mu_t(\hat{\theta},\theta);\hat{\theta},\theta \in \Theta\}$. This is according to the usual conventional assumption that when the agent has multiple optimal control processes or when the argmax set has more than one element, he chooses one from them to maximize the principal's utility.

$$\text{(iii)} \quad \forall\theta, \theta \in \underset{\hat{\theta}}{\operatorname{argmax}} \left\{ \begin{array}{l} E^B\left(-\exp\left\{-r\left[S(Y_1,\hat{\theta}) - \int_0^1 c(\mu_t(\hat{\theta},\theta),\sigma(\hat{\theta}),\theta)dt\right]\right\}\right), \\ \qquad\qquad s.t.\ dY_t = f(\mu_t(\hat{\theta},\theta),\sigma(\hat{\theta}))dt + \sigma(\hat{\theta})dB_t \end{array} \right.$$

$$\text{(iv)} \quad \forall\theta, E^B\left(-\exp\left\{-r\left[S(Y_1,\theta) - \int_0^1 c(\mu_t(\theta,\theta),\sigma(\theta),\theta)dt\right]\right\}\right) \\ \geq -\exp\{-rW_0\}.$$

In the principal's objective criterion, $E$ is the expectation operator with respect to both observed process $\{Y_0, Y_1\}$ and agent ability $\theta$; that is, $E$ is a double integral under two probability measures on $Y_1$and $\theta$. Constraint (ii) is for the incentive compatibility associated with moral hazard: Given any contract $(S(Y_1,\hat{\theta}),\sigma(\hat{\theta}))$from the menu, agent $\theta$ chooses optimal controls $\{\mu_t(\hat{\theta},\theta)\}$ based on the Brownian filtration $\{F_t\}$. $E^B$ is the expectation operator with respect to the Brownian motion process $\{B_t\}$ conditional on $(\hat{\theta},\theta) \in \Theta^2$. The superscript $B$ also emphasizes that the agent information evolves according to $\{F_t\}$. Note that in the agent's problem, the choice of $\sigma$ is determined by the contract, because $\sigma$ is observed by the principal. Constraint (iii) is for the truthtelling condition associated with adverse selection: Given a menu of contracts $\{(S(Y_1,\theta), \sigma(\theta)); \theta \in \Theta\}$, agent $\theta$ chooses $(S(Y_1,\theta), \sigma(\theta))$. This truthtelling condition is a direct consequence of the revelation principle. The last constraint (iv) is the usual participation constraint.

The information structures embedded in Problem 1 indicate that the principal makes decisions based on $\{G_0,G_1\}$ and agent $\theta$ does so based on $\{F_t\}$. Thus, there are two tiers of information asymmetry between the principal and agent; one about $\theta$ that is exogenously given before the contract is chosen and the other about intermediate outcomes $\{Y_t\}$ that are affected by the agent's endogenous effort after a contract is chosen. However, since the principal is the decision maker of the whole system of the problem, optimal solutions, if any, to the system should be adapted to the principal's information, that is, $\{G_0,G_1\}$. Throughout the article, we assume the existence of optimal controls for both the principal's and agent's problems.

## 2.1 Analysis

In this subsection, we summarize results from the analysis of Problem 1. Their proofs are omitted because they are very similar to those in Section 3.1.

The analysis justifies the use of linear contracts in discrete-time models under adverse selection and moral hazard with a risk-neutral principal and a CARA preference agent when the outcome is normally distributed. In particular, optimal controls are constant over time, and they are determined by maximizing the principal's Hamiltonian in (2) over $U \times \Sigma$. For each $\theta$, let $(\mu^*(\theta), \sigma^*(\theta))$ be a solution to the maximization of the Hamiltonian. We call $(\mu^*(\theta), \sigma^*(\theta))$ the *third-best* level of agent $\theta$'s effort, as opposed to the second best $(\mu_{2nd}(\theta), \sigma_{2nd}(\theta))$ under pure moral hazard. Then one can show that the optimal menu of contracts is given by $\{(S^*(Y_1,\theta), \sigma^*(\theta)); \theta \in \Theta\}$, where $S^*(Y_1,\theta)$, the optimal sharing rule for agent $\theta$, is short for $S(Y_1,\theta; \{(\mu^*(\theta'), \sigma^*(\theta'));\theta' \in (\theta,\theta_H)\})$ and

$$S(Y_1, \theta; \{(\mu^*(\theta'), \sigma^*(\theta'));\theta' \in [\theta, \theta_H]\}) = w_0 + \int_{\theta}^{\theta_H} c_\theta(\mu^*(\theta'), \sigma^*(\theta'), \theta')d\theta' \tag{3}$$

$$+ c(\mu^*, \sigma^*, \theta) + \frac{r}{2}\frac{c_\mu^2(\mu^*, \sigma^*, \theta)}{f_\mu^2(\mu^*, \sigma^*)}(\sigma^*)^2 - \frac{c_\mu(\mu^*, \sigma^*, \theta)}{f_\mu(\mu^*, \sigma^*)}f(\mu^*, \sigma^*) + \frac{c_\mu(\mu^*, \sigma^*, \theta)}{f_\mu(\mu^*, \sigma^*)}Y_1,$$

where $c(\mu^*,\sigma^*,\theta) \equiv c(\mu^*(\theta), \sigma^*(\theta),\theta)$ and $f(\mu^*,\sigma^*) \equiv f(\mu^*(\theta), \sigma^*(\theta))$. (Recall $c_\theta$ is the partial differentiation with respect to its third argument.) The proof of Equation (3) is essentially the same as that of Theorem A.2 in the Appendix. Note that $\mu^*(\theta)$ in Equation (3) can be interpreted as the agent $\theta$'s optimal effort level that the principal seeks to induce through the optimal menu of contracts $\{(S^*(Y_1,\theta),\sigma^*(\theta)); \theta \in \Theta\}$.

An intuition for this linearity is presented in Section 3.1. Comparing the structure of the optimal sharing rule with that of the second best in Holmstrom and Milgrom (1987), Schättler and Sung (1993), and Sung (1995), one can see that the second term in the right hand side (RHS) of Equation (3) is added. This term attributed to the adverse selection effect is called *the information rent* to be paid to agent $\theta$ by the principal in order to induce the agent to reveal his private information. Recall that the second-best levels of effort $(\mu_{2nd}(\theta), \sigma_{2nd}(\theta))$ can be computed by maximizing the Hamiltonian in (2) with $v \equiv 0$. An immediate implication is that if $c_\theta > 0$, then for given levels of effort $(\mu^*(\theta), \sigma^*(\theta))$, agent $\theta$, except for the least efficient agent $\theta_H$, has to be paid higher under moral hazard and adverse selection than he does under pure moral hazard. Thus, the third-best levels of effort are different from the second best except for $\theta = \theta_L$.

**2.1.1 Sufficient conditions.** There are a couple of subtle technical issues to have sharing rules in Equation (3) that qualify as truly optimal. Note that in the process of deriving the menu of sharing rules (3) in the Appendix, we have simplified incentive compatibility conditions, (ii) and (iii) of Problem 1, by using the first-order necessary conditions (FOCs) of the agent's problem with respect to $\{\mu_t\}$ and $\theta$. Thus, we still need to make sure the menu $\{(S^*(Y_1,\theta),\sigma^*(\theta)); \theta \in \Theta\}$ satisfies (ii) and (iii).

Before we proceed further, notational clarification is in order. Let a menu of contracts given by $\{(S(Y_1,\theta),\sigma(\theta)); \theta \in \Theta\}$, where $S(Y_1,\theta)$ is a general linear sharing rule in form (3) with $\{(\mu(\theta'),\sigma(\theta'));\ \theta' \in [\theta,\theta_H]\}$, that is, $S(Y_1,\theta) \equiv S(Y_1,\theta;\ \{(\mu(\theta'),\sigma(\theta'));\ \theta' \in [\theta,\theta_H]\})$. Here $(\mu(\theta),\sigma(\theta))$, $\theta \in \Theta$, does not have to maximize the principal's Hamiltonian in (2). It should be emphasized that $\mu(\theta)$ (with a single-$\theta$ argument) is the level of agent $\theta$'s effort the principal tries to induce through the menu $\{(S(Y_1,\theta),\sigma(\theta));\ \theta \in \Theta\}$; that $\mu(\theta,\theta)$ (with double-$\theta$ arguments) is the optimal level of effort agent $\theta$ would actually exert had he chosen the correct contract $(S(Y_1,\theta),\sigma(\theta))$ from the menu and that $\mu(\hat{\theta},\theta)$ is the optimal level of effort agent $\theta$ would actually exert had he chosen a wrong contract $(S(Y_1,\hat{\theta}),\sigma(\hat{\theta}))$ from the menu. In particular, $\mu^*(\hat{\theta},\theta)$, denotes the agent $\theta$'s optimal $\mu$ after choosing a (wrong) contract $(S^*(Y_1,\hat{\theta}),\sigma^*(\hat{\theta}))$ from the optimal menu of contracts $\{(S^*(Y_1,\theta),\ \sigma^*(\theta));\ \theta \in \Theta\}$.

We provide sufficient conditions for a menu to satisfy constraints (ii) and (iii) of Problem 1 in terms of $\mu$- and $\theta$-implementabilities. Consider a menu of contracts $\{(S(Y_1,\theta),\sigma(\theta));\ \theta \in \Theta\}$ with $S(Y_1,\theta) \equiv S(Y_1,\theta;\ \{(\mu(\theta'),\sigma(\theta'));\ \theta' \in [\theta,\theta_H]\})$. If (exogenously) given $(S(Y_1,\theta),\sigma(\theta))$ from the menu, agent $\theta$ chooses $\mu(\theta,\theta)$ so that $\mu(\theta,\theta)=\mu(\theta)$, then we call $(S(Y_1,\theta),\sigma(\theta))$ *$\mu$-implementable*. If all contracts in the menu are *$\mu$-implementable*, we call the menu *$\mu$-implementable*. If agent $\theta$, $\theta \in \Theta$, correctly chooses $(S(Y_1,\theta),\sigma(\theta))$ from the menu, then we call the menu *$\theta$-implementable*. Note that the menu of contracts $\{(S^*(Y_1,\theta),\ \sigma^*(\theta));\ \theta \in \Theta\}$ with $S^*(Y_1,\theta) \equiv S(Y_1,\theta;\ \{(\mu^*(\theta'),\sigma^*(\theta'));\ \theta' \in [\theta,\theta_H]\})$ can be truly optimal only if the two implementabilities are satisfied.

For the two implementabilities, we need to know how agent $\theta$, $\theta \in \Theta$, behaves in general given a menu $\{(S(Y_1,\theta),\ \sigma(\theta));\ \theta \in \Theta\}$. For brevity, we let the sensitivity of contract $(S(Y_1,\theta),\ \sigma(\theta))$ denoted by $\beta(\theta)$. That is,

$$\beta(\theta) := \frac{c_\mu(\mu(\theta),\sigma(\theta),\theta)}{f_\mu(\mu(\theta),\sigma(\theta))}. \tag{4}$$

Suppose agent $\theta$ chooses a contract $(S(Y_1,\hat{\theta}),\sigma(\hat{\theta}))$ from the menu and exerts the optimal level of effort $\mu(\hat{\theta},\theta)$. Then by using essentially the same argument as in the proof of Lemma A.2 in the Appendix, one can show that $\mu(\hat{\theta},\theta)$ satisfies the following equation.

$$\beta(\hat{\theta}) = \frac{c_\mu(\mu(\hat{\theta},\theta),\sigma(\hat{\theta}),\theta)}{f_\mu(\mu(\hat{\theta},\theta),\sigma(\hat{\theta}))}. \tag{5}$$

The above equation implies that given contract $(S(Y_1,\hat{\theta}),\sigma(\hat{\theta}))$, agent $\theta$ exerts efforts on $\mu$ to equate $\beta(\hat{\theta})$, the sensitivity of the contract, to the marginal cost of effort for marginal expected outcome. Since $\beta(\hat{\theta})$ is the marginal benefit to agent $\theta$ for the marginal effort to increase the expected outcome by one unit, Equation (5) simply indicates that agent $\theta$'s optimal effort decision is based on a trade-off between the marginal cost and marginal benefit.

Note that under our general assumptions that $f_\mu > 0$, $c_\mu$, $c_{\mu\mu}$, $c_{\sigma\sigma} \geq 0$, and $f_{\mu\mu}$, $f_{\sigma\sigma} \leq 0$, the agent's marginal cost for marginal expected outcome $c_\mu/f_\mu$ is nondecreasing in $\mu$, and thus Equation (5) implies that given $(S(Y_1,\theta), \sigma(\theta))$ agent $\theta$ optimally chooses $\mu(\theta)$, that is, $\mu(\theta,\theta) = \mu(\theta)$, that is, the menu $\{(S(Y_1,\theta), \sigma(\theta)); \theta \in \Theta\}$ is $\mu$-*implementable*. In other words, all contracts given in form (3) are $\mu$-*implementable* whether each $(\mu(\theta), \sigma(\theta)) \in U \times \Sigma$ maximizes the principal's Hamiltonian or not.

Although the $\mu$-implementability can easily be secured under the above general set of assumptions, the $\theta$-implementability requires an extra condition. An argument similar to the proof of Theorem A.3 in the Appendix implies that a $\mu$-implementable linear menu given in form (3) is $\theta$-implementable if and only if, for all $(\hat{\theta}, \theta)$,

$$c_{\theta\mu}(\mu(\hat{\theta},\theta),\sigma(\hat{\theta}),\theta)\mu_{\hat{\theta}}(\hat{\theta},\theta) + c_{\theta\sigma}(\mu(\hat{\theta},\theta),\sigma(\hat{\theta}),\theta)\dot{\sigma}(\hat{\theta}) \leq 0. \tag{6}$$

(Recall that $\dot{\sigma}(\theta) = d\sigma(\theta)/d\theta$ and $\mu_{\hat{\theta}}(\hat{\theta},\theta)$ are the partial derivatives of $\mu(\hat{\theta},\theta)$ with respect to its first argument.) The above condition can be related to the classical monotonicity and single-crossing conditions frequently appearing in the mechanism design literature. However, we postpone the discussion on differences and similarities between (6) and the classical conditions until Section 3.1.1.

## 2.2 Comparative statics

As discussed in section 2.1, if $\{(\mu^*(\theta), \sigma^*(\theta)); \theta \in \Theta\}$ maximizes the principal's Hamiltonian in (2) over $U \times \Sigma$ and satisfies condition (6), then the menu with sharing rules given in (3) solves Problem 1. To examine the properties of optimal controls $\{(\mu^*(\theta), \sigma^*(\theta)); \theta \in \Theta\}$, let us assume that $\{(\mu^*(\theta), \sigma^*(\theta)); \theta \in \Theta\}$ is the set of interior solutions to the principal's Hamiltonian for all $\theta$.[17] Then FOCs are

$$0 = f_\mu - c_\mu - r\sigma^{*2}\left(\frac{c_\mu}{f_\mu^2}c_{\mu\mu} - \frac{c_\mu^2}{f_\mu^3}f_{\mu\mu}\right) - \nu(\theta)c_{\mu\theta}, \tag{7}$$

$$0 = f_\sigma - c_\sigma - r\sigma^{*2}\left(\frac{c_\mu}{f_\mu^2}c_{\mu\sigma} - \frac{c_\mu^2}{f_\mu^3}f_{\mu\sigma}\right) - r\sigma^*\frac{c_\mu^2}{f_\mu^2} - \nu(\theta)c_{\sigma\theta}. \tag{8}$$

The first two terms of the RHS of Equations (7) and (8) come from necessary conditions for the first best requiring to equate the marginal products to the marginal costs, that is, $f_\mu - c_\mu = 0$ and $f_\sigma - c_\sigma = 0$. The

[17] Given that U and $\Sigma$ are open intervals, the assumption of interior solutions is equivalent to the existence of solutions, and we ignore potential corner solutions. Further note that, as is typically the case in the moral hazard literature, our principal's Hamiltonian may not be globally concave in general. Thus, the FOCs may also describe local maxima, if any, as well as the global maximum.

third term of (7) and the third and fourth terms of (8) are the marginal compensation-risk premiums originating from moral hazard. The last terms of (7) and (8) are related to the marginal information rents due to adverse selection.

When the volatility is observable, the principal chooses $\sigma$ for the agent, and the agent chooses $\mu$ on his own. Thus, the principal's optimal choice of $\sigma$ does not, in general, maximize the agent's expected utility. Consequently, one needs to look at both the principal's optimization conditions and agent's response on $\mu$ to examine comparative statics of $\sigma^*$, which can make general comparative statics highly complicated. We simplify both the principal's and agent's optimization conditions assuming $c_\sigma = 0$, which means the agent chooses $\sigma$ costlessly. This assumption can capture economic situations where the agent can affect the outcome both by costly effort and costless decision-making. For example, in a corporate finance problem, the costless decision on $\sigma$ can be interpreted as a managerial project selection decision. See Sung (1995) and also Section 4.

Comparative statics with $c_\sigma = 0$ are given in the following proposition. Since we have already assumed in the description of the general model that $c_\mu$, $c_{\mu\mu}$, $c_{\sigma\sigma} \geq 0, f_\mu > 0$, and $f_{\mu\mu}, f_{\sigma\sigma} \leq 0$, these conditions will not be repeated in all statements of results throughout the article.

**Proposition 1.** *Assume that $c_\sigma = 0$ and $c_{\mu\mu}$, $c_{\mu\theta} > 0$, that $\mu^*(\theta)$ and $\sigma^*(\theta)$ are continuously differentiable and that condition (6) is satisfied. Then* $[\mu^*(\theta), \sigma^*(\theta)]$ *has the following properties:*

- (i) If $f_{\mu\sigma} = 0$, then the more efficient the agent, the higher the sensitivity, that is, $\dot{\beta}^* \leq 0$.
- (ii) If $f_{\mu\mu} = 0, f_\sigma > 0$, and $f_{\mu\sigma}, f_{\mu\sigma\sigma} \leq 0$, and if for each $\theta$, the maximization of Hamiltonian yields a unique interior solution, then the more efficient the agent, the lower the volatility, that is, $\dot{\sigma}^* \geq 0$.

Note that in the proof in the Appendix, we utilize condition (6) which is absent in pure moral hazard models. Ironically, this extra condition plays a key role in simplifying signs of $\dot{\mu}$ *and* $\dot{\sigma}$*. In the second-best world without a condition like (6), however, those signs are not easy to establish for general* $c$ and $f$. We defer more direct comparisons of the second- and third-best solutions until Proposition 3-I in Section 4.

Part (i) of Proposition 1 is consistent with existing adverse selection literature in which risk-neutral agents are allowed to control only the mean of the outcome. Thus, Proposition 1 simply reconfirms that existing findings also hold even when the agent is risk averse and both the mean and variance are controllable. However, Part (i) requires $f_{\mu\sigma} = 0$. If $f_{\mu\sigma} \neq 0$, the volatility level can affect the marginal productivity of the production function, and thus the sensitivity can be affected by the volatility choice. For example, if $f_{\mu\sigma} < 0$, the marginal productivity drops as $\sigma$ increases, which encourages the principal to instruct the agent to choose safe projects and to provide the agent with a high incentive contract. Consequently, as $\theta$ increases, the principal may want to change the sensitivity more slowly than she may if $f_{\mu\sigma}=0$. This kind of complication makes it difficult to establish the sign of $\dot{\beta}^*$ when $f_{\mu\sigma}\neq 0$.

Part (ii) suggests that the principal instructs more efficient agents to take lower risks. The reason is as follows: She wishes to provide high incentives (with high $\beta$) for efficient agents, but there are high compensation risk premiums associated with raising $\beta$. To reduce the risk premiums, she instructs efficient agents to take low risks. She is even more motivated to discourage efficient agents from taking high risks when $f_{\mu\sigma}<0$, because their marginal effort productivity $f_\mu$ decreases as $\sigma$ increases.

Note, however, that the abovementioned tendency of risk avoidance with highly efficient agents critically depends on our assumption that $f_{\mu\sigma}\leq 0$, that is, the agent's effort productivity is either adversely affected or unaffected by taking high risks. In Section 4, we shall see that if this assumption is violated in a particular way, then it is sometimes possible to have $\dot{\sigma}^*<0$: the more efficient the agent, the higher the volatility. Roughly speaking, if the agent's effort productivity is so favorably affected by high $\sigma$ that the cost of high compensation-risk premium can be outweighed by the benefit of improved effort productivity, then the principal may sometimes want to force efficient agents to take high risks.

## 3. The Optimal Menu of Contracts with Unobservable Volatility

In this section, the volatility process is unobservable and incontractible, as described in Case 2 of Section 1. The principal's problem is stated as follows:

**Problem 2.** *Choose $G_0$-measurable controls $\{(\mu_t(\theta,\theta),\sigma_t(\theta,\theta))\};\theta\in\Theta\}$ and $G_1$-measurable menu ofcontracts $\{S(Y_1,\theta);\ \theta\in\Theta\}$ to maximize$E[Y_1-S(Y_1,\theta)]$ subject to the following four constraints:*

(i) $\forall\theta,\ dY_t=f(\mu_t(\theta,\theta),\sigma_t(\theta,\theta))dt+\sigma_t(\theta,\theta)dB_t,$

(ii) $\forall(\hat{\theta},\theta)\in\Theta^2,[(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta)]$

$$\in\underset{\{(\mu_t',\sigma_t')\in F_t\}}{\operatorname{argmax}}\left\{\begin{array}{c}E^B\left(-\exp\left\{-r\left[S(Y_{1,}\hat{\theta})-\int\limits_0^1 c(\mu_t',\sigma_t',\theta)dt\right]\right\}\right),\\ s.t.\ dY_t=f(\mu_t',\sigma_t')dt+\sigma_t' dB_t\end{array}\right.$$

$$\text{(iii)}\quad\forall\theta,\theta\in\underset{\hat{\theta}}{\operatorname{argmax}}\left\{\begin{array}{c}E^B\left(-\exp\left\{-r\left[S(Y_1,\hat{\theta})-\int\limits_0^1 c(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta),\theta dt\right]\right\}\right),\\ s.t.\ dY_t=f(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta))dt+\sigma_t(\hat{\theta},\theta)dB_t\end{array}\right.$$

$$\text{(iv)} \quad \forall\theta, E^B\left(-\exp\left\{-r\left[S(Y_1,\theta) - \int_0^1 c(\mu_t(\theta,\theta),\sigma_t(\theta,\theta),\theta)dt\right]\right\}\right)$$
$$\geq -\exp\{-rw_0\}.$$

As compared to Problem 1, constraint (ii) now allows the agent to privately choose both the drift and volatility processes over time. All other aspects of Problem 2 are basically the same as those of Problem 1.

## 3.1 Analysis

As in Section 2.1, results are first summarized. Details and their proofs are given in the Appendix. Again, the analysis suggests that optimal controls are constant over time, and thus it justifies the use of linear contracts in analogous discrete-time models when the outcome is normally distributed. In particular, Theorem A.2 in the Appendix indicates that the principal's problem for each $\theta$ can be relaxed to the maximization of the Hamiltonian in (2) subject to the ($\mu$, $\sigma$)-implementability constraint. By Theorem A.3 in the Appendix, if the solution to this constrained maximization problem satisfies the $\theta$-implementability, then the solution also solves the principal's original problem. Both the ($\mu$, $\sigma$)- and $\theta$-implementability conditions will be explained in Section 3.1.1.

The principal's constrained problem implied by Theorem A.2 can further be relaxed as the maximization of the following Lagrangian for each $\theta$ over $U$ x $\Sigma$:

$$\mathcal{L} = f(\mu,\sigma) - c(\mu,\sigma,\theta) - \frac{r}{2}\frac{c_\mu^2(\mu,\sigma,\theta)}{f_\mu^2(\mu,\sigma)}\sigma^2 - \nu(\theta)c_\theta(\mu,\sigma,\theta)$$
$$+q\left[\frac{c_\mu(\mu,\sigma,\theta)}{f_\mu(\mu,\sigma)}f_\sigma(\mu,\sigma) - c_\sigma(\mu,\sigma,\theta) - r\frac{c_\mu^2(\mu,\sigma,\theta)}{f_\mu^2(\mu,\sigma)}\sigma\right], \tag{9}$$

where $q$ is the Lagrange multiplier for the ($\mu$, $\sigma$)-implementability. Note that the constraint part of the Lagrangian is from the agent's FOC with respect to $\sigma$, which constrains the principal's problem because it is now the agent, not the principal, that chooses the volatility. For each $\theta$, let ($\mu^*(\theta)$, $\sigma^*(\theta)$) be a solution to the Lagrangian maximization. By Theorems A.2 and A.3, if $\{(\mu^*(\theta), \sigma^*(\theta)); \theta \in \Theta\}$ satisfy sufficient conditions discussed in Section 3.1.1, then the optimal menu of contracts $\{S(Y_1{}^*,\theta); \theta \in \Theta\}$ consists of linear contracts of form (3) with $\{(\mu^*(\theta), \sigma^*(\theta)); \theta \in \Theta\}$.

Our linearity basically follows from the informational constraint imposed on the principal's problem which requires that she observe only the final outcome and implement $G_0$-measurable controls on ($\mu_t$, $\sigma_t$) over time. However, the linearity can also be related to the existing literature.[18] Recall that in pure moral hazard models of both Holmstrom and Milgrom (1987) and Sung (1995), the optimal contract is linear because of the stationarity of the principal's dynamic problem. The stationarity results from the fact that both the principal and agent exhibit CARA preferences. Under the CARA utility assumptions, wealth effects influence neither the principal's nor the agent's decisions at any intermediate time.

When an adverse selection problem is added to the abovementioned pure moral hazard models, the principal is concerned with one more constraint related to the information rent, an amount that the principal has to pay the agent to reveal his private information. In our continuous time, the rent can be expressed as the expected value of a stochastic process, which we call the information rent process. In fact, the information rent process at any intermediate time $t(<1)$ is the instantaneous rate of an extra cumulative utility gain up to time $t$ the agent would enjoy had he chosen a (marginally) wrong contract at time 0. The rate turns out to be $c_\theta L_t$, a history-dependent integral, where $L_t$ is a history-dependent exponential martingale and $-L_t$ is, in effect, the realized cumulative utility at time $t$. (See Theorem A.1 in the Appendix.) Consequently, the principal maximizes the expected wealth subject to an integral constraint defined in terms of the history-dependent information rent process. If she is allowed to observe the whole history of the outcome, her constrained maximization would no longer exhibit the stationarity over time because of the history-dependent integral constraint and thus could result in history-dependent optimal controls and highly nonlinear whole-history–dependent contracts.

However, when allowed to observe the final outcome only, the risk-neutral principal is only concerned with the expected value of the rent process. Since the principal's information does not change over time, neither does the expected value, and thus the integral constraint in the principal's problem reduces to a deterministic integral, independent of the history of the rent process. Therefore, the principal's problem becomes stationary, resulting in constant optimal controls and linear contracts.

[18] A remark is in order. Recall that linear contracts cannot be optimal in an analogous discrete-time setting, because the first-best solution can be achieved in the limit by a set of the famous Mirrlees (penalty-reward type) forcing contracts. In a continuous-time framework, however, it is not possible to achieve the first best using the Mirrlees forcing contracts, because in continuous time, his contracts do not induce the agent to choose the first-best level of effort. Consider a Mirrlees penalty-reward contract that implements the first-best effort in discrete time. Suppose that the same contract is given to the agent in continuous time. When an intermediate outcome happens to be deeply into the penalty zone, the agent loses the hope of bringing the outcome process back to the reward zone, and thus abandons the effort, that is, no first-best effort is made. Even when the intermediate outcome happens to be deeply into the reward zone, no first-best effort is expected either, because even with zero effort, the agent may find it to be highly unlikely that the final outcome will drop down into the penalty zone.

**3.1.1 Sufficient conditions.** We again discuss the sufficient conditions in terms of the ($\mu$, $\sigma$)- and $\theta$-implementabilities. Consider a menu of contracts $\{S(Y_1,\theta);\, \theta \in \Theta\}$ with $S(Y_1,\theta) \equiv S(Y_1,\theta;\, \{(\mu(\theta'), \sigma(\theta'));\, \theta' \in [\theta,\theta_H]\})$ as in (3). Here the control pair ($\mu(\theta)$, $\sigma(\theta)$), $\theta \in \Theta$, does not have to maximize the principal's Lagrangian in (9). If given $S(Y_1,\theta)$, agent $\theta$ chooses ($\mu(\theta,\theta)$, $\sigma(\theta,\theta)$) such that ($\mu(\theta,\theta)$, $\sigma(\theta,\theta)$) $\equiv$ ($\mu(\theta)$, $\sigma(\theta)$), then we call $S(Y_1,\theta)$ ($\mu$, $\sigma$)-implementable. As in Section 2.1.1, $\mu(\theta,\theta)$ and $\sigma(\theta,\theta)$ are controls determined by the agent and $\mu(\theta)$ and $\sigma(\theta)$ by the principal. The definition of the $\theta$-implementability is identical to that of Section 2.1.1.

To provide sufficient conditions for the ($\mu$, $\sigma$)- and $\theta$-implementabilities, let us define function $\Phi$ as follows:

$$\Phi(\mu, \sigma; \hat{\theta}, \theta) := \beta(\hat{\theta})f(\mu, \sigma) - c(\mu, \sigma, \theta) - \frac{r}{2}(\beta(\hat{\theta}))^2\sigma^2, \tag{10}$$

where $\beta$ is as defined by (4). The above definition is motivated by Lemma A.1 in the Appendix, which gives the ($\mu$, $\sigma$)-implementability condition for an admissible salary function. In fact, $\Phi$ is the quantity agent $\theta$ wishes to maximize after choosing a contract $S(Y_1, \hat{\theta})$ from menu $\{S(Y_1,\theta);\, \theta \in \Theta\}$. Thus, the pair of the agent's optimal response functions $(\mu(\hat{\theta}, \theta), \sigma(\hat{\theta}, \theta))$ is given by

$$(\mu(\hat{\theta}, \theta), \sigma(\hat{\theta}, \theta)) \in \underset{(\mu,\sigma)\in U\times\Sigma}{\operatorname{argmax}}\ \Phi(\mu, \sigma; \hat{\theta}, \theta).$$

The above implies that each $S(Y_1,\theta)$ is ($\mu$, $\sigma$)-implementable, that is, ($\mu(\theta, \theta)$, $\sigma(\theta, \theta)$) $\equiv$ ($\mu(\theta)$, $\sigma(\theta)$), if and only if

$$(\mu(\theta), \sigma(\theta)) \in \underset{(\mu,\sigma)\in U\times\Sigma}{\operatorname{argmax}}\ \Phi(\mu, \sigma; \theta, \theta). \tag{11}$$

In other words, given contract $S(Y_1,\theta)$, agent $\theta$ is successfully induced to choose ($\mu(\theta)$, $\sigma(\theta)$) if and only if the contract is ($\mu$, $\sigma$)-implementable, or if and only if (11) holds.

For an insight into the ($\mu$, $\sigma$)-implementable set of controls, or the set of controls that can make $S(Y_1,\theta)$ ($\mu$, $\sigma$)-implementable, we look at the FOCs and the second-order necessary conditions (SOCs) for the agent's problem in (11). Since $f_{\mu\mu}$, $f_{\sigma\sigma} \leq 0$, and $c_{\mu\mu}$, $c_{\sigma\sigma} \geq 0$ by assumption, the FOCs and SOCs are summarized as follows: with arguments of $c$ and $f$ suppressed,

$$\Phi_\mu = \beta f_\mu - c_\mu = 0, \tag{12}$$

$$\Phi_\sigma = \beta f_\sigma - r\beta^2\sigma(\theta, \theta) - c_\sigma = 0, \tag{13}$$

$$|\Phi''| = (\beta f_{\mu\mu} - c_{\mu\mu})(\beta f_{\sigma\sigma} - c_{\sigma\sigma} - r\beta^2) - (\beta f_{\mu\sigma} - c_{\mu\sigma})^2 \geq 0. \tag{14}$$

If for each $\beta(\geq 0)$, the agent's SOC (14) holds for all ($\mu$, $\sigma$) $\in U \times \Sigma$ [i.e., if $\Phi(\mu, \sigma; \theta,\theta)$is globally concave in ($\mu$, $\sigma$)], then ($\mu(\theta,\theta)$, $\sigma(\theta,\theta)$) is unique, and

Equation (12) is trivially satisfied. Thus, under the global concavity assumption, the ($\mu$, $\sigma$)-implementability reduces to Equation (13), which is the same as the constraint in the principal's Lagrangian (9). Then, contract $S(Y_1,\theta)$ constructed in form (3) by using the solutions to Lagrangian (9) is guaranteed to be ($\mu$, $\sigma$)-implementable and is written as $S^*(Y_1,\theta)$.

Next, we turn to the $\theta$-implementability. By Theorem A.3 in the Appendix, a menu of ($\mu$, $\sigma$)-implementable contracts $\{S(Y_1,\theta);\ \theta \in \Theta\}$, where $S(Y_1,\theta) \equiv S(Y_1,\theta;\ \{(\mu(\theta'),\ \sigma(\theta'));\ \theta' \in [\theta,\theta_H]\})$ is given in form (3), satisfies the $\theta$-implementability if and only if the following inequality holds for all $(\hat{\theta},\theta)$:

$$c_{\theta\mu}(\mu(\hat{\theta},\theta),\sigma(\hat{\theta},\theta),\theta)\mu_{\hat{\theta}}(\hat{\theta},\theta) + c_{\theta\sigma}(\mu(\hat{\theta},\theta),\sigma(\hat{\theta},\theta),\theta)\sigma_{\hat{\theta}}(\hat{\theta},\theta) \leq 0. \quad (15)$$

As a corollary, one can show that if $c_\sigma = 0$, then condition (15) can be simplified to the familiar usual constraint $\dot{\beta} \leq 0$, that is, the sensitivities of optimal contracts are nonincreasing in $\theta$. See Proposition A.2-ii in the Appendix.

Another corollary to condition (15) is that if $c_{\theta\mu} > 0$, $c_{\theta\sigma} \leq 0$, $\mu_{\hat{\theta}}(\hat{\theta},\theta) \leq 0$, and $\sigma_{\hat{\theta}}(\hat{\theta},\theta) \geq 0$, then (15) is satisfied. These conditions correspond to the well known sufficient conditions in discrete-time models: $c_{\theta\mu} > 0$ and $c_{\theta\sigma} \leq 0$ are analogous to the Spence–Mirrlees single-crossing condition and $\mu_{\hat{\theta}}(\hat{\theta},\theta) \leq 0$ and $\sigma_{\hat{\theta}}(\hat{\theta},\theta) \geq 0$ to the monotonicity condition. [See Fudenberg and Tirole (1991) or Stole (1997) for an excellent introduction to these conditions.)[19]

Their single-crossing and monotonicity conditions are developed for a one-dimensional outcome. For multidimensional outcomes, Garcia (2002) provides necessary conditions for the $\theta$-implementability. In fact, Garcia's necessary conditions more closely resemble ours in (15) in the sense that in this article we have two-dimensional actions ($\mu$, $\sigma$). Since we have an agency problem embedded in the adverse-selection problem, it is not straightforward to provide a direct comparison of $\theta$-implementability conditions between the literature and ours. Nevertheless, one may say the main difference lies in the fact that conditions in both Spence–Mirrlees and Garcia require restrictions on $(\dot{\mu}(\theta),\dot{\sigma}(\theta))$,whereas in this article, a restriction is imposed on $(\mu_{\hat{\theta}}(\hat{\theta},\theta),\sigma_{\hat{\theta}}(\hat{\theta},\theta))$. The reason is that in pure adverse-selection models, the principal chooses effort levels for the agent, whereas in this article, the agent does. This additional opportunistic behavior leads to a restriction on $(\mu_{\hat{\theta}}(\hat{\theta},\theta),\sigma_{\hat{\theta}}(\hat{\theta},\theta))$, instead of $(\dot{\mu}(\theta),\dot{\sigma}(\theta))$.

The effect of the agent's opportunistic behaviors on the $\theta$-implementability conditions becomes clear when one tries to compare (6) and (15).

[19] In fact, the single-crossing conditions for our agent's problem can be expressed in terms of either the signs of $\Phi_{\mu\theta}$ and $\Phi_{\sigma\theta}$ or those of $\Phi_{\mu\hat{\theta}}$ and $\Phi_{\sigma\hat{\theta}}$. The inequality condition in (15) is presented in terms of $\Phi_{\mu\theta}(=-c_{\mu\theta})$ and $\Phi_{\sigma\theta}(=-c_{\sigma\theta})$. In order to facilitate comparison with frequent results in the mechanism design literature, Proposition A.1-i in the Appendix alternatively describes a necessary condition for the $\theta$-implementability in terms of $\Phi_{\mu\hat{\theta}}(=\dot{\beta}f_\mu)$ and $\Phi_{\sigma\hat{\theta}}(=\dot{\beta}f_\sigma - 2r\beta\dot{\beta}\sigma)$. However, Proposition A.1-ii tells us that under certain conditions, the two expressions, (15) and Proposition A.1-i, are directly related to each other.

For the case of (6), the principal uses two instruments to screen agents, the compensation contracts and the volatility, whereas for (15), there is only one instrument, the compensation contract. Thus, in the latter case, the agent is more opportunistic than in the former case. Consequently, the principal's choice of menu becomes severely limited as she takes into account agent $\theta$'s all possible opportunistic choices of $\sigma(\hat{\theta}, \theta)$ for all $\hat{\theta}$ given $\theta$, resulting in an additional stringent restriction on $\sigma_{\hat{\theta}}(\hat{\theta}, \theta)$ for all $(\hat{\theta}, \theta)$, not just on $\dot{\sigma}(\theta)$ for all $\theta$. Latter case requires a restriction on $\sigma_{\hat{\theta}}(\hat{\theta}, \theta)$, whereas the former is concerned with $\dot{\sigma}(\theta)$.

### 3.2 Comparative statics

The following comparative statics shed light on the agent's optimal choices of $\mu$ and $\sigma$.

**Proposition 2.** *Suppose that the optimal menu consists of contracts in form* (3) *with sensitivities $\beta^*(\theta) > 0$ for all $\theta \in \Theta$ and that for all $(\hat{\theta}, \theta)$, the agent's Hessian in* (14) *holds with strict inequality. Also suppose that $(c_\mu/f_\mu)f_{\mu\sigma} - c_{\mu\sigma} \le 0$, $f_\sigma - 2r(c_\mu/f_\mu)\sigma < 0$, $c_{\theta\sigma} \le 0$, and $c_{\mu\theta} > 0$. Then the more efficient the agent,* (*i*) *the higher the sensitivity of the contract, that is, $\dot{\beta}^* \le 0$; and* (*ii*) *the lower the volatility, that is, $\dot{\sigma}^*(\theta) \ge 0$.*

Note that the condition $(c_\mu/f_\mu)f_{\mu\sigma} - c_{\mu\sigma} \le 0$ implies $\Phi_{\mu\sigma} \le 0$, that is, the agent's marginal utility with respect to $\mu$ decreases as $\sigma$ increases. In other words, as $\sigma$ increases, the effectiveness of $\mu$ in marginally increasing the agent's utility diminishes. This condition can be satisfied if $c_\sigma = 0$ and $f_{\mu\sigma} \le 0$. On the other hand, the condition $f_\sigma - 2r(c_\mu/f_\mu)\sigma < 0$ can be related to a single-crossing condition that $\Phi_{\sigma\hat{\theta}}$ should be either positive or negative. (See footnote 19 and the discussion following Proposition A.1 in the Appendix.) In Proposition 2, it turns out to be $\Phi_{\sigma\hat{\theta}} \ge 0$, that is, given a contract, the agent's marginal utility with respect to $\sigma$ increases as he pretends to be less efficient. Or the effectiveness of $\sigma$ in marginally increasing the agent's utility increases, as the agent pretends to be less efficient. Also note that the agent's FOC with respect to $\sigma$ immediately implies that the condition $f_\sigma - 2r(c_\mu/f_\mu)\sigma < 0$ holds if $c_\sigma = 0$.

Proposition 2 is similar to Proposition 1, except that conditions to support $\dot{\sigma} \ge 0$ in Proposition 2 are slightly more general than those in Proposition 1. General comparative statics with unobservable volatility are easier to establish than those with observable volatility. The reason is that when the volatility is unobservable, given a contract, comparative statics for both $\mu$ and $\sigma$ are mostly determined by the agent's problem alone, whereas when the principal observes the volatility, one needs to look at the agent's problem for comparative statics of optimal $\mu$ and the principal's problem for comparative statics of optimal $\sigma$.

The intuition for Proposition 2-(ii) is essentially the same as that for Proposition 1-(ii), except that now efficient agents are induced, instead of "instructed," to take low risks. Consequently, if $f_{\mu\sigma} \leq 0$ (or $(c_\mu/f_\mu)f_{\mu\sigma} - c_{\mu\sigma} \leq 0$) and if the volatility is unobservable, then the more efficient the agent, the higher the sensitivity and the lower the volatility. However, as is the case with observable volatility, the sign of $f_{\mu\sigma}$ can be important to establish $\dot{\sigma}^*(\theta) \geq 0$ with unobservable volatility. That is, if $f_{\mu\sigma} > 0$, then the marginal product of effort $\mu$ increases with the volatility so fast that it may sometimes be possible to have the opposite result, that is, $\dot{\sigma}^*(\theta) < 0$. We rely on a specific example in Section 4 to further explore the effect of the sign of $f_{\mu\sigma}$ on decision variables $\mu^*$ and $\sigma^*$, and we compare observable and unobservable volatilities in Section 5.

## 4. An Example

As mentioned in previous sections, we consider corporate management problems where the manager can affect the corporate profit by exerting costly effort $\mu$ and by making costless decision $\sigma$, to examine comparative statics on optimal sensitivities and volatilities under different assumptions about $f_{\mu\sigma}$. We assume that $\nu(\theta)$ is strictly increasing in $\theta$. The second- and third-best solutions are denoted by $(\mu_{2\text{nd}}, \sigma_{2\text{nd}})$ and $(\mu^*, \sigma^*)$, respectively.

Let the cost and production functions be given as follows:

$$c(\mu, \theta) = \begin{cases} \frac{1}{2}(a\mu + b\theta)^2 & \text{if } a\mu + b\theta > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$f(\mu, \sigma) = \gamma\mu + \delta\mu\sigma + g(\sigma).$$

Assume that $a$, $b > 0$, that $g$ is strictly concave real-valued function and that $\gamma$, $\delta \in R$ are parameters such that $f_\mu = \gamma + \delta\sigma > 0$. Given $\mu \in U$, we call a pair $(\sigma, \delta\mu\sigma + g(\sigma))$ a project. Also we call $\delta\mu\sigma + g(\sigma)$ *the total NPV of the project*, $\delta\mu\sigma$ *the effort-productivity gain from volatility,* and $g(\sigma)$ *the project NPV*. Of course these NPVs are before managerial compensation. Note that given $\mu$, choosing a level of $\sigma$ is the same as choosing a project. We say $g$ is quadratic if for $A$, $C \geq 0$ and $B > 0$,

$$g(\sigma) = A - \frac{B}{2}(\sigma - C)^2.$$

The expected total profit, $f$, of the firm before managerial compensation consists of two parts: one from managerial effort, $\gamma\mu$, and the other from the chosen project, $\delta\mu\sigma + g(\sigma)$. When $\delta > (<)0$, the managerial effort not only increases the profit of the firm by $\gamma\mu$ but also improves (reduces) the project opportunity set $\{(\sigma, \delta\mu\sigma + g(\sigma)); \sigma \in \Sigma\}$ by $\delta\mu\sigma$. In the following subsections, for brevity, we use $m$ to denote $a\mu + b\theta$.

### 4.1 Observable project selection

As seen in Section 2, investors maximize the Hamiltonian in (2) over $(\mu, \sigma) \in U \times \Sigma$. The FOCs are as follows:

$$0 = \gamma + \delta\sigma^* - am^* - \frac{ra^3m^*}{(\gamma + \delta\sigma^*)^2}\sigma^*2 - \nu(\theta)ab, \tag{16}$$

$$0 = \frac{\delta}{a}(m^* - b\theta) + g_\sigma - \frac{ra^2\gamma m^{*2}\sigma^*}{(\gamma + \delta\sigma^*)^3}. \tag{17}$$

Equation (16) implies that given $\sigma$, the sensitivity of each contract in the menu is as follows: Let $\nu^{-1}$ be the inverse of $\nu$. Then for $\theta < \nu^{-1}((\gamma + \delta\sigma^*)/ab)$,

$$\frac{am^*}{\gamma + \delta\sigma^*} = \frac{(\gamma + \delta\sigma^*)(\gamma + \delta\sigma^* - \nu(\theta)ab)}{(\gamma + \delta\sigma^*)^2 + ra^2\sigma^*2} \tag{18}$$

and for $\theta \geq \nu^{-1}((\gamma + \delta\sigma)/ab)$, $c_\mu = am^* = 0$, that is, the sensitivities are zero.[20] That is, for managers with sufficiently low abilities, flat-wage contracts are optimal.[21]

Recall that in the first-best world, investors maximize the Hamiltonian in (2) with $r = \nu = 0$ and that in the second-best world, they maximize the same with $\nu = 0$. Thus, the third term of the RHS of the Hamiltonian is attributed to moral hazard and the fourth term to adverse selection.

Equation (17), the FOC with respect to $\sigma$, describes the third-best project selection rule investors instruct the manager to use. Since there are no terms involving $\nu$ in Equation (17), one can argue that the third-best rule is identical to the second-best rule. However, optimal projects for the second- and third-best worlds, in general, can be different, because the second- and third-best sensitivities are different owing to the presence of a term involving $\nu$ in the FOC in (16) and because the sensitivity affects the project selection.

For an intuitive understanding of project selection rules, we consider two benchmark cases. One benchmark is a well-known case with $\delta = 0$ [see Sung (1995, p. 735, Equation (19))]. In this case, the FOC in (17) reduces to

$$g_\sigma - r(\beta^*)^2\sigma^* = 0, \tag{19}$$

where $\beta^* = am^*/\gamma$, the sensitivity of the contract. The above rule simply states that in both second- and third-best worlds, investors' project selection is based on a simple trade-off between the marginal NPV, $g_\sigma$, and the marginal compensation-risk premium, $r(\beta^*)^2\sigma^*$. The marginal premium is the cost of forcing the risk-averse manager to choose riskier projects,

[20] The possibility of such a flat wage contract is also discussed in Bernardo, Cai, and Luo (2001) in the context of capital-budgeting problems with the risk-neutral investors and manager. Their case is a special case of our observable project selection problem with $\gamma = 0$, $\delta > 0$, and $g$ being quadratic.

[21] In an adverse-selection world, if it were believed that those low ability managers who were expected to choose flat-wage contracts were not sufficiently qualified, then investors could easily screen the managers out simply by assigning their certainty equivalent wealth $Q(\theta)$ below their reservation certainty equivalent wealth $W_0$.

because he demands higher risk premium for making his performance measure $Y_1$ riskier. On the other hand, the first-best project selection rule is to maximize the total NPV of the project, that is, $g_\sigma = 0$, because the manager receives a fixed salary and thus demands zero risk premium.

The other benchmark case occurs when $\gamma = 0$ and $\delta > 0$: The FOC in (17) becomes

$$f_\sigma = \mu^* \delta + g_\sigma = 0. \tag{20}$$

That is, both the second- and third-best project selection rules reduce to the first-best rule: Given $\mu$, choose the project maximizing the total NPV. In other words, the first-, second-, and third-best project selection rules are all the same. In particular, the compensation-risk premium does not affect the investors' project selection decision even when the manager is risk averse, because given $\mu$, the risk premium stays constant as $\sigma$ marginally increases. Note that the risk premium is a product of the price of risk $(r/2)(c_\mu/f_\mu)^2$ and the size of risk $\sigma^2$. As $\sigma$ increases, the size increases but the price decreases. The two opposite effects exactly cancel each other to make the risk premium remain constant for all $\sigma > 0$. This precise offset leads to the first-best project selection rule even when the manager is risk averse.

It is instructive to rewrite the project selection rule in (20) as follows:

$$\frac{\delta^2}{a^2}\beta^* \sigma^* - \frac{b\delta\theta}{a} + g_\sigma = 0. \tag{21}$$

Note that the first two terms, $(\delta^2/a^2)\beta^*\sigma^* - b\delta\theta/a$, are the marginal benefit of effort-productivity gain from volatility. Interestingly, given $\beta^*$ and $\theta$, this marginal benefit increases with risks. Therefore, when $g$ is quadratic, investors have incentives to increase risks until they get sufficiently penalized by the negative portion of $g_\sigma$, the marginal project NPV. We use this observation later in Section 4.2 to compare the manager's private project selection rule against investors'.

Comparative statics for the second- and third-best solutions are given below.

**Proposition 3.** *Assume that there are no flat-wage contracts in the menu of contracts and that both the second- and third-best solutions uniquely satisfy the FOCs of corresponding investors' Hamiltonians with equality.*[22] *Then their solutions have the following properties.*

[22] In general, the Hamiltonian may not be globally concave. When $\gamma > 0$ and $\delta \leq 0$, the SOC is algebraically too complex to express in terms of model primitives. However, when $\gamma = 0$ and $\delta > 0$, it can be shown that the SOC is globally satisfied if $a^2(\delta^2 + ra^2)g_{\sigma\sigma} + \delta^2 < 0$, which implies, for instance, that the project frontier has to be sufficiently concave to ensure the global strict concavity of the Hamiltonian.

*(I) Suppose $\gamma > 0$ and $\delta \leq 0$*

*(i) The third-best solutions: $\dot{\beta}^* < 0$ and $\dot{\sigma}^* > 0$.*

*(ii) The second-best solutions: (1) If $\delta < 0$, then $\dot{\beta}_{2nd} < 0$ and $\dot{\sigma}_{2nd} > 0$. (2) If $\delta = 0$, then sensitivities and volatilities are the same across all $\theta \in \Theta$, that is, $\dot{\beta}_{2nd} = \dot{\sigma}_{2nd} = 0$.*

*(iii) Comparison: If $\delta = 0$, then the third-best volatility is higher than the second best, that is, for all $\theta \in (\theta_L, \theta_H]$, $\sigma^*(\theta) > \sigma_{2nd}(\theta)$; and $\sigma^*(\theta_L) = \sigma_{2nd}(\theta_L) = \sigma_{2nd}(\theta)$.*

*(II) Suppose $\gamma = 0$ and $\delta > 0$*

*(i) The third-best solutions: $\dot{\beta}^* < 0$ and $\dot{\sigma}^* < 0$. If g is quadratic,*

$$\sigma^* = \frac{(a^2BC - ab\delta\theta)(\delta^2 + ra^2) - \delta^3\nu(\theta)ab}{a^2B(\delta^2 + ra^2) - \delta^4}. \tag{22}$$

*(ii) The second-best solutions: $\dot{\beta}_{2nd} = 0$ and $\dot{\sigma}_{2nd} < 0$. If g is quadratic,*

$$\sigma_{2nd} = \frac{(a^2BC - ab\delta\theta)(\delta^2 + ra^2)}{a^2B(\delta^2 + ra^2) - \delta^4}. \tag{23}$$

*(iii) Comparison: If g is quadratic, then for all $\theta \in [\theta_L, \theta_H]$, $\sigma^*(\theta) < \sigma_{2nd}(\theta)$ and $\sigma^*(\theta_L) = \sigma_{2nd}(\theta_L)$.*

In the second-best case with $\gamma > 0$ and, $\delta < 0$, Proposition 3-I-ii-(1) is consistent with the conventional understanding that the higher the volatility, the lower the sensitivity. This negative relationship is intuitive in the sense that with $\gamma > 0$ and $\delta < 0$, high volatility accompanies low managerial effort productivity and high compensation risk premium, which in turn implies low sensitivity. However, Propositions 3-I-ii-(2) and -II-ii suggest that either when $\gamma > 0$ and $\delta = 0$ or when $\delta > 0$ and $\gamma = 0$, the second-best sensitivities are the same for all $\theta \in \Theta$.[23] Thus, in these second-best cases, there are no relationships between the sensitivity and volatility.

[23] The second-best sensitivity (with known $\theta$) remains the same for all $\theta$ because of the particular structure of cost and production functions considered in the example. The reason for this can be best understood by looking at the equivalently transformed problem. Using $m = a\mu + \sigma\theta$, one can write $c = (1/2)m^2$, for $m > 0$ and $c = 0$ otherwise. Then

$$dY_t = \left(\frac{\gamma}{a}m + \frac{\delta}{a}m\sigma - \frac{b\gamma}{a}\theta - \frac{b\delta}{a}\sigma\theta + g(\sigma)\right)dt + \sigma_t dB_t.$$

That is, after the change of variable, the cost functions are the same for all $\theta$'s, but the outcome processes are different across $\theta$'s. When $\delta = 0$ and $\gamma > 0$, in the above outcome process, $\theta$ does not interact with effort $m_t$ and $\sigma_t$ at all. Thus, the second-best sensitivities are the same for all $\theta$. When $\gamma = 0$ and $\delta > 0$ (see Part II-ii), $\theta$ does affect the second-best effort and volatility. However with $g$ being quadratic, $c_\mu/f_\mu$ coincidentally happens to be constant because the optimal $c_\mu = am^*$ becomes proportional to $f_\mu = \delta\sigma^*$.

On the other hand, Proposition 3 indicates that the third-best world preserves most qualitative implications of the second-best world, except that when $\delta = 0$ and $\gamma > 0$, high sensitivity is associated with low volatility, and that when $\delta > 0$ and $\gamma = 0$, high sensitivity is associated with high volatility. The positive relationship in the latter case is in contrast with the conventional negative relationship between the volatility and sensitivity.

Behind the rationale for the conventional negative relationship is an implicit assumption that $\delta = 0$ or the managerial effort productivity remains the same regardless of the levels of risks. However, when $\delta > 0$ and $\gamma = 0$, the positive relationship holds in the third best, because unlike the second best, the third-best sensitivity is decreasing in $\theta$ to induce the manager to reveal his private information. (Recall that the second-best sensitivity remains the same across all $\theta \in \Theta$.) As the sensitivity increases, the managerial compensation-risk burden (the marginal cost of managerial project choice) increases. However, as the sensitivity increases, so does effort (the marginal benefit of the managerial project choice). Proposition 3-II-i implies that the net benefit for an efficient manager is larger than that for an inefficient manager, and thus the efficient manager chooses riskier projects. Hence, we have a positive relationship between the sensitivity and volatility.

In comparing the second- and third-best quantitatively, the second-best volatility can be higher or lower than the third best depending on values in $\gamma$ and $\delta$. If $\gamma > 0$ and $\delta = 0$ for each $\theta \in (\theta_L, \theta_H]$, the second-best volatility is pointwise lower than the third best. From now on, whenever we compare the first-, second-, and third-best solutions, we will always do so "pointwise in $\theta \in (\theta_L, \theta_H]$," without using this qualifier. With $\delta = 0$, the managerial effort productivity is unaffected by volatility. Since the second-best sensitivity is typically higher than the third best, investors are less willing to take risks in the second-best world than in the third best.

If $\gamma = 0$, $\delta > 0$ and $g$ is quadratic, then the second-best volatility is higher than the third best. As noted before, with $\delta > 0$, the effort productivity increases with volatility. Thus, for high productivity, a high volatility environment is necessary. In the third-best world, because of "information rent" paid to the manager, an efficiency loss is necessary, which makes the third-best sensitivity and volatility lower than the second best.

### 4.2 Unobservable project selection

Now, the manager privately makes project-selection decisions. With $c_\sigma = 0$, investors' Lagrangian (9) can be simplified as follows: For each $\theta$

$$\mathcal{L} = f - c - \frac{r}{2}\frac{c_\mu^2}{f_\mu^2}\sigma^2 - \nu(\theta)c_\theta + q\left(f_\sigma - r\frac{c_\mu}{f_\mu}\sigma\right). \tag{24}$$

With the quadratic cost and $f = \gamma\mu + \delta\mu\sigma + g(\sigma)$, the above Lagrangian yields the following FOCs.

$$\mathcal{L}_\mu = \gamma + \delta\sigma^* - am^* - \frac{ra^3 m^*}{(\gamma + \delta\sigma^*)^2}\sigma^{*2} - \nu ab + q\left(\delta - \frac{ra^2\sigma^*}{\gamma + \delta\sigma^*}\right) = 0, \quad (25)$$

$$\mathcal{L}_\sigma = \frac{\delta}{a}(m^* - b\theta) + g_\sigma - \frac{ra^2\gamma m^{*2}\sigma^*}{(\gamma + \delta\sigma^*)^3} + q\left[g_{\sigma\sigma} - \frac{ra\gamma m^*}{(\gamma + \delta\sigma^*)^2}\right] = 0, \quad (26)$$

$$\mathcal{L}_q = \frac{\delta}{a}(m^* - b\theta) + g_\sigma - \frac{ram^*\sigma^*}{\gamma + \delta\sigma^*} = 0. \quad (27)$$

Note that by the construction of cost and production functions, the optimal sensitivity cannot be negative. Thus, by the FOC with respect to $\mu$, the optimal sensitivity

$$\frac{am^*}{\gamma + \delta\sigma^*} = \frac{(\gamma + \delta\sigma^*)^2 - \nu ab(\gamma + \delta\sigma^*) + q\left[\delta(\gamma + \delta\sigma^*) - ra^2\sigma^*\right]}{(\gamma + \delta\sigma^*)^2 + ra^2\sigma^{*2}} \quad (28)$$

is well defined for all $\theta$ with which the above quantity is positive, that is, for all $\theta$'s such that $\nu(\theta)ab(\gamma + \delta\sigma^*) < (\gamma + \delta\sigma^*)^2 + q\left[\delta(\gamma + \delta\sigma^*) - ra^2\sigma^*\right]$. For all other $\theta$'s, the sensitivities are zero, that is, $c_\mu = am^* = 0$. That is, for managers with sufficiently low abilities, flat-wage contracts are optimal.

The last two of the above three FOCs describe how the project-selection decision is made. Since the two are free of terms involving $\nu$, again the third-best project selection rule is the same as the second best. The FOC in (27) implies that the manager chooses a project equating the marginal expected compensation to the marginal compensation-risk burden. The FOC in (26) with $q = 0$ suggests that investors wish the manager to choose a project equating the marginal NPV to the marginal compensation-risk premium. (Note that at the optimum, the compensation-risk burden and premium are equal.) However, since the marginal expected compensation is a fraction of the marginal NPV, given a contract, the manager tends to be more conservative in project selection than desired by investors. Hence, there are conflicts of interests about the project selection between the manager and investors.

To be specific, we look at two benchmark cases: When $\delta = 0$ and $\gamma > 0$, we have the projection selection considered in Sung [1995, p. 735, Equation (22)], and by the FOC (27), the project-selection rule can be written as

$$\beta^* g_\sigma - r(\beta^*)^2\sigma^* = 0. \quad (29)$$

The above rule equates the marginal expected managerial compensation, $\beta^* g_\sigma$, to the marginal compensation-risk burden on the manager, $r(\beta^*)^2\sigma^*$. Given a contract with $\beta^*$, the marginal expected compensation for each $\theta$ is a fraction $\beta^*$ of $g_\sigma$, the marginal total NPV of the optimal project. As compared with the observable project- selection rule in (19), the rule in (29) suggests that when the project choice is unobservable, the manager has incentives to choose safer projects than desired by investors, because the expected marginal benefits to the manager from a risky project is only a fraction of the marginal NPV from the project.

When $\delta > 0$ and $\gamma = 0$, the project-selection rule given by the FOC (25) can be rewritten as follows

$$\left(\delta^2 - ra^2\right)\beta^*\sigma - ab\delta\theta + a^2 g_\sigma = 0. \tag{30}$$

In this case, the marginal total NPV consists of two parts: $\mu^*\delta = (\delta^2/a^2)\beta^*\sigma^* - (b\delta/a)\theta$, the marginal effort-productivity gain from volatility; and $g_\sigma$, the marginal project NPV. In fact, if $\delta^2 - ra^2 > 0$, then as $\sigma$ increases, the marginal expected compensation due to the marginal effort-productivity gain grows faster than marginal compensation risk burden. Thus, when $g$ is quadratic, the manager has incentives to choose risky projects until he is penalized by a significant reduction of the project NPV as $g$ starts decreasing beyond a certain level, that is, $\sigma = C$. If $\delta^2 - ra^2 < 0$, then as $\sigma$ increases, the marginal expected compensation grows slower than marginal compensation risk burden, and thus the manager has incentives to choose safe projects.

Comparative statics for the second- and third-best solutions are given below.

**Proposition 4.** *Assume that there are no flat-wage contracts in the menu of contracts and that both the second- and third-best solutions uniquely satisfy the FOCs of corresponding investors' constrained Hamiltonian with equality.*[24] *Then the solution has the following properties.*

*(I)* *Suppose $\gamma > 0$ and $\delta = 0$*

- *(i)* *The third-best solutions: $\dot{\beta}^*$<0 and $\dot{\sigma}^*$>0.*[25]
- *(ii)* *The second-best solutions: For all $\theta \in \Theta$ $\dot{\beta}_{2nd} = \dot{\sigma}_{2nd}(\theta) = 0$.*[26]
- *(iii)* *Comparison: $\sigma^*(\theta) > \sigma_{2nd}(\theta)$ for all $\theta \in (\theta_L, \theta_H]$; and $\sigma^*(\theta_L) = \sigma_{2nd}(\theta)$.*

*(II)* *Suppose $\gamma = 0$ and $\delta > 0$. Also suppose that $\dot{\nu}\left(\delta^2 - ra^2\right) + ra^2 \geq 0$ and $(\delta^2 - ra^2)g_{\sigma\sigma\sigma} \geq 0$ and that either $ra^2 - \delta^2 \geq 0$ or $a^2B - \delta^2 \geq 0$.*[27]

---

[24] See footnote 22. The constrained Hamiltonian may not be globally concave in general. When $\gamma = 0$ and $\delta > 0$, the strict concavity holds if $\delta^2 - ra^2 + a^2 g_{\sigma\sigma} < 0$.

[25] If the $\theta$-implementability is assumed, then by Proposition 2, one can immediately conclude that $\dot{\beta}^*$<0 and $\dot{\sigma}^*$>0 for $\gamma > 0$ and $\delta \leq 0$. Note however, that both Propositions 3 and 4 do not *a priori* assume the $\theta$-implementability.

[26] See footnote 23 for an explanation of the same sensitivities.

[27] Assumptions of $ra^2 - \delta^2 \geq 0$, or $a^2B - \delta^2 \geq 0$, are technical conditions to ensure the ($\mu$, $\sigma$)-implementability, or the manager's SOC. If $\theta$ is uniformly distributed, then $\dot{\nu} = 1$, and the condition $\dot{\nu}(\delta^2 - ra^2) + ra^2 \geq 0$ is satisfied.

(i) *The third-best solutions:* $\dot{\beta}^* < 0$ *and* $\dot{\sigma}^* < 0$. *If g is quadratic, then for* $\theta \in \Theta$,

$$\sigma^*(\theta) = \frac{\left(a^2BC - \delta ab\theta\right)\left[B\left(\delta^2 + ra^2\right) - r\left(\delta^2 - ra^2\right)\right] - B\left(\delta^2 - ra^2\right)\nu(\theta)ab\delta}{B\left(\delta^2 + ra^2\right)\left[a^2B - \left(\delta^2 - ra^2\right)\right]} \tag{31}$$

(ii) *The second-best solutions:* $\dot{\beta}_{2nd} = 0$ *and* $\dot{\sigma}_{2nd}(\theta) < 0$.[28] *If g is quadratic, then for* $\theta \in \Theta$,

$$\sigma_{2nd}(\theta) = \frac{\left(a^2BC - \delta ab\theta\right)\left[B\left(\delta^2 + ra^2\right) - r\left(\delta^2 - ra^2\right)\right]}{B(\delta^2 + ra^2)[a^2B - (\delta^2 - ra^2)]}. \tag{32}$$

(iii) *Comparison: If* $\delta^2 - ra^2$ $(=, <)0$, *then we have* $\sigma^*(\theta) < (=, >)$ $\sigma_{2nd}(\theta)$ *for all* $\theta \in (\theta_L, \theta_H]$, *and* $\sigma^*(\theta_L) = \sigma_{2nd}(\theta_L)$.

All results in Proposition 4 except part II-iii are qualitatively similar to those of Proposition 3. The differences between Propositions 3 and 4 occur simply because of project-selection rules (17) and (27). As noted before, since that he privately recognizes a smaller marginal benefit from the project NPV than investors do, the manager has stronger incentives to choose safe projects when his project choice is unobservable than when observable. In most cases, qualitatively, the stronger incentives alone do not much alter the relationships of the sensitivity and volatility between the second- and third-best worlds. For example, both Propositions 3 and 4 agree that if the volatility does not affect the effort productivity, the third-best volatility is greater than the second best. However, a disagreement can occur when the volatility positively affects the effort productivity.

To understand the disagreement, recall a couple of facts. First, the third-best sensitivity is typically lower than the second best. Thus, the risk-averse manager demands lower risk premium in the third best to choose a risky project than he does in the second best. Second, consider cases with $\gamma = 0$ and $\delta > 0$. When the project choice is observable, investors instruct the manager to choose the highest NPV project given $\mu$, because the positive volatility effect completely offsets the managerial compensation risk premium. When the project selection is unobservable, the manager may still have incentive to choose a low NPV project to reduce the compensation-risk burden on himself, unless the positive volatility effect is sufficiently high.

The disagreement depends on quantity $\delta^2 - ra^2 < 0$. If $\delta^2 - ra^2 < 0$, or if the volatility effect is not sufficiently positive, then the marginal positive volatility effect on the expected managerial compensation does not completely offset the managerial marginal risk burden, and thus given $\mu$, the manager tends to choose safer projects than investors desire. Since the

[28] See footnote 23.

third-best sensitivity is typically lower than the second best, the third-best contract typically gives lower managerial risk burden, and therefore the third-best volatility becomes higher than the second best, which results in a disagreement between Propositions 3 and 4. That is, when observable, the third-best volatility is lower than the second best, whereas when unobservable, the opposite holds.

However, if $\delta^2 - ra^2 > 0$, or if the volatility effect is sufficiently positive, then the managerial marginal compensation risk burden is completely offset by the marginal benefit of positive volatility effect. Thus, both the manager's and investors' interests are aligned to choose the highest NPV project given $\mu$. Hence, in this case, both Propositions 3 and 4 agree, that is, observable or not, the third-best volatility is lower than the second best.

## 5. Effects of Unobservability

Given our results thus far, we compare observable and unobservable project-selection cases. First, we discuss a simple benchmark where at the optimum, both observable and unobservable volatilities turn out to be the same. Then we present a case deviating from the benchmark.

**5.1 A simple case where the observability is irrelevant.** A glance at the Lagrangian reveals simple, albeit restrictive, conditions under which the unobservability does not affect the principal's wealth, as in the following proposition, which can also be viewed as a corollary to Theorem A.2 in the Appendix.

**Proposition 5.** (1) *If* $c_{\mu\sigma} = f_\sigma = 0$ *and* $c_{\theta\sigma} = 0$, *then the observability of* $\sigma$ *is irrelevant for* $\theta \in \Theta$. (2) *If* $c_{\mu\sigma} = f_\sigma = 0$ *and* $c_{\theta\sigma} \neq 0$, *then the observability of* $\sigma$ *is irrelevant only for* $\theta = \theta_L$, *the most efficient agent.*

Under pure moral hazard [see Sung (1995, Corollary 1)], if $c_{\mu\sigma} = 0$ and $f_\sigma = 0$, then the observability of $\sigma$ is irrelevant. However, Proposition 5 indicates that if an adverse selection problem is added, then the observability matters even when $c_{\mu\sigma} = 0$ and $f_\sigma = 0$, unless $c_{\theta\sigma} = 0$.

If $c_{\theta\sigma} \neq 0$, and if the adverse selection problem is added to the moral hazard, then given a contract, $\sigma(\theta)$ affects the marginal information rent $c_\theta$. If $c_{\theta\sigma} > (<)0$, then the information rent to be paid to each agent increases (decreases) with $\sigma$, a negative (positive) effect to the principal expected wealth. Thus, when $\sigma$ is observable, given sensitivities, the principal chooses $\sigma$ by striking a balance among marginal changes in four factors: marginal expected outcome, agent's cost of effort, compensation-risk premium, and information rent. [See (8), the investors' FOC with respect to $\sigma$.] However, when $\sigma$ is unobservable, each agent completely ignores the marginal information rent and bases a project decision on a trade-off among three factors: marginal compensation, marginal cost of agent's effort, and marginal compensation-risk burden. [See (13), the agent's FOC with respect to $\sigma$.] That is, conflicts can arise between the principal and agent over the choice of $\sigma$. Therefore, if $c_{\theta\sigma} \neq 0$ under adverse selection and moral hazard, the observability matters even when $c_{\mu\sigma}=0$ and $f_{\sigma}=0$.

**5.2 A case where the observability matters.** Recall from Section 4 that given a contract, the manager has incentives to privately choose safer projects than desired by investors, and thus unobservable volatility is expected to be lower than observable volatility. Also recall that the total expected outcome can be increased either by increasing managerial effort or by choosing risky projects. When effort incentives can conflict project-selection incentives, investors should weigh the importance between the two and decide which incentive has to be sacrificed in favor of the other.

In the presence of adverse selection, the sensitivity is typically lower than the second best, and thus the contribution of managerial effort to the expected outcome becomes less important than it is in the second best. As a result, investors can be more willing to sacrifice managerial effort incentives in the third best than they are in the second best, which can in turn make it more likely that unobservable volatility is greater than observable volatility in the third best than in the second best.

We reinforce the above intuition by using the case of $f=\mu\sigma\delta + g(\sigma)$ considered in Propositions 3-II and 4-II. In this case, we have explicit solutions for optimal volatilities, which makes the comparison easier.

**Proposition 6.** *Suppose the assumptions for Proposition 4-II-i hold, then optimal observable and unobservable volatilities in the second- and third-best worlds are compared as follows.*

1. $\forall\theta, \sigma_{2\text{nd}}^{\text{obs}}(\theta) > \sigma_{2\text{nd}}^{\text{unobs}}(\theta)$.
2. $\sigma_{\text{obs}}^{*}(\theta) > (=,<)\sigma_{\text{unobs}}^{*}(\theta)$ *iff* $(aBC - b\delta\theta)\delta - \nu(\theta)Ba^2b > (=,<)0$.
3. If $\theta$ is uniformly distributed on $[\theta_L,\theta_H]$, then

$$\sigma_{\text{obs}}^{*}(\theta) > (=,\ <)\sigma_{\text{unobs}}^{*}(\theta)\ \textit{iff}\ \theta < (=,\ >)\theta_K := \frac{aBC\delta + Ba^2b\theta_L}{b\delta^2 + Ba^2b}.$$

When $f=\mu\sigma\delta + g(\sigma)$ with quadratic $c$ and $g$, Proposition 6 indicates that it does not happen in the second best that the unobservable volatility is greater than the observable volatility. However, in the third best it can happen, because the managerial effort is less important than it is in the second best, and thus investors are more willing to sacrifice the efficiency of managerial effort incentives.

More clearly, under the uniform distribution assumption, for sufficiently high $\theta, \sigma_{\text{obs}}^{*}(\theta)<\sigma_{\text{unobs}}^{*}(\theta)$. That is, the degree of the willingness to sacrifice effort incentives increases as the managerial effort efficiency decreases. This suggests that sufficiently inefficient managers such that $\theta > \theta_K$ are induced to choose higher volatilities when their choices are unobservable than they are when observable. This implication conditionally supports the conjecture by DeMarzo and Duffie (1995, p. 769) that inefficient managers may choose riskier projects in a "hope of a lucky draw." Of course, if $\theta_H < \theta_K$, then DeMarzo–Duffie case cannot occur. In general, their conjecture does not hold for sufficiently efficient managers such that $\theta < \theta_K$, because managers are risk-averse and they are efficient enough that investors do not want to sacrifice the efficiency of effort-incentive contracts too much.

## 6. Conclusion

We have presented a continuous-time principal-agent model under adverse selection and moral hazard, where the principal is risk-neutral, the agent exhibits constant absolute risk aversion, and the agent controls both the drift and volatility of the outcome process. The principal may or may not observe the volatility and is allowed to observe the final outcome only. We have shown that in the principal's problem, the truthtelling condition inducing the agent to reveal his private information can be reduced to an information-rent integral constraint. Thus, solving the principal's problem becomes similar to solving, under the integral constraint, a pure moral hazard problem as seen in Holmstrom and Milgrom (1987), Schättler and Sung (1993), and Sung (1995). Given the principal's informational restriction to observe the final outcome only and to implement controls based on her initial information, this integral constraint introduces no wealth effect to her problem, and thus optimal contracts remain linear in the final outcome. We have also shown that the optimal menu is required to satisfy a modified version of the well-known Spence–Mirrlees single-crossing and monotonicity conditions, to prevent the agent's opportunistic effort choices after choosing a wrong contracts from the menu.

As an application of the general model, we have provided in Section 4 an example that can be used to address various corporate contracting problems and project-selection decisions under adverse selection and moral hazard. Note that our project-selection problems can also be easily transformed into capital budgeting problems by interpreting the volatility choice as the choice of a dollar amount of investment: the higher the volatility, the larger the investment.

The example in Section 4 reveals that corporate project choice or investment decisions can be affected not only by the observability of managerial decisions and information asymmetry, but also by interactions between the managerial cost and corporate production functions. We have shown that contrary to conventional wisdom, it is sometimes possible to have a positive relationship between the volatility and the sensitivity of the contract. Such a case can sometimes happen, if the managerial effort productivity is favorably affected by the riskiness of the firm. In this case, efficient managers choose riskier projects or invest more than inefficient managers do.

If the effort productivity is either unaffected or unfavorably affected by the volatility, then we have the conventional negative relationship between the volatility and the sensitivity. In this case, efficient managers choose safer projects or invest less than efficient managers do.

We have argued that in most cases, the manager tends to be more conservative in project-selection decisions when they are unobservable than when they are observable. However, under adverse selection and moral hazard, when investors believe there is a positive probability that the manager is highly efficient but in fact he is not sufficiently efficient up to a threshold, then he may be optimally induced to choose riskier projects when the choice is unobservable than when the choice is observable. This result conditionally supports DeMarzo–Duffie's (1995) conjecture.

On the other hand, it is well known that above results on project selection/capital budgeting decisions under adverse selection and moral hazard presented in Section 4 can also equivalently be applicable to the same decisions under hidden information and moral hazard, in which information asymmetry occurs after the managerial contract is signed but before the manager exerts effort. [See Fudenberg and Tirole (1991, Chapter 7)]. Under hidden information and moral hazard, the manager with good (bad) news behaves like an efficient (inefficient) manager under adverse selection and moral hazard.

For instance, the example in Section 4 can be reinterpreted in the context of hidden information and moral hazard as follows: Regardless of the observability of volatility (investment decisions), (1) if managerial effort productivity is unaffected by the volatility (by the amount of investment), then the firm, after receiving good news, chooses safer projects (invests less) than it does after receiving bad news; and (2) if the managerial effort productivity is sufficiently positively affected by the volatility (by the amount of investment), then the firm, after receiving good news, can sometimes choose riskier projects (invest more) than it does after receiving bad news.

For simplicity, let us focus on unobservable capital budgeting implications. In the first case, the firm's conservatism in investment decisions with good news appears to be counterintuitive. When good news arrives, the manager chooses a high-sensitivity contract as a consequence of the revelation mechanism. Since the high sensitivity increases the compensation-risk burden on the manager, he invests less. Thus, good news is associated with a small investment. However, in the second case, the high sensitivity does not necessarily increase the compensation-risk burden as much as it does in the first case, because a large investment improves his effort productivity and reduces the marginal price of the compensation-risk burden per one unit of volatility. (Note that the compensation-risk burden depends on the marginal cost of effort and the volatility.) As a result, good news can be associated with a large investment.

The continuous-time model of this article may be extended to many interesting economic settings. In our adverse selection model, when the information asymmetry exists before contracting, managerial career concern and renegotiation can potentially emerge as important considerations in long-term contracting. [See Fudenberg, Holmstrom and Milgrom (1990) and Rey and Salanié (1996).] It would be interesting to see how the information asymmetry in terms of managerial ability can affect managerial contracting if managerial career concern and the possibility of renegotiation are allowed to dynamically interact with adverse selection. A continuous-time model that can successfully tackle these issues with risk-averse agents allowed to control both the mean and volatility of the outcome might help provide new economic insights into many existing long-term agency problems in financial economics.

The optimal menu in the article may possibly be extended to an optimal screening menu. In this article, the principal has already decided to hire the agent and tries to offer a menu of contracts which the agent will accept. However, one can think of a more complex setting where the principal tries to hire one agent from a pool of many candidates with different ability levels and would like to screen out bad candidates from the pool by using a menu of contracts. This problem is more complex than ours in the sense that it involves an optimal screening menu. That is, it requires the principal to consider the probability of successfully hiring one qualified agent and to solve an additional optimization problem, say, with respect to $\theta_H$, after solving our problem. We leave all these potentially interesting extensions for future research.

## Appendix

**A Proof of Proposition 1.** (i) Since $c_\sigma = 0$ and $c_{\mu\theta} > 0$, the $\theta$-implementability condition in (6) implies that $\mu^*_{\hat{\theta}}(\hat{\theta}, \theta) \leq 0$. On the other hand, by varying Equation (5) over $\theta$ with $\hat{\theta}$ fixed, we have

$$(\beta^*(\hat{\theta})f_{\mu\mu} - c_{\mu\mu})\mu^*_\theta - c_{\mu\theta} = 0. \tag{A1}$$

Therefore,

$$\begin{aligned}\dot{\beta}^* &= \frac{d}{d\theta}\left(\frac{c_\mu}{f_\mu}\right) = \frac{1}{f_\mu}\left[(c_{\mu\mu}\dot{\mu}^* + c_{\mu\theta}) - \frac{c_\mu}{f_\mu}(f_{\mu\mu}\dot{\mu}^* + f_{\mu\sigma}\dot{\sigma}^*)\right] \\ &= \frac{1}{f_\mu}\left[(c_{\mu\mu} - \frac{c_\mu}{f_\mu}f_{\mu\mu})\dot{\mu}^* + (\frac{c_\mu}{f_\mu}f_{\mu\mu} - c_{\mu\mu})\mu^*_\theta - \frac{c_\mu}{f_\mu}f_{\mu\sigma}\dot{\sigma}^*\right] \\ &= \frac{1}{f_\mu}\left[(c_{\mu\mu} - \frac{c_\mu}{f_\mu}f_{\mu\mu})\mu^*_{\hat{\theta}} - \frac{c_\mu}{f_\mu}f_{\mu\sigma}\dot{\sigma}^*\right] \leq 0.\end{aligned}$$

For the second equality, we have utilized (A1) evaluated at $\hat{\theta} = \theta$. The last inequality follows because $f_{\mu\sigma} = 0$ by assumption and $\mu_{\hat{\theta}} \leq 0$ by the $\theta$-implementability.

(ii) Since $c_\sigma = 0$, the FOC with respect to $\sigma$ is, for each $\theta$,

$$0 = f_\sigma + r\sigma^{*2}\frac{c_\mu^2}{f_\mu^3}f_{\mu\sigma} - r\sigma^*\frac{c_\mu^2}{f_\mu^2}.$$

Thus,

$$0 = \left\{f_{\mu\sigma} - 3r\sigma^{*2}\frac{c_\mu^2}{f_\mu^4}f_{\mu\sigma}f_{\mu\mu} + r\sigma^{*2}\frac{c_\mu^2}{f_\mu^3}f_{\mu^2\sigma} + 2r\sigma^*\frac{c_\mu^2}{f_\mu^3}f_{\mu\mu}\right\}\dot{\mu}^*$$

$$+\left\{f_{\sigma\sigma} + 2r\sigma^*\frac{c_\mu^2}{f_\mu^3}f_{\mu\sigma} - 3r\sigma^{*2}\frac{c_\mu^2}{f_\mu^4}f_{\mu\sigma}^2 + r\sigma^{*2}\frac{c_\mu^2}{f_\mu^3}f_{\mu\sigma^2} - r\frac{c_\mu^2}{f_\mu^2} + 2r\sigma^*\frac{c_\mu^2}{f_\mu^3}f_{\mu\sigma}\right\}\dot{\sigma}^*$$

$$+\left\{2r\sigma^{*2}\frac{c_\mu}{f_\mu^3}f_{\mu\sigma} - 2r\sigma^*\frac{c_\mu}{f_\mu^2}\right\}\dot{c}_\mu.$$

However, the second curly bracket term is equal to $H_{\sigma\sigma}$. By the second-order condition with respect to $\sigma$ to maximize $H$ with a unique solution, we know $H_{\sigma\sigma} < 0$. Note also that the first and third curly bracket terms are nonpositive. Thus, in order to conclude $\dot{\sigma} \geq 0$, it suffices to show $\dot{c}_\mu \leq 0$ and $\dot{\mu}^*<0$.

Note that Equation (A1) implies $\mu^*_\theta(\hat{\theta},\theta)<0$, because $c_{\mu\mu}$, $c_{\mu\theta} > 0$. Therefore, by the differentiability of $\mu^*(\theta)$,

$$\dot{\mu}^*(\theta,\theta) = \mu^*_{\hat{\theta}}(\hat{\theta},\theta) + \mu^*_\theta(\hat{\theta},\theta)|_{\hat{\theta}=\theta} < 0.$$

For the last inequality, we have utilized the fact that $\mu_{\hat{\theta}} \leq 0$ by the $\theta$-implementability. Next, in order to see $\dot{c}_\mu \leq 0$, note that $\dot{c}_\mu = c_{\mu\mu}\dot{\mu}^* + c_{\mu\theta} = c_{\mu\mu}(\mu^*_{\hat{\theta}} + \mu^*_\theta) + c_{\mu\theta} = (c_\mu/f_\mu)f_{\mu\mu}\mu^*_\theta +$ $c_{\mu\mu}\mu^*_{\hat{\theta}}$, by Equation (A1). However since $f_{\mu\mu}=0$ and $\mu^*_{\hat{\theta}} \leq 0$, we have $\dot{c}_\mu = c_{\mu\mu}\mu^*_{\hat{\theta}} \leq 0$. ■

## B Appendix to Section 3.

Sections B.1 and B.2 provide the detailed derivation of the optimal menu. Section B.3 contains the proof of Proposition 2.

### B.1 The agent's problem.

In this section, we consider a class of $Y_1$-measurable salary functions $S(Y_1, \theta)$ that are representable in the following form.

$$S(Y_1,\theta) = F(Y_1,\theta) + \int_0^1 \xi(t, Y_t, \theta)dt + \int_0^1 \eta(t, Y_t, \theta)dB_t, \tag{A2}$$

where $\xi$ and $\eta$ are bounded Borel-measurable functions. In other words, we are interested in $Y_1$-measurable salary functions $S$ that can be expressed in the form of the RHS of (A2).[29] This class is much larger than that of the principal's admissible salary functions described in Section 1. In Section B.2 where we discuss the principal's problem, we will narrow down the above class to the principal's admissible set of salary functions.

We start the analysis of Problem 2 with constraint (ii). Suppose that given a menu of contracts $\{S(Y_1,\theta);\theta \in \Theta\}$, agent $\theta$ chooses a contract $S(Y_1,\hat{\theta})$ and exerts optimal effort

[29] Note that the results in the section will exactly hold, even if S in (A2) is allowed to depend on the whole history of the process $\{Y_t\}$, as in Schättler and Sung (1993).

$\{(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta))\}$. Let $-e^{-rQ(\hat{\theta},\theta)}$ be agent $\theta$'s optimal expected utility at time zero given $S(Y_1,\hat{\theta})$. The quantity $Q(\hat{\theta},\theta)$ is agent $\theta$'s optimal certainty equivalent wealth given $S(Y_1,\hat{\theta})$. When $\hat{\theta}=\theta$, we also write $Q(\theta,\theta)\equiv Q(\theta)$.

It can then be shown that the FOC of agent $\theta$'s problem given by constraint (ii) implies[30]

$$\begin{aligned} S(Y_1,\hat{\theta}) = Q(\hat{\theta},\theta) &+ \int_0^1 c(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta),\theta)dt + \frac{r}{2}\int_0^1 \frac{c_\mu^2(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta),\theta)}{f_\mu^2(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta))}\sigma_t^2(\hat{\theta},\theta)dt \\ &- \int_0^1 \frac{c_\mu(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta),\theta)}{f_\mu(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta))} f(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta))dt + \int_0^1 \frac{c_\mu(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta),\theta)}{f_\mu(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta))}dY_t. \end{aligned} \tag{A3}$$

Note that the LHS of Equation (A3), a member of the optimal menu, is specifically designed for agent $\hat{\theta}$. The RHS describes the agent's optimal responses given contract $S(Y_1,\hat{\theta})$. In other words, if he were given $S(Y_1,\hat{\theta})$, agent $\theta$ would choose effort levels $[(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta))]$ so that the LHS is equal to the RHS.

Under the optimal menu, however, constraint (iii) implies that agent $\theta$ always chooses $S(Y_1,\theta)$. Let $\{(\mu_t(\theta),\sigma_t(\theta))\}$ be the effort levels that the principal *expects* agent $\theta$ to choose given $S(Y_1,\theta)$. That is, the principal expects $\{(\mu_t(\theta),\sigma_t(\theta))\}\equiv\{(\mu_t(\theta,\theta),\sigma_t(\theta,\theta))\}$. Then, the optimal menu consists of contracts of the following form:

$$\begin{aligned} S(Y_1,\theta) = Q(\theta) &+ \int_0^1 c(\mu_t(\theta),\sigma_t(\theta),\theta)dt + \frac{r}{2}\int_2^1 \frac{c_\mu^2(\mu_t(\theta),\sigma_t(\theta),\theta)}{f_\mu^2(\mu_t(\theta),\sigma_t(\theta))}\sigma_t^2(\theta)dt \\ &- \int_0^1 \frac{c_\mu(\mu_t(\theta),\sigma_t(\theta),\theta)}{f_\mu(\mu_t(\theta),\sigma_t(\theta))} f(\mu_t(\theta),\sigma_t(\theta))dt + \int_0^1 \frac{c_\mu(\mu_t(\theta),\sigma_t(\theta),\theta)}{f_\mu(\mu_t(\theta),\sigma_t(\theta))}dY_t. \end{aligned} \tag{A4}$$

Equation (A4) specifies a necessary form of optimal contracts under constraints (ii) and (iii), that is, under moral hazard and adverse selection. However, this form represents too large a family of candidates for optimal contracts. The reason is that contract $S(Y_1,\theta)$ of form (A4) does not always guarantee that agent $\theta$ *actually* chooses $\{(\mu_t(\theta),\sigma_t(\theta))\}$. Note that a menu of contracts of form (A4) can be a candidate for an optimal menu only when given the menu, agent $\theta$ optimally chooses $S(Y_1,\theta)$ and optimally exerts effort $\{(\mu_t(\theta),\sigma_t(\theta))\}$.

We refine this family of contracts of form (A4) in two steps. First, we use the following lemma, which characterizes the necessary and sufficient condition under which given contract $S(Y_1,\theta)$ of form (A4), agent $\theta$ optimally chooses $\{(\mu_t(\theta),\sigma_t(\theta))\}$. We call this condition the $(\mu,\sigma)$-implementability.

**Lemma A.1.** [$(\mu,\sigma)$*-implementability, Sung (1995, Proposition A.1)*]. *Contract* $S(Y_1,\theta)$ *of form* (A4) *is* $(\mu,\sigma)$*-implementable, if and only if for* $t\in[0,1]$ *a.e.,*

$$(\mu_t(\theta),\sigma_t(\theta)) \in \operatorname*{argmax}_{(\mu,\sigma)\in U\times\Sigma} \Phi(\mu,\sigma;\mu_t(\theta),\sigma_t(\theta),\theta) \tag{A5}$$

*where* $\Phi$ *is defined in* (10).

[30] If $f_\mu\equiv 1$, and if $\sigma_t$ is constant, then (A3) reduces to Holmstrom-Milgrom's (1987) original representation of the salary function. Schättler and Sung (1993, Corollary 4.1) proved Equation (A3) by using a martingale method.

Lemma A.1 defines all possible levels of agent $\theta$'s effort that the principal can induce through the optimal menu of form (A4). Motivated by Lemma A.1, let us define the $(\mu, \sigma)$-implementable control set $\overline{U}(\theta) \times \overline{\Sigma}(\theta)$ for each $\theta$ as follows:

$$\overline{U}(\theta) \times \overline{\Sigma}(\theta) := \left\{ (\mu, \sigma) \middle| (\mu, \sigma) \in \underset{(\mu', \sigma') \in U \times \Sigma}{\operatorname{argmax}} \Phi(\mu', \sigma'; \mu, \sigma, \theta) \right\}. \tag{A6}$$

Then Equation (A4) defined on $\overline{U}(\theta) \times \overline{\Sigma}(\theta)$ characterizes all contracts that satisfy the $(\mu, \sigma)$-implementability. In other words, given $S(Y_1,\theta)$ of form Equation (A4) with $\{(\mu_t(\theta), \sigma_t(\theta)) \in \overline{U}(\theta) \times \overline{\Sigma}(\theta)\}$, agent $\theta$ optimally chooses effort levels so that $\{(\mu_t(\theta,\theta),\sigma_t(\theta,\theta))\} \equiv \{(\mu_t(\theta), \sigma_t(\theta))\}$.

For the second step of the refinement, note that the representation in Equation (A4) defined on $\overline{U}(\theta) \times \overline{\Sigma}(\theta)$, however, is still incomplete because the certainty equivalent wealth $Q(\theta)$ is not specified. Theorem A.1 reveals a necessary structure of $Q(\theta)$ to satisfy constraint (iii) and (iv). For economy of notation, let $Q(\hat{\theta}), c(\mu_t(\hat{\theta}), \sigma_t(\hat{\theta}), \hat{\theta})$, $\mu_t(\hat{\theta})$, and $\sigma_t(\hat{\theta})$ be equivalently denoted by $\hat{Q}, \hat{c}, \hat{\mu}_t$, and $\hat{\sigma}_t$, respectively.

**Theorem A.1.** *Let $\{S(Y_1, \theta)\}$ be the optimal menu of contracts of form* (A4)*. Assume that for all $t \in [0, 1]$, $\mu_t(\hat{\theta}, \theta)$ and $\sigma_t(\hat{\theta}, \theta)$ are differentiable in $(\hat{\theta}, \theta)$ a.e. Then constraints (iii) and (iv) imply that for all $\theta$,*

$$Q(\theta) = W_0 + \int_{\theta}^{\theta_H} E^B \left[ \int_0^1 c_\theta(\mu_t, \sigma_t, \theta') L_t(\theta') dt \right] d\theta', \tag{A7}$$

*where*

$$L_t(\theta) = \exp\left\{ -\frac{1}{2} \int_0^1 \left( r \frac{c_\mu(\mu_t, \sigma_t, \theta)}{f_\mu(\mu_t, \sigma_t)} \sigma_t(\theta) \right)^2 dt - \int_0^1 \left( r \frac{c_\mu(\mu_t, \sigma_t, \theta)}{f_\mu(\mu_t, \sigma_t)} \sigma_t(\theta) \right) dB_t \right\},$$

*where $(\mu_t, \sigma_t, \theta')$, $(\mu_t, \sigma_t, \theta)$, and $L_t(\theta)$ are short for $(\mu_t(\theta'), \sigma_t(\theta'), \theta')$, $(\mu_t(\theta), \sigma_t(\theta), \theta)$, and $L_t(\{(\mu_s(\theta), \sigma_s(\theta)); 0 \le s \le t\}, \theta)$, respectively. (Note that since $\{S(Y_1, \theta)\}$ is optimal, $[\mu_t(\theta), \sigma_t(\theta)] \equiv [\mu_t(\theta,\theta), \sigma_t(\theta,\theta)]$.)*

**Proof.** Suppose that agent $\theta$ has chosen $S(Y_1, \hat{\theta})$ satisfying (A4). Then the agent's expected utility is

$$\begin{aligned} u(\hat{\theta}; \theta) &= \max_{\{\mu_t, \sigma_t\}} E^B \Bigg( -\exp\Bigg\{ -r \Bigg[ \hat{Q} + \int_0^1 \left( \hat{c} - \frac{\hat{c}_\mu}{\hat{f}_\mu} \hat{f} + \frac{r}{2} \frac{\hat{c}_\mu^2}{\hat{f}_\mu^2} \hat{\sigma}_t^2 \right) dt \\ &\quad + \int_0^1 \frac{\hat{c}_\mu}{\hat{f}_\mu} (f dt + \sigma_t dB_t) - \int_0^1 c(\mu_t, \sigma_t, \theta) dt \Bigg] \Bigg\} \Bigg) \\ &= e^{-r\hat{Q}} \max_{\{\mu_t, \sigma_t\}} E^B \left\{ -\exp\left[ -r \int_0^1 \Psi(\mu_t, \sigma_t, \hat{\theta}, \theta) dt \right] \tilde{L}_1(\hat{\theta}, \theta) \right\}, \end{aligned} \tag{A8}$$

where

$$\Psi(\mu_t, \sigma_t, \hat{\theta}, \theta) := \hat{c} - c(\mu_t, \sigma_t, \theta) - \frac{\hat{c}_\mu}{\hat{f}_\mu} (\hat{f} - f(\mu_t, \sigma_t)) + \frac{r}{2} \frac{\hat{c}_\mu^2}{\hat{f}_\mu^2} (\hat{\sigma}_t^2 - \sigma_t^2) \tag{A9}$$

and

$$\tilde{L}_t(\hat{\theta},\theta) = \exp\left[-\frac{1}{2}\int_0^t \left(-r\frac{\hat{c}_\mu}{\hat{f}_\mu}\sigma_t\right)^2 dt + \int_0^t \left(-r\frac{\hat{c}_\mu}{\hat{f}_\mu}\sigma_t\right) dB_t\right].$$

Let $\{(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta))\}$ be the optimal levels of effort of agent $\theta$, given $S(Y_1,\hat{\theta})$ from the optimal menu. That is, $\{(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta))\}$ solves the maximization problem in (A8), satisfying the incentive compatibility condition (ii). By the truthtelling condition (iv), for all $\theta,\hat{\theta}\in[\theta_L,\theta_H]$,

$$u(\theta;\theta) \geq u(\hat{\theta};\theta).$$

Since given $\theta$, the above should hold for all $\hat{\theta}\in[\theta_L,\theta_H]$, we have

$$u(\theta;\theta) = \max_{\hat{\theta}} u(\hat{\theta};\theta).$$

That is,

$$-e^{-rQ(\theta)} = \max_{\hat{\theta}} E^B\left\{-\exp\left[-r\left(Q(\hat{\theta}) + \int_0^1 \Psi(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta),\hat{\theta},\theta)dt\right)\right]L_1(\hat{\theta},\theta)\right\},$$

where $L_1(\hat{\theta},\theta)$ is $\tilde{L}_1(\hat{\theta},\theta)$ evaluated at $(\mu_t,\sigma_t) = (\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta))$.

Momentarily, let us assume that $Q$ is differentiable. (We will later show that $Q$ is indeed differentiate a.e.) By the uniform integrability, the first-order condition with respect to $\hat{\theta}$ evaluated at $\hat{\theta}=\theta$ is

$$\begin{aligned}
&E^B\left(-\exp\left\{-r\left[Q(\theta) + \int_0^1 \Psi(\mu_t(\theta,\theta),\sigma_t(\theta,\theta),\theta,\theta)dt\right]\right\}\right.\\
&\quad\left.\times\left\{\left[-r\frac{dQ}{d\theta}(\theta) - r\int_0^1 \frac{\partial\Psi}{\partial\hat{\theta}}(\mu_t(\theta,\theta),\sigma_t(\theta,\theta),\theta,\theta)dt\right]L_1(\theta,\theta) + \frac{\partial L_1(\theta,\theta)}{\partial\hat{\theta}}\right\}\right) = -\exp\{\\
&\quad -rQ(\theta)\}E^B\left\{\left[-r\frac{dQ}{d\theta}(\theta) - r\int_0^1 \frac{\partial\Psi}{\partial\hat{\theta}}(\mu_t(\theta,\theta),\sigma_t(\theta,\theta),\theta,\theta)dt\right]L_1(\theta,\theta) + \frac{\partial L_1(\theta,\theta)}{\partial\hat{\theta}}\right\}\\
&\quad = r\exp\{-rQ(\theta)\}E^B\left\{\left[\frac{dQ}{d\theta}(\theta) + \int_0^1 \frac{\partial\Psi}{\partial\hat{\theta}}(\mu_t(\theta,\theta),\sigma_t(\theta,\theta),\theta,\theta)dt\right]L_1(\theta,\theta)\right\} = 0,
\end{aligned}$$

where $(\partial\Psi/\partial\hat{\theta})(\mu_t(\theta,\theta),\sigma_t(\theta,\theta),\theta,\theta) = (\partial\Psi/\partial\hat{\theta})(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta),\hat{\theta},\theta)|_{\hat{\theta}=\theta}$, and $\partial L_1(\theta,\theta)/\partial\hat{\theta} = \partial L_1(\hat{\theta},\theta)/\partial\hat{\theta}|_{\hat{\theta}=\theta}$. For the above, we have utilized the fact that the derivative of the expected value of the Girsanov density $L_1$with respect to $\hat{\theta}$ is always zero and that $\Psi(\mu_t(\theta,\theta),\ \sigma_t(\theta,\theta),\theta,\theta)=0$. (Note: Since $f_\mu$ is bounded away from zero and since $c_\mu$ is bounded, we have $E^B[L_1]=1$. See Karatzas and Shreve (1987, p. 200, Corollary 5.16). Thus, $\lim_{\Delta\hat{\theta}\to 0}(1/\Delta\hat{\theta})(E[L_1(\hat{\theta},\theta) - L_1(\theta,\theta)]) = 0$. The uniform integrability follows because $E^B[L_1^p]<\infty$ for $p<\infty$. See Chung (1974, Theorems 4.5.2 and 4.5.4).) However,

$$\frac{\partial}{\partial\hat{\theta}}\Psi(\mu_t(\hat{\theta},\theta),\sigma_t(\hat{\theta},\theta),\hat{\theta},\theta)|_{\hat{\theta}=\theta} = c_\theta(\mu_t,\sigma_t,\theta), \tag{A10}$$

where the RHS of the above equation is the partial derivative of $c$ with respect to its third argument. To see this,

$$\begin{aligned}\frac{\partial\Psi}{\partial\hat\theta}(\mu(\hat\theta,\theta),\sigma(\hat\theta,\theta),\hat\theta,\theta) &= \hat c_\theta + \hat c_\mu\frac{d\hat\mu}{d\hat\theta} + \hat c_\sigma\frac{d\hat\sigma}{d\hat\theta} - c_\mu\frac{\partial\mu}{\partial\hat\theta} - c_\sigma\frac{\partial\sigma}{\partial\hat\theta} - \left(\hat f - f(\mu_t,\sigma_t)\right)\frac{d}{d\hat\theta}\left(\frac{\hat c_\mu}{\hat f_\mu}\right)\\
&\quad - \left(\frac{\hat c_\mu}{\hat f_\mu}\right)\left(\hat f_\mu\frac{d\hat\mu}{d\hat\theta} + \hat f_\sigma\frac{d\hat\sigma}{d\hat\theta} - f_\mu\frac{\partial\mu}{\partial\hat\theta} - f_\sigma\frac{\partial\sigma}{\partial\hat\theta}\right) + \frac{r}{2}(\hat\sigma_t^2-\sigma_t^2)\frac{d}{d\hat\theta}\left(\frac{\hat c_\mu^2}{\hat f_\mu^2}\right)\\
&\quad + r\frac{\hat c_\mu^2}{\hat f_\mu^2}\left(\hat\sigma_t\frac{d\hat\sigma}{d\hat\theta} - \sigma_t\frac{\partial\sigma}{\partial\hat\theta}\right)\\
&= \hat c_\theta + \left[\hat c_\sigma - \left(\frac{\hat c_\mu}{\hat f_\mu}\right)\hat f_\sigma + r\hat\sigma_t\frac{\hat c_\mu^2}{\hat f_\mu^2}\right]\frac{d\hat\sigma}{d\hat\theta} - \left[c_\mu - \left(\frac{\hat c_\mu}{\hat f_\mu}\right)f_\mu\right]\frac{\partial\mu}{\partial\hat\theta}\\
&\quad - \left[c_\sigma - \left(\frac{\hat c_\mu}{\hat f_\mu}\right)f_\sigma + r\sigma_t\frac{\hat c_\mu^2}{\hat f_\mu^2}\right]\frac{\partial\sigma}{\partial\hat\theta}\\
&\quad - [\hat f - f(\mu_t,\sigma_t)]\frac{d}{d\hat\theta}\left(\frac{\hat c_\mu}{\hat f_\mu}\right) + \frac{r}{2}(\hat\sigma_t^2-\sigma_t^2)\frac{d}{d\hat\theta}\left(\frac{\hat c_\mu^2}{\hat f_\mu^2}\right).\end{aligned}$$

When $\hat\theta=\theta$, we have $\hat f = f$, $\hat\sigma=\sigma$, and $\hat\mu=\mu$, and thus the FOC for (A5) with respect to $\sigma$ [or the $(\mu,\sigma)$-implementability] implies that the above equation reduces to Equation (A10).

Furthermore, since $\Psi(\mu(\theta,\theta), \sigma(\theta,\theta),\theta,\theta)=0$ and $Q(\theta)$ is independent of the Brownian motion $B$, the above first-order condition reduces to

$$\begin{aligned}0 &= \frac{dQ}{d\theta}(\theta) + E^B\left[L_1(\theta)\int_0^1 c_\theta(\mu_t(\theta,\theta),\sigma_t(\theta,\theta),\theta)dt\right]\\
&= \frac{dQ}{d\theta}(\theta) + E^B\left[\int_0^1 c_\theta(\mu_t(\theta,\theta),\sigma_t(\theta,\theta),\theta)L_t(\theta)dt\right],\end{aligned}\tag{A11}$$

where $L_t(\theta)=L_t(\theta,\theta)$ which is the same as the definition following Equation (A7). The last equality is a standard result that can be easily established using Itô's formula and the fact that $L_t$ is a martingale.

To justify Equation (A11), we still need to check the differentiability of $Q$. In fact, we claim that $Q$ is nonincreasing in $\theta$ and thus differentiable with respect to $\theta$ a.e. Constraint (iii) implies that given $\theta$, we have for all $\hat\theta$ and admissible $\{(\mu_t,\sigma_t)\}$

$$-e^{-rQ(\theta)} \ge E^B\left\{-\exp\left[-r\left(Q(\hat\theta)+\int_0^1\Psi(\mu_t,\sigma_t,\hat\theta,\theta)dt\right)\right]L_1(\hat\theta,\theta)\right\}.\tag{A12}$$

Choose $\theta$ and $\hat\theta$ such that $\theta<\hat\theta$. Since Equation (A12) holds for all admissible $\{(\mu_t,\sigma_t)\}$,

$$-e^{-rQ(\theta)} \ge -e^{-rQ(\hat\theta)}E^B\left\{-\exp\left[-r\int_0^1\Psi(\hat\mu_t,\hat\sigma_t,\hat\theta,\theta)dt\right]L_1(\hat\theta,\theta)\right\} \ge -e^{-rQ(\hat\theta)}.$$

For the last inequality, note

$$\Psi(\hat\mu_t,\hat\sigma_t\hat\theta,\theta) = \Psi(\hat\mu_t,\hat\sigma_t,\hat\theta,\hat\theta) + c(\hat\mu_t,\hat\sigma_t,\hat\theta) - c(\hat\mu_t,\hat\sigma_t,\theta) \ge 0,$$

because $\Psi(\hat\mu_t,\hat\sigma_t,\hat\theta,\hat\theta)=0$ and $c_\theta\ge 0$. Therefore, $Q(\theta)\ge Q(\hat\theta)$ for $\theta<\hat\theta$ and $Q(\theta)$ is differentiable a.e.

Finally, since $Q$ is nonincreasing in $\theta$, it suffices to have $Q(\theta_H) \geq W_0$ in order to satisfy constraint (iv) for all $\theta$. Since $Q$ values do not affect the agent's effort decisions, we have $Q(\theta_H) = W_0$ to maximize the principal's wealth. Therefore, Equation (A11) and $Q(\theta_H) = W_0$ imply Equation (A7). ■

We will use Equation (A7) to simplify the principal's problem of adverse selection. As an immediate corollary, the representation of the salary function under moral hazard by Holmstrom and Milgrom (1987) and Schättler and Sung (1993) can be modified as follows.

**Corollary A.1.** *Under adverse selection and moral hazard, the salary function* $S(Y_1, \theta)$ *is represented as follows*:

$$S(Y_1,\theta) = W_0 + \int_\theta^{\theta_H} E^B\left[\int_0^1 c_\theta(\mu_t,\sigma_t,\theta')L_t(\theta')dt\right]d\theta' + \int_0^1 c(\mu_t,\sigma_t,\theta)dt + \frac{r}{2}\int_0^1 \frac{c_\mu^2(\mu_t,\sigma_t,\theta)}{f_\mu^2(\mu_t,\sigma_t)}\sigma_t^2(\theta)dt$$
$$-\int_0^1 \frac{c_\mu(\mu_t,\sigma_t,\theta)}{f_\mu(\mu_t,\sigma_t)}f(\mu_t,\sigma_t)dt + \int_0^1 \frac{c_\mu(\mu_t,\sigma_t,\theta)}{f_\mu(\mu_t,\sigma_t)}dY_t. \tag{A13}$$

The above two-step refinement suggests that Equation (A13) defined on $\overline{U}(\theta) \times \overline{\Sigma}(\theta)$ represents a necessary condition to satisfy constraints (ii), (iii), and (iv) of Problem 2. It is also sufficient for constraints (ii) and (iv). Later in Theorem A.3, we will discuss a sufficient condition for constraint (iii). Now we are ready to move on to the principal's problem.

**B.2 The principal's problem.** The results thus far have been obtained regardless of the principal's risk preferences. Now we assume the principal is risk-neutral. We relax the principal's problem by replacing the original constraints (ii), (iii), and (iv) with Equation (A13) defined on $\overline{U}(\theta) \times \overline{\Sigma}(\theta)$.

Then the risk-neutral principal's problem in a relaxed form can be stated as follows:

**Problem 2′.** *Choose $G_0$-adapted controls* $\{(\mu_t(\theta),\sigma_t(\theta)) \in \overline{U}(\theta)\times\overline{\Sigma}(\theta)\};\ \theta \in \Theta\}$ *to maximize*

$$\int_{\theta_L}^{\theta_H} E^{Y_1}\left[Y_1 - Q(\theta) - \int_0^1\left(c + \frac{r}{2}\frac{c_\mu^2(\mu_t(\theta),\sigma_t(\theta),\theta)}{f_\mu^2(\mu_t(\theta),\sigma_t(\theta))}\sigma_t^2(\theta)\right)dt\right]h(\theta)d\theta,$$

$$s.t.\ \text{(i)}\quad \forall\theta, dY_t = f(\mu_t(\theta),\sigma_t(\theta))dt + \sigma_t(\theta)dB_t,$$

$$\text{(ii)}\ \forall\theta, Q(\theta) = W_0 + \int_\theta^{\theta_H} E^B\left[\int_0^1 c_\theta(\mu_t(\theta'),\sigma_t(\theta'),\theta')L_t(\theta')d\right]d\theta',$$

*where $E^{Y_1}$ is the expectation operator with respect to $Y_1$ conditional on $\theta$.*

A close inspection of Problem 2′ reveals that it is a mixture of deterministic and stochastic control problems: deterministic in $\theta$ and stochastic in $\{Y_t\}$. The problem can be greatly simplified because the principal can only observe $\{Y_0, Y_1\}$. Since $G_0 \equiv G_t$ for $t \in [0,1)$, implementable controls from the principal's perspective are restricted to be $G_0$-measurable.

**Theorem A.2.** *For each* $(t, \theta) \in [0,1] \times \Theta$ *a.e., optimal controls for Problem 2′ are given by* $(\mu_t^*(\theta),\sigma_t^*(\theta)) \in \text{argmax}_{(\mu,\sigma)\in\overline{U}(\theta)\times\overline{\Sigma}(\theta)}H(\mu,\sigma,\theta)$, *where $H$ is as defined in* (2). *That is, the optimal controls are constant over time.*

**Proof.** Since the principal observes only $Y_1$, the resulting salary function for each $\theta$ has to be both $Y_1$-measurable and ($\mu$, $\sigma$)-implementable for each $\theta$. Since $G_0 \equiv F_0$, constraint (ii) is equivalent to

$$\begin{aligned} Q(\theta) &= W_0 + \int_{\theta}^{\theta_H} E^{Y_1}\left[\int_0^1 c_\theta(\mu_t(\theta'), \sigma_t(\theta'), \theta') L_t(\theta') dt\right] d\theta' \\ &= W_0 + \int_{\theta}^{\theta_H}\int_0^1 c_\theta(\mu_t(\theta'), \sigma_t(\theta'), \theta') dt d\theta'. \end{aligned}$$

For the last equality we have used Fubini's theorem, the fact that ($\mu_t(\theta')$, $\sigma_t(\theta')$) are $G_0$-measurable and $E[L_t] = 1$.

Then Problem 2′ can be solved in two steps. First we separate the stochastic control problem from the deterministic control problem as follows. Let the Hamiltonian be

$$\begin{aligned} &H(Q(\theta), \{\mu_t\}, \{\sigma_t\}, \theta, \lambda) : \\ &= E^{Y_1}\left[Y_1 - Q(\theta) - \int_0^1 \left(c + \frac{r}{2}\frac{c_\mu^2}{f_\mu^2}\sigma_t^2\right) dt\right] h(\theta) - \lambda E^{Y_1}\left[\int_2^1 c_\theta(\mu_t, \sigma_t, \theta) dt\right]. \end{aligned} \tag{A14}$$

Then by using the Pontryagin maximum principle, the necessary conditions are

$$\frac{d\lambda}{d\theta} = -\frac{\partial H}{\partial Q} = h(\theta), \tag{A15}$$

$$\lambda(\theta_L) = 0, \tag{A16}$$

$$\{(\mu_t^*, \sigma_t^*)\} \in \underset{\{(\mu_t,\sigma_t) \in \overline{U}(\theta)\times\overline{\Sigma}(\theta)\}}{\operatorname{argmax}} H(Q, \{\mu_t\}, \{\sigma_t\}, \theta, \lambda). \tag{A17}$$

By (A15) and (A16), we have $\lambda(\theta) = \int_{\theta_L}^{\theta} h(\theta') d\theta'$.

Now, we need to solve the stochastic control problem (A17). However, note that since ($\mu_t$, $\sigma_t$), $\forall t \in$ [0,1), is $G_0$-measurable, the Hamiltonian in (A14) can be restated as follows:

$$\begin{aligned} H(Q(\theta), \{\mu_t\}, \{\sigma_t\}, \theta, \lambda) &= h(\theta)\int_0^1 \left[f - Q(\theta) - \left(c + \frac{r}{2}\frac{c_\mu^2}{f_\mu^2}\sigma_t^2\right)\right] dt - \lambda\int_2^1 c_\theta(\mu_t, \sigma_t, \theta) dt \\ &= h(\theta)\int_\theta^1 H(\mu_t, \sigma_t, \theta) dt - h(\theta) Q(\theta), \end{aligned}$$

where $H(.)$ is given by Equation (2). Then, it becomes a simple pointwise maximization problem. Thus, $(\mu_t^*(\theta), \sigma_t^*(\theta)) \in \operatorname{argmax}_{(\mu,\sigma)\in\overline{U}(\theta)\times\overline{\Sigma}(\theta)} H(\mu_t, \sigma_t, \theta)$ are optimal controls for all $t \in$ [0,1]. ■

By formulation, it should be clear that if solutions to Problem 2′ for all $\theta$ satisfy the truthtelling constraint (iii) of Problem 2, then they also solve Problem 2. That is, if for each $\theta$, the control pair $(\mu_t^*(\theta), \sigma_t^*(\theta)) \in \operatorname{argmax}_{(\mu,\sigma)\in\overline{U}(\theta)\times\overline{\Sigma}(\theta)} H(\mu_t, \sigma_t, \theta)$ is substituted into the salary function for $\theta$ in Equation (A13), and if for all $\theta$ the menu of the salary functions satisfy the truthtelling constraint (iii) of Problem 2, then the control pairs $(\mu_t^*(\theta), \sigma_t^*(\theta))$'s for all $\theta$ and the resulting menu are optimal solutions to Problem 2. Moreover, since the optimal controls are constant over time for all $\theta$, the menu consists of contracts that are linear in $Y_1$. Of course, if $\operatorname{argmax}_{(\mu,\sigma)\in\overline{U}(\theta)\times\overline{\Sigma}(\theta)} H(\mu_t, \sigma, \theta)$ is singleton for all $\theta$, then the optimal controls are unique for all $\theta$ and so is the menu.

In order to compute optimal control pairs $(\mu_t^*(\theta), \sigma_t^*(\theta))$ for all $\theta$, note that the constrained maximization of the Hamiltonian, that is, $\max_{(\mu,\sigma)\in\overline{U}(\theta)\times\overline{\Sigma}(\theta)} H(\mu_t, \sigma, \theta)$, can be relaxed to the maximization of Lagrangian (9) over $U \times \Sigma$. The reason is as follows. By Lemma A.1 or by the set defined in (A6),

$$\max_{(\mu,\sigma)\in\overline{U}(\theta)\times\overline{\Sigma}(\theta)} H(\mu,\sigma,\theta) = \begin{cases} \max_{(\mu,\sigma)\in U\times\Sigma} H(\mu,\sigma,\theta) \\ s.t.\ (\mu,\sigma) \in \operatorname{argmax}_{(\mu',\sigma')\in U\times\Sigma} \Phi(\mu',\sigma';\mu,\sigma,\theta) \end{cases}$$

Thus, using the agent's FOCs, the RHS can be relaxed to

$$\begin{array}{cc} \max & H(\mu,\sigma,\theta) \\ (\mu,\sigma)\in U\times\Sigma & \Phi_\mu(\mu,\sigma;\mu,\sigma,\theta)=0 \\ s.t. & \Phi_\sigma(\mu,\sigma;\mu,\sigma,\theta)=0. \end{array}$$

However, $\Phi_\mu(\mu, \sigma; \mu, \sigma, \theta) = 0$ is trivially satisfied. Therefore, the above relaxed problem is equivalent to

$$\begin{array}{cc} \max & \\ (\mu,\sigma)\in U\times\Sigma & H(\mu,\sigma,\theta) \\ s.t. & \Phi_\sigma(\mu,\sigma;\mu,\sigma,\theta)=0. \end{array}$$

Hence, we have Lagrangian (9).

The next step is to check whether $\{S^*(Y_1,\theta); \theta \in \Theta\}$ satisfies the truthtelling constraint which requires that given the optimal menu, agent $\theta \in \Theta$ should optimally choose $S^*(Y_1,\theta)$. We call this requirement the θ-implementability. To establish the sufficiency for the θ-implementability, the following lemma will be useful.

**Lemma A.2.** *Suppose that agent $\theta$ is given any linear contract $S(Y_1,\hat{\theta})$ of form* (3) *with* $\{(\mu(\theta), \sigma(\theta));\ \theta \in \Theta)\}$. (*Note that S does not have to be $S^*$, that is, $(\mu(\theta), \sigma(\theta), \theta)$ is not required to maximize the principal's Lagrangian given $\theta$.*) *Then the agent's optimal responses $\{(\mu_t(\hat{\theta},\theta), \sigma_t(\hat{\theta},\theta))\}$ are constant over time such that for all t, $(\mu_t(\hat{\theta},\theta), \sigma_t(\hat{\theta},\theta)) = (\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta))$, where*

$$(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta)) \in \operatorname*{argmax}_{(\mu',\sigma')\in U\times\Sigma} \Phi(\mu',\sigma';\hat{\mu},\sigma,\hat{\theta};\theta),$$

*and $\Phi$ is as defined in* (10).

**Proof.** Given $S(Y_1,\hat{\theta})$, agent $\theta$ solves the following problem.

$$\max_{\{\mu_t',\sigma_t'\}} E^B\left\{-\exp\left[-r\left(\hat{Q}+\hat{c}-\frac{\hat{c}_\mu}{\hat{f}_\mu}\hat{f}+\frac{r}{2}\frac{\hat{c}_\mu^2}{\hat{f}_\mu^2}\hat{\sigma}^2+\frac{\hat{c}_\mu}{\hat{f}_\mu}Y_1-\int_0^1 c(\mu_t',\sigma_t',\theta)dt\right)\right]\right\},$$

$$s.t.\ dY_t = f(\mu_t',\sigma_t')dt + \sigma_t' dB_t.$$

The Hamilton-Jacobi-Bellman (HJB) equation for the above problem [see Schättler and Sung (1993, Theorem 5.1) or Sung (1992, Lemma A.1)] is

$$0 \equiv \frac{\partial V}{\partial t} + \max_{(\mu_t',\sigma_t')}\left[\frac{\partial V}{\partial y}f(\mu_t',\sigma_t') + \frac{1}{2}\frac{\partial^2 V}{\partial y^2}\sigma_t'^2 + rc(\mu_t',\sigma_t',\theta)V\right]$$

with the terminal condition

$$V(1,y) = -\exp\left[-r\left(\hat{Q}+\hat{c}-\frac{\hat{c}_\mu}{\hat{f}_\mu}\hat{f}+\frac{r}{2}\frac{\hat{c}_\mu^2}{\hat{f}_\mu^2}\hat{\sigma}^2+\frac{\hat{c}_\mu}{\hat{f}_\mu}y\right)\right].$$

Then one can show that the following function satisfies the above HJB equation:

$$V(t, Y_t) = -\exp\left[-r\left(\hat{Q} + \hat{c} - \frac{\hat{c}_\mu}{\hat{f}_\mu}\hat{f} + \frac{r}{2}\frac{\hat{c}_\mu^2}{\hat{f}_\mu^2}\hat{\sigma}^2 + \frac{\hat{c}_\mu}{\hat{f}_\mu} Y_t + (1-t)\Phi(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta); \hat{\mu}, \hat{\sigma}, \hat{\theta}; \theta)\right)\right].$$

Thus, by the verification theorem [Sung (1992, Lemma A.2)], the assertion follows. ■

Lemma A.2 tells us that given any linear contract of form (3), agent $\theta$ still exerts constant levels of effort on $\mu$ and $\sigma$ throughout the whole contracting period. Thus, for all $t \in [0,1]$, we write $(\mu_t(\hat{\theta},\theta), \sigma_t(\hat{\theta},\theta))$ satisfying Equation (A18) as $(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta))$ without the time subscript. In particular, agent $\theta$'s responses given $S^*(Y_1, \hat{\theta})$ are written as $(\mu^*(\hat{\theta},\theta), \sigma^*(\hat{\theta},\theta))$.

The following theorem gives the necessary and sufficient condition for a linear menu to satisfy the $\theta$-implementability.

**Theorem A.3.** ($\theta$-*implementability*) *Let a menu of* ($\mu$, $\sigma$)-*implementable linear contracts be* $\{S(Y_1, \theta); \theta \in \Theta\}$, *each of which is given in the form of Equation* (3) *with* $[(\mu(\theta), \sigma(\theta)); \theta \in \Theta]$. (*Note S does not have to be* $S^*$.) *The menu satisfies condition* (iii) *of Problem 2 if and only if for all* $(\hat{\theta},\theta)$, *condition* (15) *holds, that is,*

$$c_{\theta\mu}(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta), \theta)\mu_{\hat{\theta}}(\hat{\theta},\theta) + c_{\theta\sigma}(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta), \theta)\sigma_{\hat{\theta}}(\hat{\theta},\theta) \le 0.$$

**Proof.** For Equation (3) to satisfy condition (iii) of Problem 2, agent θ should not be able to find a contract $S(Y_1, \hat{\theta}), \hat{\theta} \neq \theta$ such that $u(\hat{\theta};\theta) > u(\theta;\theta)$.

***Sufficiency.*** Suppose that under condition (15), there is $S(Y_1, \hat{\theta})$ such that $u(\hat{\theta};\theta) > u(\theta;\theta)$. Without loss of generality, we assume that $\theta < \hat{\theta}$, that is, the agent pretends to be a less efficient agent $\hat{\theta}$. Given $S(Y_1, \hat{\theta})$, the agent's optimal response $\mu(\hat{\theta},\theta)$ is, by Lemma A.2, constant over time, and the expected utility of agent $\theta$ is $u(\hat{\theta};\theta)$, where

$$\begin{aligned} u(\hat{\theta},\theta) &= E^B\left\{-\exp\left[-r\left(\hat{Q} + \hat{c} - \frac{\hat{c}_\mu}{\hat{f}_\mu}\hat{f} + \frac{r}{2}\frac{\hat{c}_\mu^2}{\hat{f}_\mu^2}\hat{\sigma}^2 + \frac{\hat{c}_\mu}{\hat{f}_\mu} Y_1 - c(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta), \theta)\right)\right]\right\} \\ &= -\exp\left\{-r(\hat{Q} + \Psi(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta), \hat{\theta}, \theta)\right\}, \end{aligned}$$

where $\Psi$ is defined by Equation (A9), and $\mu(\hat{\theta},\theta)$ and $\sigma(\hat{\theta},\theta)$ satisfy (A18). That is,

$$(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta)) \in \underset{(\mu,\sigma)}{\operatorname{argmax}}\, \Psi(\mu, \sigma, \hat{\theta}, \theta).$$

Since $u(\hat{\theta};\theta) > u(\theta;\theta)$ by the hypothesis,

$$u(\theta;\theta) = -e^{-rQ(\theta)} < -\exp\left[-r\left(\hat{Q} + \Psi(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta), \hat{\theta}, \theta)\right)\right],$$

or

$$Q(\theta) - Q(\hat{\theta}) < \Psi(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta), \hat{\theta}, \theta).$$

But since $\Psi(\mu(\hat{\theta},\hat{\theta}), \sigma(\hat{\theta},\hat{\theta}), \hat{\theta}, \hat{\theta}) = 0$,

$$\begin{aligned} &\Psi(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta), \hat{\theta}, \theta) = \Psi(\mu(\hat{\theta},\theta), \sigma(\hat{\theta},\theta), \hat{\theta}, \theta) - \Psi(\mu(\hat{\theta},\hat{\theta}), \sigma(\hat{\theta},\hat{\theta}), \hat{\theta}, \hat{\theta}) \\ &= -\int_\theta^{\hat{\theta}} [\Psi_\mu(\mu(\hat{\theta},\theta'), \sigma(\hat{\theta},\theta'), \hat{\theta}, \theta')\mu_{\theta'}(\hat{\theta},\theta') + \Psi_\sigma(\mu(\hat{\theta},\theta'), \sigma(\hat{\theta},\theta'), \hat{\theta}, \theta')\sigma_{\theta'}(\hat{\theta},\theta') \end{aligned}$$

$$+\Psi_{\theta'}(\mu(\hat{\theta},\theta'),\sigma(\hat{\theta},\theta'),\hat{\theta},\theta')]\ d\theta' = -\int_{\theta}^{\hat{\theta}} \Psi_{\theta'}(\mu(\hat{\theta},\theta'),\sigma(\hat{\theta},\theta'),\hat{\theta},\theta')\ d\theta'$$

$$= \int_{\theta}^{\hat{\theta}} c_{\theta}(\mu(\hat{\theta},\theta'),\sigma(\hat{\theta},\theta'),\theta')\ d\theta'.$$

For the above, we have utilized the FOCs, that is,

$$\Psi_{\mu}(\mu(\hat{\theta},\theta'),\sigma(\hat{\theta},\theta'),\hat{\theta},\theta') = \Psi_{\sigma}(\mu(\hat{\theta},\theta'),\sigma(\hat{\theta},\theta'),\hat{\theta},\theta') = 0.$$

Therefore, we must have

$$Q(\theta) - Q(\hat{\theta}) = \int_{\theta}^{\hat{\theta}} c_{\theta}(\mu(\hat{\theta}),\sigma(\theta'),\theta')\ d\theta' < \int_{\theta}^{\hat{\theta}} c_{\theta}(\mu(\hat{\theta},\theta'),\sigma(\hat{\theta},\theta'),\theta')\ d\theta'.$$

The LHS of the above is from (A7) and the fact that $\mu$ and $\sigma$ are constant by Theorem A.2. Moreover, by the ($\mu$, $\sigma$)-implementability, we have $\mu(\theta')=\mu(\theta', \theta')$ and $\sigma(\theta')=\sigma(\theta', \theta')$. Thus, we have

$$\int_{\theta}^{\hat{\theta}} \left[c_{\theta}(\mu(\theta'),\sigma(\theta'),\theta') - c_{\theta}(\mu(\hat{\theta},\theta'),\sigma(\hat{\theta},\theta'),\theta')\right] d\theta'$$

$$= -\int_{\theta}^{\hat{\theta}}\int_{\theta'}^{\hat{\theta}} \left[c_{\theta\mu}(\mu(\tilde{\theta},\theta'),\sigma(\tilde{\theta},\theta'),\hat{\theta})\mu_{\tilde{\theta}}(\tilde{\theta},\theta') + c_{\theta\sigma}(\mu(\tilde{\theta},\theta'),\sigma(\tilde{\theta},\theta'),\theta')\sigma_{\tilde{\theta}}(\tilde{\theta},\theta')\right] d\tilde{\theta} d\theta' < 0,$$

contradicting condition (15).

***Necessity.*** Condition (iii) of Problem 2 implies that for all $(\hat{\theta},\theta), u(\hat{\theta};\theta) \le u(\theta;\theta)$. Then, by using essentially the same argument as in the proof of the sufficiency, we have

$$\int_{\theta}^{\hat{\theta}}\int_{\theta'}^{\hat{\theta}} [c_{\theta\mu}(\mu(\tilde{\theta},\theta'),\sigma(\tilde{\theta},\theta'),\hat{\theta})\mu_{\tilde{\theta}}(\tilde{\theta},\theta') + c_{\theta\sigma}(\mu(\tilde{\theta},\theta'),\sigma(\tilde{\theta},\theta'),\theta')\sigma_{\tilde{\theta}}(\tilde{\theta},\theta')] d\tilde{\theta} d\theta' \le 0.$$

Since the above holds for an arbitrary choice of pair $(\hat{\theta},\theta)$ such that $\theta < \hat{\theta}$, (15) should be satisfied. ∎

In the mechanism design literature, a necessary condition for the $\theta$-implementability condition is often alternatively presented using the well-known Spence-Mirrlees single crossing (or sorting) and monotonicity conditions. In the following proposition, we give a connection between the necessary condition and our condition (15).

**Proposition A.1.** *Let a menu of* ($\mu$, $\sigma$)*-implementable linear contracts be* $\{S(Y_1, \theta); \theta \in \Theta\}$, *each of which is given in the form of Equation* (3) with $\{(\mu(\theta), \sigma(\theta)); \theta \in \Theta\}$. (*Note S does not have to be* $S^*$.)

*i. If the menu satisfies condition (iii) of Problem 2, then for almost all* $\theta$,

$$\dot{\beta}(\theta)\left[f_{\mu}(\mu(\theta,\theta),\sigma(\theta,\theta))\mu_{\theta}(\theta,\theta) + (f_{\sigma}(\mu(\theta,\theta),\sigma(\theta,\theta)) - 2r\ \beta\sigma(\theta,\theta))\sigma_{\theta}(\theta,\theta))\right] \ge 0, \quad \text{(A.19)}$$

*where* $\mu_{\theta}(\theta, \theta)$ *and* $\sigma_{\theta}(\theta, \theta)$ *are partial derivatives with respect to their second arguments.*

*ii. Suppose that given $S(Y_1, \hat{\theta})$, the Hessian for agent $\theta$'s problem is strictly negative, that is, $|\Phi''| < 0$. Then (15) implies (A19).*

**Proof.** In this proof, we use the following notational convention. Consider a function $y(x(a,b),\ a,b)$. We use a subscript to note the partial differentiation with respect to the argument of the subscripted variable, for example, $y_a$ is the partial differentiation with respect to the second argument of function $y$. We also use $\frac{\partial}{\partial a} y(x(a,b),a,b)$ to mean $y_x x_a + y_a$.
*Part i.* Let $\Upsilon(\mu,\sigma,\hat{\theta},\theta) := Q(\hat{\theta}) + \Psi(\mu,\sigma,\hat{\theta},\theta)$. Recall

$$\Psi(\mu,\sigma,\hat{\theta},\theta) = \hat{c} - c(\mu_t,\sigma_t,\theta) - \beta(\hat{\theta})\left(\hat{f} - f(\mu_t,\sigma_t)\right) + \frac{r}{2}\beta^2(\hat{\theta})(\hat{\sigma}_t^2 - \sigma_t^2).$$

We know that given $S(Y_1,\hat{\theta})$, the agent $\theta$ chooses $(\mu,\sigma)$ to maximize $\Upsilon(\mu,\sigma,\hat{\theta},\theta)$. Let the solution be $(\mu(\hat{\theta},\theta),\sigma(\hat{\theta},\theta))$. Then the FOCs with respect to $(\mu,\sigma)$ are

$$\Upsilon_\mu(\mu(\hat{\theta},\theta),\sigma(\hat{\theta},\theta),\hat{\theta},\theta) = -c_\mu + \beta(\hat{\theta})f_\mu = 0$$
$$\Upsilon_\sigma(\mu(\hat{\theta},\theta),\sigma(\hat{\theta},\theta),\hat{\theta},\theta) = -c_\sigma + \beta(\hat{\theta})f_\sigma - r\beta^2(\hat{\theta})\sigma(\hat{\theta},\theta) = 0.$$

Note that in fact $\Upsilon_\mu = \Psi_\mu$ and $\Upsilon_\sigma = \Psi_\sigma$. By the $\theta$-incentive compatibility

$$\frac{\partial}{\partial\hat{\theta}}\Upsilon(\mu(\hat{\theta},\theta),\sigma(\hat{\theta},\theta),\hat{\theta},\theta) = \dot{Q}(\hat{\theta}) + \frac{\partial}{\partial\hat{\theta}}\Psi(\mu(\hat{\theta},\theta),\sigma(\hat{\theta},\theta),\hat{\theta},\theta)|_{\hat{\theta}=\theta} = 0.$$

The SOC with respect to $\hat{\theta}$ is

$$\frac{d^2Q}{d\hat{\theta}^2} + \frac{\partial}{\partial\hat{\theta}}\{\Psi_\mu\mu_{\hat{\theta}} + \Psi_\sigma\sigma_{\hat{\theta}} + \Psi_{\hat{\theta}}\} \le 0.$$

Since $\Psi_\mu = \Psi_\sigma = 0$ for all $(\hat{\theta},\theta)$, $\frac{\partial}{\partial\theta}\Psi_\mu = \frac{\partial}{\partial\theta}\Psi_\sigma = 0$ and the SOC reduces to

$$\frac{d^2Q}{d\hat{\theta}^2} + \frac{\partial}{\partial\hat{\theta}}\Psi_{\hat{\theta}} = \frac{d^2Q}{d\hat{\theta}^2} + \Psi_{\mu\hat{\theta}}\mu_{\hat{\theta}} + \Psi_{\sigma\hat{\theta}}\sigma_{\hat{\theta}} + \Psi_{\hat{\theta}\hat{\theta}} \le 0.$$

Since $\frac{\partial}{\partial\hat{\theta}}\Upsilon(\mu(\hat{\theta},\theta),\sigma(\hat{\theta},\theta),\hat{\theta},\theta) = 0$ for all $\theta$, differentiation with respect to $(\hat{\theta},\theta)$ at $\hat{\theta}=\theta$ yields

$$\frac{d}{d\theta}\left(\frac{\partial}{\partial\hat{\theta}}\Upsilon|_{\hat{\theta}=\theta}\right) = \frac{d^2Q}{d\theta^2}Q(\theta) + \frac{\partial}{\partial\hat{\theta}}\left(\Psi_\mu\mu_{\hat{\theta}} + \Psi_\sigma\sigma_{\hat{\theta}} + \Psi_{\hat{\theta}}\right) + \frac{\partial}{\partial\theta}\left(\Psi_\mu\mu_{\hat{\theta}} + \Psi_\sigma\sigma_{\hat{\theta}} + \Psi_{\hat{\theta}}\right) = 0.$$

Therefore, using the SOC and the fact that $\Psi_\mu = \Psi_\sigma = 0$ for all $(\hat{\theta},\theta)$, the above reduces to $\frac{\partial}{\partial\theta}\Psi_{\hat{\theta}} \ge 0$. That is,

$$\frac{\partial}{\partial\theta}\Psi_{\hat{\theta}} = \Psi_{\mu\hat{\theta}}\mu_\theta + \Psi_{\sigma\hat{\theta}}\sigma_\theta + \Psi_{\theta\hat{\theta}} \ge 0.$$

Since $\Psi_{\theta\hat{\theta}} = 0$ and since $\Psi_{\mu\hat{\theta}} = \dot{\beta}f_\mu$ and $\Psi_{\sigma\hat{\theta}} = \dot{\beta}f_\sigma - 2r\beta\dot{\beta}\sigma(\theta,\theta)$, we obtain the statement of the proposition.
*Part ii.* Note that using the FOCs with respect to $(\mu,\sigma)$, one can show that

$$\mu_{\hat{\theta}}(\hat{\theta},\theta) = \frac{\dot{\beta}(\hat{\theta})}{|\Phi''|}\left[-f_\mu\Phi_{\sigma\sigma} + \Phi_{\mu\sigma}(f_\sigma - 2r\beta\sigma(\hat{\theta},\theta))\right]$$
$$\sigma_{\hat{\theta}}(\hat{\theta},\theta) = \frac{\dot{\beta}(\hat{\theta})}{|\Phi''|}\left[\Phi_{\mu\mu}(2r\beta\sigma(\hat{\theta},\theta) - f_\sigma) + f_\mu\Phi_{\mu\sigma}\right]$$

If the above two are substituted into (15), with arguments suppressed, we have

$$\dot{\beta}(\hat{\theta})\left[(c_{\mu\theta}f_\mu\Phi_{\sigma\sigma} - c_{\sigma\theta}f_\mu\Phi_{\mu\sigma}) + (f_\sigma - 2r\beta(\hat{\theta})\sigma)(c_{\sigma\theta}\Phi_{\mu\mu} - c_{\mu\theta}\Phi_{\mu\sigma})\right] \ge 0.$$

On the other hand, the FOCs with respect to ($\mu$, $\sigma$) imply that

$$\mu_\theta = \frac{1}{|\Phi''|}(c_{\mu\theta}\Phi_{\sigma\sigma} - \Phi_{\mu\sigma}c_{\sigma\theta})$$

$$\sigma_\theta = \frac{1}{|\Phi''|}(c_{\sigma\theta}\Phi_{\mu\mu} - c_{\mu\theta}\Phi_{\mu\sigma})$$

Therefore, rearranging the LHS of inequality (A20) using the above two equations and letting $(\hat{\theta}, \theta) = (\theta, \theta)$, we have (A19). ■

The condition in (A19), consistent with Fudenberg and Tirole (1991, Theorem 7.1), is a necessary condition for the $\theta$-implementability.[31] Typically in the literature, signs of $\Upsilon_{\mu\hat{\theta}}$ and $\Upsilon_{\sigma\hat{\theta}}$ are assumed to be either positive or negative. [Note $\Upsilon_{\mu\hat{\theta}} = \Psi_{\mu\hat{\theta}} = \Psi_{\mu\hat{\theta}} = \dot{\beta}f_\mu(\mu, \sigma)$ and $\Upsilon_{\sigma\hat{\theta}} = \Psi_{\sigma\hat{\theta}} = \Psi_{\sigma\hat{\theta}} = \dot{\beta}(\theta)(f_\sigma(\mu, \sigma) - 2r\beta\sigma(\theta, \theta))$] These sign conditions are referred to as the Spence-Mirrlees single crossing conditions. Under the single crossing conditions, the signs of $\mu_\theta(\theta, \theta)$ and $\sigma_\theta(\theta, \theta)$ are examined as a necessary condition for the $\theta$-implementability.

When $c_\sigma = 0$, both the ($\mu$, $\sigma$)- and $\theta$-implementability conditions are simplified as follows.

**Proposition A.2.** *Suppose that $c_\sigma = 0$, that the menu of contracts is given by Equation* (3) *with sensitivities of $\beta^*(\theta) = c_\mu(\mu^*(\theta),\theta)/f_\mu(\mu^*(\theta),\sigma^*(\theta)) > 0$, for $\theta \in \Theta$, and that $(\mu^*(\theta), \sigma^*(\theta))$ maximizes the principal's Lagrangian* (9).

*i. The ($\mu$, $\sigma$)-implementability is satisfied if for all* $(\mu, \sigma, \theta) \in U \times \Sigma \times \Theta$

$$(\beta^*(\theta)f_{\mu\mu}(\mu, \sigma) - c_{\mu\mu}(\mu, \theta))(f_{\sigma\sigma}(\mu, \sigma) - r\beta^*(\theta)) - \beta^*(\theta)f^2_{\mu\sigma}(\mu, \sigma) > 0.$$

*ii. Suppose that the ($\mu$, $\sigma$)-implementability condition in part i. holds. If for all* $(\mu^*(\theta), \sigma^*(\theta)), \theta \in \Theta$,

$$f_\mu(\mu^*, \sigma^*)(f_{\sigma\sigma}(\mu^*, \sigma^*) - r\beta^*(\theta)) + f_{\mu\sigma}(\mu^*, \sigma^*)r\sigma^*)\beta^*(\theta) < 0,$$

*and if contracts in the menu are nonincreasing in their sensitivities, that is,* $\dot{\beta}^*(\theta) \leq 0$*, then the $\theta$-implementability is satisfied.*

**Remark.** Proposition A.2 implies that if $f_{\mu\sigma}$, $f_{\sigma\sigma} \leq 0$ and if sensitivities in the menu are nonincreasing in $\theta$, then the $\theta$-implementability is achieved.

**Proof.** By Lemma A.2, given a contract with a sensitivity of $\beta^*(\hat{\theta})$, agent $\theta \in \Theta$ solves the following problem.

$$\max_{(\mu', \sigma') \in U \times \Sigma} \Phi(\mu', \sigma'; \mu^*(\hat{\theta}), \sigma^*(\hat{\theta}), \hat{\theta}; \theta) = \beta^*(\hat{\theta})f(\mu', \sigma') - \frac{r}{2}(\beta^*(\hat{\theta}))^2\sigma'^2 - c(\mu', \theta).$$

The first-order conditions are

$$0 = \beta^*(\hat{\theta})f_\mu - c_\mu, \; 0 = \beta^*(\hat{\theta})f_\sigma - r(\beta^*(\hat{\theta}))^2\sigma^*.$$

Note that $\Phi_{\mu\mu} = \beta^*(\hat{\theta})f_{\mu\mu} - c_{\mu\mu} < 0$, $\Phi_{\mu\sigma} = \beta^*(\hat{\theta})f_{\mu\sigma}$, and $\Phi_{\sigma\sigma} = \beta^*(\hat{\theta})f_{\sigma\sigma} - r(\beta^*(\hat{\theta}))^2 < 0$.
*Part i.* When $\beta^*(\hat{\theta}) = \beta^*(\theta)$, $(\mu^*(\theta), \sigma^*(\theta))$ trivially satisfies the first FOC. It also satisfies the second FOC since $(\mu^*(\theta), \sigma^*(\theta))$ solves the principal's Lagrangian maximization. Moreover,

[31] To derive a sufficient condition, one may use the line of reasoning found in the proof of Fudenberg and Tirole (1991, Theorem 7.3). However, their proof is essentially the same as that of our inequality (15). Thus, of course, the reasoning of Fudenberg and Tirole (1991, Theorem 7.3) should also lead to the sufficient condition identical to (15).

by the stated hypothesis, the agent's SOC is satisfied as well. Therefore, given $\beta^*(\hat{\theta}) = \beta^*(\theta)$, agent $\theta$ chooses $(\mu^*(\theta), \sigma^*(\theta))$, satisfying the $(\mu, \sigma)$-implementability.
*Part ii*. Let us differentiate the agent's FOCs with respect to $\hat{\theta}$,

$$\Phi_{\mu\mu}\mu^*_{\hat{\theta}} + \Phi_{\mu\sigma}\sigma^*_{\hat{\theta}} = -\dot{\beta}^* f_\mu,\ \ \Phi_{\mu\sigma}\mu^*_{\hat{\theta}} + \Phi_{\sigma\sigma}\sigma^*_{\hat{\theta}} = -\dot{\beta}^* f_\sigma + 2r\beta^*\dot{\beta}^*\sigma^*.$$

By Theorem A.3, the $\theta$-implementability is also satisfied if $\mu^*_{\hat{\theta}}(\hat{\theta}, \theta) \leq 0$. Since, by part i., the determinant of the Hessian is strictly positive, we have

$$\begin{aligned} sign(\mu^*_{\hat{\theta}}(\hat{\theta}, \theta)) &= sign\{\dot{\beta}^*\left[-f_\mu(f_{\sigma\sigma} - r\beta^*) - f_{\mu\sigma}(2r\sigma^*\beta^* - f_\sigma)\right]\} \\ &= sign\{\dot{\beta}^*\left[-f_\mu(f_{\sigma\sigma} - r\beta^*) - f_{\mu\sigma}r\sigma^*\beta^*\right]\} \leq 0. \end{aligned}$$

For the last equality, we have utilized the agent's FOC with respect to $\sigma$. The last inequality is implied by the stated hypothesis. Thus, by Theorem A.3, the $\theta$-implementability is satisfied.

## B.3 Proof of Proposition 2

*Part i*. Suppose agent $\theta$ chooses contract $S^*(Y_1, \hat{\theta})$. Then the agent's FOCs are

$$\begin{aligned} \Phi_\mu &= \beta^*(\hat{\theta})f_\mu - c_\mu = 0 \\ \Phi_\sigma &= \beta^*(\hat{\theta})f_\sigma - r(\beta^*(\hat{\theta}))^2\sigma^*(\hat{\theta}, \theta) - c_\sigma = 0. \end{aligned}$$

Note that the above conditions must hold for all $(\hat{\theta}, \theta)$. Thus, by differentiating the FOCs with respect to $\hat{\theta}$, we have

$$\begin{bmatrix} \Phi_{\mu\mu} & \Phi_{\mu\sigma} \\ \Phi_{\mu\sigma} & \Phi_{\sigma\sigma} \end{bmatrix} \begin{bmatrix} \mu^*_{\hat{\theta}}(\hat{\theta}, \theta) \\ \sigma^*_{\hat{\theta}}(\hat{\theta}, \theta) \end{bmatrix} = \begin{bmatrix} -\dot{\beta}^* f_\mu \\ -\dot{\beta}^* f_\sigma + 2r\beta^*\dot{\beta}^*\sigma^*(\hat{\theta}, \theta) \end{bmatrix},$$

where $\Phi_{\mu\mu} = \beta^* f_{\mu\mu} - c_{\mu\mu}$, $\Phi_{\mu\sigma} = \beta^* f_{\mu\sigma} - c_{\mu\sigma}$, and $\Phi_{\sigma\sigma} = \beta^* f_{\sigma\sigma} - r(\beta^*)^2 - c_{\sigma\sigma}$. Let the Hessian be denoted by $|\Phi''|$. By assumptions, $|\Phi''| > 0$, $\Phi_{\mu\mu} < 0$, $\Phi_{\mu\sigma} \leq 0$ and $\Phi_{\sigma\sigma} < 0$.

$$\begin{aligned} \mu^*_{\hat{\theta}}(\hat{\theta}, \theta) &= \frac{\dot{\beta}^*}{|\Phi''|}\left[-f_\mu\Phi_{\sigma\sigma} + \Phi_{\mu\sigma}(f_\sigma - 2r\beta^*\sigma^*(\hat{\theta}, \theta))\right] \\ \sigma^*_{\hat{\theta}}(\hat{\theta}, \theta) &= \frac{\dot{\beta}^*}{|\Phi''|}\left[\Phi_{\mu\mu}(2r\beta^*\sigma^*(\hat{\theta}, \theta) - f_\sigma) + f_\mu\Phi_{\mu\sigma}\right] \end{aligned}$$

Therefore, $sign(\mu^*_{\hat{\theta}}(\hat{\theta}, \theta)) = sign(\dot{\beta}^*)$ and $sign(\sigma^*_{\hat{\theta}}(\hat{\theta}, \theta)) = -sign(\dot{\beta}^*)$.

Now, suppose $\dot{\beta}^* > 0$. Then we have $\mu^*_{\hat{\theta}}(\hat{\theta}, \theta) > 0$ and $\sigma^*_{\hat{\theta}}(\hat{\theta}, \theta) < 0$, which imply $c_{\mu\theta}\mu^*_{\hat{\theta}}(\hat{\theta}, \theta) + c_{\theta\sigma}\sigma^*_{\hat{\theta}}(\hat{\theta}, \theta) > 0$, contradicting the $\theta$-implementability condition (15). Therefore, we must have $\dot{\beta}^* \leq 0$, $\mu^*_{\hat{\theta}}(\hat{\theta}, \theta) \leq 0$ and $\sigma^*_{\hat{\theta}}(\hat{\theta}, \theta) \geq 0$.
*Part ii*. Note that the FOCs must hold for all $(\hat{\theta}, \theta)$. Differentiating the FOCs with respect to θ, we have

$$\begin{bmatrix} \Phi_{\mu\mu} & \Phi_{\mu\sigma} \\ \Phi_{\mu\sigma} & \Phi_{\sigma\sigma} \end{bmatrix} \begin{bmatrix} \mu^*_{\theta}(\hat{\theta}, \theta) \\ \sigma^*_{\theta}(\hat{\theta}, \theta) \end{bmatrix} = \begin{bmatrix} c_{\mu\theta} \\ c_{\sigma\theta} \end{bmatrix}.$$

Thus,

$$\mu^*_{\theta}(\hat{\theta}, \theta) = \frac{1}{|\Phi''|}\{c_{\mu\theta}\Phi_{\sigma\sigma} - c_{\sigma\theta}\Phi_{\mu\sigma}\} < 0.$$

$$\sigma^*_{\theta}(\hat{\theta}, \theta) = \frac{1}{|\Phi''|}\{\Phi_{\mu\mu}c_{\mu\theta} - \Phi_{\mu\sigma}c_{\sigma\theta}\} \geq 0.$$

However, using results from the proof of part i,

$$\dot{\mu}^*(\theta) = \mu^*_{\hat{\theta}} + \mu^*_{\theta} < 0$$
$$\dot{\sigma}^*(\theta) = \sigma^*_{\hat{\theta}} + \sigma^*_{\theta} \geq 0,$$

as claimed in the statement. ■

# C Appendix to Section 4

## C.1 Proof of Proposition 3

From the FOCs given in Equations (16) and (17),

$$H_{\mu\mu} = -a^2 - \frac{ra^4\sigma^{*2}}{(\gamma+\delta\sigma^*)^2},\ H_{\mu\sigma} = \delta - \frac{2ra^3\gamma m^*\sigma^*}{(\gamma+\delta\sigma^*)^3},$$
$$H_{\sigma\sigma} = g_{\sigma\sigma} - \frac{ra^2\gamma^* m^{*2}}{(\gamma+\delta\sigma^*)^3} + \frac{3ra^2\gamma\delta m^{*2}\sigma^*}{(\gamma+\delta\sigma^*)^4} = g_{\sigma\sigma} - \frac{ra^3\gamma m^{*2}}{(\gamma+\delta\sigma^*)^4}(\gamma - 2\delta\sigma^*).$$

By the SOC for a unique solution

$$H_{\mu\mu} < 0,\ H_{\sigma\sigma} < 0,\ H_{\mu\mu}H_{\sigma\sigma} - H^2_{\mu\sigma} > 0.$$

Differentiating with respect to $\theta$

$$H_{\mu\mu}\frac{\dot{m}^*}{a^2} + H_{\mu\sigma}\dot{\sigma}^* = \dot{\nu}ab,\ H_{\mu\sigma}\frac{\dot{m}^*}{a^2} + H_{\sigma\sigma}\dot{\sigma}^* = \frac{b\delta}{a}.$$

Therefore,

$$sign(\dot{c}_\mu) = sign(\dot{\nu}abH_{\sigma\sigma} - H_{\mu\sigma}\frac{b\delta}{a}),\ sign(\dot{\sigma}^*) = sign(H_{\mu\mu}\frac{b\delta}{a} - \dot{\nu}abH_{\mu\sigma}).$$

Part I with $\gamma > 0$, $\delta \leq 0$.

(i) The third-best solutions: Note that $H_{\mu\sigma} \leq 0$. Thus,

$$\dot{c}_\mu < 0 \quad \text{and} \quad \dot{\sigma}^* > 0.$$

The following computation indicates that the sensitivity is decreasing in $\theta$. (Recall that $f_\mu > 0$ is always assumed and that the sensitivity in Equation (18) is defined for $\gamma + \delta\sigma^* - \nu ab > 0$.)

$$\frac{d}{d\theta}\left(\frac{am^*}{\gamma+\delta\sigma^*}\right)\left[(\gamma+\delta\sigma^*)^2 + ra^2\sigma^{*2}\right]^2$$
$$= (\gamma+\delta\sigma^*)\Big[2ra^2\delta\sigma^{*2}\dot{\sigma}^* - \dot{\nu}ab\Big((\gamma+\delta\sigma^*)^2 + ra^2\sigma^{*2}\Big) \tag{A21}$$
$$+2\nu ab(\gamma+\delta\sigma^*)\delta\dot{\sigma}^* - 2(\gamma+\delta\sigma^* - \nu ab)ra^2\sigma^*\dot{\sigma}^*\Big] < 0.$$

Since the sensitivity is strictly decreasing in $\theta$, by Equation (6) and an argument similar to Proposition A.2-ii, one can show that the $\theta$-implementability holds. Moreover, $\mu$-implementability is also ensured because by construction $f_\mu > 0$, $c_\mu$, $c_{\mu\mu} \geq 0$, and $f_{\mu\mu} \leq 0$ are satisfied.

(ii) The second-best solutions: By setting $\dot{\nu} = \nu = 0$ in the above computations, the assertion follows.

(iii) Obvious from parts I-i and I-ii-(2).

Part II with $\gamma = 0$ and $\delta > 0$.

(i) The third-best solutions: note that

$$\dot{\nu}abH_{\sigma\sigma} - H_{\mu\sigma}\frac{b\delta}{a} = \dot{\nu}abg_{\sigma\sigma} - \frac{b\delta^2}{a} < 0,\ H_{\mu\mu}\frac{b\delta}{a} - \dot{\nu}abH_{\mu\sigma} = H_{\mu\mu}\frac{b\delta}{a} - \dot{\nu}ab\delta < 0.$$

Hence,

$$\dot{c}_\mu < 0 \quad \text{and} \quad \dot{\sigma} < 0.$$

Moreover, using the same computation as in (A21), one can easily check that the sensitivity is strictly decreasing. Thus, by the same reasonings as in part I, both $\theta$- and $\mu$-implementabilities are satisfied.

Since $0 = \delta\mu + g_\sigma$ and $g$ is quadratic, the closed-form solution for $\sigma$ is

$$B(\sigma - C) = -\frac{b\delta\theta}{a} + \frac{\delta^3(\delta\sigma - \nu(\theta)ab)}{a^2\delta^2 + ra^4}.$$

Thus,

$$\sigma = \frac{(a^2BC - ab\delta\theta)(\delta^2 + ra^2) - \delta^3\nu(\theta)ab)}{a^2B(\delta^2 + ra^2) - \delta^4}.$$

The denominator is positive by the SOC.

(ii) The second-best solutions: Again, by setting $\dot{\nu} = \nu = 0$ in the above computations, the assertion follows.

(iii) Obvious from parts II-i and II-ii. ■

## C.2 Proof of Proposition 4

Using the FOCs,

$$\mathcal{L}_{\mu\mu} = -a^2 - \frac{ra^4\sigma^{*2}}{(\gamma + \delta\sigma^*)^2},\ \mathcal{L}_{\mu\sigma} = \delta - \frac{2ra^3\gamma m^*\sigma^*}{(\gamma + \delta\sigma^*)^3} - q\frac{ra^2\gamma}{(\gamma + \delta\sigma^*)^2},$$

$$\mathcal{L}_{\sigma\sigma} = g_{\sigma\sigma} - \frac{ra^2\gamma m^{*2}}{(\gamma + \delta\sigma^*)^4}(\gamma - 2\delta\sigma^*) + q\left(g_{\sigma\sigma\sigma} + \frac{2ra\gamma\delta m^*}{(\gamma + \delta\sigma^*)^3}\right),$$

$$\mathcal{L}_{\mu g} = \delta - \frac{ra^2\sigma^*}{\gamma + \delta\sigma^*},\ \mathcal{L}_{\sigma q} = g_{\sigma\sigma} - \frac{ra\gamma m^*}{(\gamma + \delta\sigma^*)^2}.$$

By the assumption of unique solutions, $\mathcal{L}_{\mu\mu} < 0$, $\mathcal{L}_{\sigma\sigma} < 0$, and the determinant of the bordered Hessian is strictly positive.

Differentiating the FOCs with respect to $\theta$, we have

$$\mathcal{L}_{\mu m}\dot{m}^* + \mathcal{L}_{\mu\sigma}\dot{\sigma}^* + \mathcal{L}_{\mu q}\dot{q} = \nu ab,\ \mathcal{L}_{\sigma m}\dot{m}^* + \mathcal{L}_{\sigma\sigma}\dot{\sigma}^* + \mathcal{L}_{\sigma q}\dot{q} = \frac{b\delta}{a},\ \mathcal{L}_{qm}\dot{m}^* + \mathcal{L}_{q\sigma}\dot{\sigma}^* = \frac{b\delta}{a}.$$

Since $\mathcal{L}_{\mu\mu} = \mathcal{L}_{\mu m}a,\ \mathcal{L}_{\sigma\mu} = \mathcal{L}_{\sigma m}a$, and $\mathcal{L}_{q\mu} = \mathcal{L}_{qm}a$, we have

$$\begin{bmatrix} \mathcal{L}_{\mu\mu} & \mathcal{L}_{\mu\sigma} & \mathcal{L}_{\mu q} \\ \mathcal{L}_{\sigma\mu} & \mathcal{L}_{\sigma\sigma} & \mathcal{L}_{\sigma q} \\ \mathcal{L}_{q\mu} & \mathcal{L}_{q\sigma} & 0 \end{bmatrix} \begin{bmatrix} \dot{m}^*/a \\ \dot{\sigma}^* \\ \dot{q} \end{bmatrix} = \begin{bmatrix} \dot{\nu}ab \\ b\delta/a \\ b\delta/a \end{bmatrix}.$$

Therefore,

$$sign(\dot{m}^*) = sign\left(\frac{b\delta}{a}\mathcal{L}_{q\sigma}\mathcal{L}_{\mu q} + \mathcal{L}_{\mu\sigma}\mathcal{L}_{\sigma q}\frac{b\delta}{a} - \mathcal{L}_{\mu q}\mathcal{L}_{\sigma\sigma}\frac{b\delta}{a} - \dot{\nu}ab\mathcal{L}_{\sigma q}^2\right),$$

$$sign(\dot{\sigma}^*) = sign\left(\mathcal{L}_{\mu\sigma}\frac{b\delta}{a}\mathcal{L}_{\mu q} + \dot{\nu}ab\mathcal{L}_{\sigma q}\mathcal{L}_{q\mu} - \mathcal{L}_{\mu q}^2\frac{b\delta}{a} - \mathcal{L}_{\mu\mu}\mathcal{L}_{\sigma q}\frac{b\delta}{a}\right).$$

Part I with $\gamma > 0$ and $\delta = 0$.

(i) The third-best solutions:

$$\mathcal{L}_{\mu\mu} < 0,\ \mathcal{L}_{\mu\sigma} = -\frac{2ra^3m^*\sigma^*}{\gamma^2} - q\frac{ra^2}{\gamma},\ \mathcal{L}_{\sigma\sigma} = g_{\sigma\sigma} - \frac{ra^2m^{*2}}{\gamma^2} + qg_{\sigma\sigma\sigma},$$

$$\mathcal{L}_{\mu q} = -\frac{ra^2\sigma^*}{\gamma} < 0,\ \mathcal{L}_{\sigma q} = g_{\sigma\sigma} - \frac{ram^*}{\gamma} < 0.$$

Thus,

$$sign(\dot{m}^*) = sign(-\dot{\nu}ab\mathcal{L}_{\sigma q}^2) < 0, sign(\dot{\sigma}^*) = sign(\dot{\nu}ab\mathcal{L}_{\sigma q}\mathcal{L}_{q\mu}) > 0.$$

Since $c_\mu/f_\mu = am^*/\gamma$, the sensitivity behaves like $m^*$, that is, $(d/d\theta)(c_\mu/f_u) = (a/\gamma)\dot{m}^* < 0$. Since the sensitivity is (strictly) decreasing, $\theta$-implementability is satisfied by Proposition A.2.

The ($\mu$, $\sigma$)-implementability is ensured because the agent's SOC in Proposition A.2-i is satisfied. That is, since $\delta = 0$, we have for all ($m^*$, $\sigma^*$)

$$-a^2\left(g_{\sigma\sigma} - r\frac{am^*}{\gamma + \delta\sigma^*}\right) - \frac{am^*}{\gamma + \delta\sigma^*}\delta^2 \geq 0.$$

(ii) The second-best solutions: By setting $\dot{\nu} = 0$, we have

$$\dot{c}_\mu(\mu_{2\text{nd}}, \theta) = \dot{\sigma}_{2\text{nd}}(\theta) = (d/d\theta)(c_\mu(\mu_{2\text{nd}}, \theta)/f_\mu(\mu_{2\text{nd}}, \sigma_{2\text{nd}}) = 0.$$

Thus, $c_\mu(\mu_{2\text{nd}}, \theta)$, $\sigma_{2\text{nd}}(\theta)$, and the second-best sensitivities are the same across all $\theta \in \Theta$.

(iii) Comparison: Since $\mu^*(\theta_L) = \mu_{2\text{nd}}(\theta)$, $\sigma^*(\theta_L) = \sigma_{2\text{nd}}(\theta)$, and $\dot{\sigma}^* > 0$, we have $\sigma^*(\theta) > \sigma_{2\text{nd}}(\theta)$ for all $\theta \in (\theta_L, \theta_H]$.

Part II with $\gamma = 0$ and $\delta > 0$.

(i) The third-best solutions:

$$\mathcal{L}_{\mu\mu} = -a^2 - \frac{ra^4}{\delta^2} < 0,\ \mathcal{L}_{\mu\sigma} = \delta > 0,\ \mathcal{L}_{\sigma\sigma} = g_{\sigma\sigma} + qg_{\sigma\sigma\sigma},\ \mathcal{L}_{\mu q} = \delta - \frac{ra^2}{\delta},\ \mathcal{L}_{\sigma q} = g_{\sigma\sigma} < 0.$$

$$sign(\dot{m}^*) = sign\left(\frac{b\delta}{a}\mathcal{L}_{q\sigma}\mathcal{L}_{\mu q} + \mathcal{L}_{\mu\sigma}\mathcal{L}_{\sigma q}\frac{b\delta}{a} - \mathcal{L}_{\mu q}\mathcal{L}_{\sigma\sigma}\frac{b\delta}{a} - \dot{\nu}ab\mathcal{L}_{\sigma q}^2\right),$$

$$sign(\dot{\sigma}^*) = sign\bigg(\mathcal{L}_{\mu\sigma}\frac{b\delta}{a}\mathcal{L}_{\mu q} + \dot{\nu}ab\mathcal{L}_{\sigma q}\mathcal{L}_{q\mu} - \mathcal{L}_{\mu q}^2\frac{b\delta}{a} - \mathcal{L}_{\mu\mu}\mathcal{L}_{\sigma q}\frac{b\delta}{a}\bigg).$$

Note that the FOCs (26) and (27) imply that $q = -ram/\delta g_{\sigma\sigma} > 0$. Using the assumption $(\delta^2 - ra^2)g_{\sigma\sigma\sigma} \geq 0$, we have $\dot{m}^*$<0, because

$$\mathcal{L}_{q\sigma}\mathcal{L}_{\mu q} + \mathcal{L}_{\mu\sigma}\mathcal{L}_{\sigma q} - \mathcal{L}_{\mu q}\mathcal{L}_{\sigma\sigma} - \frac{\dot{\nu}a^2}{\delta}\mathcal{L}_{\sigma q}^2 = \delta g_{\sigma\sigma} - \left(\delta - \frac{ra^2}{\delta}\right)qg_{\sigma\sigma\sigma} - \frac{\dot{\nu}a^2}{\delta}\mathcal{L}_{\sigma q}^2 < 0.$$

We also have $\dot{\sigma} < 0$, because

$$\begin{aligned}\mathcal{L}_{\mu\sigma}\mathcal{L}_{\mu q} + \frac{\dot{\nu}a^2}{\delta}\mathcal{L}_{\sigma q}\mathcal{L}_{q\mu} - \mathcal{L}_{\mu q}^2 - \mathcal{L}_{\mu\mu}\mathcal{L}_{\sigma q} &= \delta^2 - ra^2 + \frac{\dot{\nu}a^2}{\delta^2}g_{\sigma\sigma}(\delta^2 - ra^2) \\ &\quad - \frac{1}{\delta^2}(\delta^2 - ra^2)^2 + \frac{a^2}{\delta^2}(\delta^2 + ra^2)g_{\sigma\sigma} \\ = \delta^2 - ra^2 + a^2 g_{\sigma\sigma} - \frac{1}{\delta^2}(\delta^2 - ra^2)^2 + \frac{a^2}{\delta^2}g_{\sigma\sigma}\{\dot{\nu}(\delta^2 - ra^2) + ra^2\} < 0.\end{aligned}$$

For the above inequality, we have utilized our assumption that $\dot{\nu}(\delta^2 - ra^2) + ra^2 \geq 0$, and the SOC, $\delta^2 - ra^2 + a^2 g_{\sigma\sigma} < 0$.
For the SOC, note that by the assumption of unique solutions, the determinant of the bordered Hessian is strictly positive, that is,

$$0 < 2\mathcal{L}_{\mu\sigma}\mathcal{L}_{q\sigma}\mathcal{L}_{q\mu} - \mathcal{L}_{q\mu}^2\mathcal{L}_{\sigma\sigma} - \mathcal{L}_{q\sigma}^2\mathcal{L}_{\mu\mu} = \frac{g_{\sigma\sigma}}{\delta^2}(\delta^2 + ra^2)(\delta^2 - ra^2 + a^2 g_{\sigma\sigma}).$$

Since $g_{\sigma\sigma} < 0$, we have

$$\delta^2 - ra^2 + a^2 g_{\sigma\sigma} < 0.$$

On the other hand, in order to explicitly solve the third-best volatility, let us substitute $q = -ram/\delta g_{\sigma\sigma}$ into the FOC with respect to $\mu$. Then we have

$$am^* = \frac{\delta\sigma^* - \nu(\theta)ab}{\left(1 + r\frac{a^2}{\delta^2}\right) + \frac{r}{g_{\sigma\sigma}}\left(1 - \frac{ra^2}{\delta^2}\right)}.$$

However, by the constraint

$$\delta\mu^* + g_\sigma = \frac{r}{\sigma}am^*,$$

which implies

$$g_\sigma - \frac{\delta b}{a}\theta = \frac{ra^2 - \delta^2}{a^2\delta}am^*.$$

When $g$ is quadratic, and Equations (A22) and (A23), we have Equation (31).
In order to look at changes in the sensitivity over $\theta$, let us use Equation (A.22).

$$\frac{am^*}{\delta\sigma^*} = \frac{B\delta\left(\delta - \frac{\nu(\theta)}{\sigma^*}ab\right)}{B(\delta^2 + ra^2) - r(\delta^2 - ra^2)} < 1.$$

The SOC is used for the above inequality. Thus,

$$\frac{d}{d\theta}\left(\frac{am^*}{\delta\sigma^*}\right) = -\frac{Bab\delta}{B(\delta^2 + ra^2) - r(\delta^2 - ra^2)}\frac{d}{d\theta}\left(\frac{\nu(\theta)}{\sigma^*}\right).$$

Since $B(\delta^2 + ra^2) - r(\delta^2 - ra^2) > 0$ by the SOC, $\dot{\sigma}^* < 0$ and

$$\frac{d}{d\theta}\left(\frac{\nu(\theta)}{\sigma^*}\right) = \frac{1}{\sigma^{*2}}(\dot{\nu}\sigma^* - \nu\dot{\sigma}^*) > 0,$$

we have $am/\delta\sigma < 0$, satisfying the $\theta$-implementability condition.

To check the $(\mu, \sigma)$-implementability, note that the agent's SOC is

$$a^2\left(B + r\frac{am^*}{\delta\sigma^*}\right) - \frac{am^*}{\delta\sigma^*}\delta^2 = a^2B + \frac{am^*}{\delta\sigma^*}(ra^2 - \delta^2) \geq 0.$$

Since we know that $0 \leq am^*/\delta\sigma^* < 1$ from the previous computation, the assumption of either $a^2B - \delta^2 \geq 0$ or $ra^2 - \delta^2 \geq 0$ implies that the agent's SOC is always satisfied. Thus, the $(\mu, \sigma)$-implementability is satisfied.

ii. The second-best solutions: By setting $\dot{\nu} = \nu = 0$, we have

$$\dot{c}_\mu(\mu_{\text{2nd}}, \theta) < 0, \ \ \dot{\sigma}_{\text{2nd}}(\theta) < 0, \ \text{and} \ \ \frac{d}{d\theta}\frac{c_\mu(\mu_{\text{2nd}}, \theta)}{f_\mu(\mu_{\text{2nd}}, \sigma_{\text{2nd}})} = 0.$$

When $g$ is quadratic, Equation (31) implies that the second-best volatility is explicitly given by Equation (32).

iii. Comparison: By comparing Equations (31) and (32), we have the following relationship. If $\delta^2 - ra^2 > (=,<)0$, then we have $\sigma^*(\theta) < (=,>)\sigma_{\text{2nd}}(\theta)$ for all $\theta \in (\theta_L, \theta_H]$, and $\sigma^*(\theta_L) = \sigma_{\text{2nd}}(\theta_L)$. ■

## D Appendix to Section 5

### D.1 Proof of Proposition 5

It suffices to show that under the stated assumptions, the constraint of the principal's Lagrangian is not binding. Fix $\theta \in \Theta$. Suppose $(\mu^*, \sigma^*)$ maximizes the principal's unconstrained Hamiltonian. Since $c_{\mu\sigma} = f_\sigma = 0$, the FOC with respect to $\sigma$ is

$$H_\sigma = -c_\sigma - r\sigma^*\frac{c_\mu^2}{f_\mu^2} - \nu(\theta)c_{\theta\sigma} = 0.$$

On the other hand, since $f_\sigma = 0$, the constraint from the principal's Lagrangian (9), given a contract with a sensitivity of $c_\mu(\mu^*, \sigma^*)/f_\mu(\mu^*, \sigma^*)$, is

$$\mathcal{L}_q = \Phi_\sigma = -r\frac{c_\mu^2}{f_\mu^2}\sigma - c_\sigma = 0.$$

Thus, when $c_{\theta\sigma}=0$, both $H_\sigma$ and $\Phi_\sigma$ are identical. Moreover, since $c_{\mu\sigma}=f_\sigma=0$, $\Phi$ is strictly concave in $(\mu, \sigma)$, and the agent's FOC with respect to $\mu$, $\Phi_\mu=0$ is satisfied with $\mu^*$. Hence, the agent optimally chooses $(\mu^*, \sigma^*)$. That is, the constraint of the Lagrangian is not binding.

If $c_{\theta\sigma}\neq 0$, $\sigma^*(\theta)$ satisfies both $H_\sigma$ and $\Phi_\sigma$ that are identical only if $\theta=\theta_L$, because $\nu(\theta_L)=0$. ■

## D.2 Proof of Proposition 6

We prove the second claim first. Let

$$X = a^2B(\delta^2+ra^2)-\delta^4,$$

$$Y = (a^2BC-ab\delta\theta)(\delta^2+ra^2)-\delta^3\nu(\theta)ab$$

$$= (a^2BC-ab\delta\theta)\frac{1}{a^2B}(X+\delta^4)-\delta^3\nu(\theta)ab.$$

Some algebraic manipulations using Equations (31) and (22) yield

$$\sigma^*_{\text{obs}}(\theta)-\sigma^*_{\text{unobs}}(\theta)$$

$$=\frac{Y}{X}-\frac{BY-(a^2BC-ab\delta\theta)r(\delta^2-ra^2)+Br\nu a^3b\delta}{B(X+r^2a^4)}$$

$$=(ra^2\delta^2+X)ra\delta\left[(aBC-b\delta\theta)\delta-\nu(\theta)Ba^2b\right].$$

That is,

$$\sigma^*_{\text{obs}}(\theta)>(=,\ <)\sigma^*_{\text{unobs}}(\theta)\ \textit{iff}\ (aBC-b\delta\theta)\delta-\nu(\theta)Ba^2b>(=,\ <)0.$$

Thus, the second part is proven. Note that the second-best volatilities should also satisfy the above inequality with $\nu(\theta)=0$. But by assumption of interior solutions, we always have $(aBC-b\delta\theta)\delta>0$, and thus $\sigma^{\text{2nd}}_{\text{obs}}(\theta)>\sigma^{\text{2nd}}_{\text{unobs}}(\theta)$.

Under the uniform distribution, $\nu(\theta)=\theta-\theta_L$, and

$$(aBC-b\delta\theta)\delta-\nu(\theta)Ba^2b>(=,<)0\ \textit{iff}\ \theta<(=,>)\frac{aBC\delta+Ba^2b\theta_L}{b\delta^2+Ba^2b}.$$

Therefore, we are done. ■

**References**

Baron, D., 1982, "A Model of the Demand for Investment Banking and Advising and Distribution Services for New Issues," *Journal of Finance*, 37, 955–976.

Baron, D., and D. Besanko, 1987, "Monitoring, Moral Hazard, Asymmetric Information, and Risk Sharing in Procurement Contracting," *Rand Journal of Economics*, 18, 509–532.

Baron, D., and B. Holmstrom, 1980, "The Investment Banking Contract for New Issues Under Asymmetric Information: Delegation and the Incentive Problem," *Journal of Finance*, 35, 1115–1138.

Bernardo, A. E., H. Cai, and J. Luo, 2001, "Capital Budgeting and Compensation with Asymmetric Information and Moral Hazard," *Journal of Financial Economics*, 61, 311–344.

Caillaud, B., R. Guesnerie, and P. Rey, 1992, "Noisy Observation in Adverse Selection Models," *Review of Economic Studies*, 59, 595–615.

Chung, K. L., 1974, *A Course in Probability Theory* (2d ed.), Academic Press, San Diego, CA.

Darrough, M. N., and N. M. Stoughton, 1986, "Moral Hazard and Adverse Selection: The Question of Financial Structure," *Journal of Finance*, XLI(2), 501–513.

DeMarzo, P., and D. Duffie, 1995, "Corporate Incentives for Hedging and Hedge Accounting," *Review of Financial Studies*, 8, 743–771.

Demski, J. S., and R. A. Dye, 1999, "Risk, Return and Moral Hazard," *Journal of Accounting Research*, 37, 27–55.

Detemple, J., S. Govindaraj, and M. Loewenstein, 2001, "Hidden Actions, Agents with Non-Separable Utility and Wealth-Robust Intertemporal Incentive Contracts," Working Paper, Boston University, Boston, MA.

Dybvig, P., H. Farnsworth, and J. Carpenter, 2001, "Portfolio Performance and Agency," Working Paper, Washington University, St. Louis, MO.

Faynzilberg, P. S., and P. Kumar, 2000, "On the Generalized Principal-Agent Problem: Decomposition and Existence Results," *Review of Economic Design*, 5, 23–58.

Fudenberg, D., B. Holmstrom, and P. Milgrom, 1990, "Short-Term Contracts and Long-Term Agency Relationships," *Journal of Economic Theory*," 51, 1–31.

Fudenberg, D., and J. Tirole, 1991, *Game Theory*, MIT Press, Cambridge, MA.

Garcia, D., 2002, "Monotonicity in Direct Revelation Mechanisms," Working Paper, Dartmouth College, Hanover, NH.

Hellwig, M., and K. M. Schmidt, 2002, "Discrete-Time Approximations of Holmstrom-Milgrom Brownian-Motion Model of Intertemporal Incentive Provision," *Econometrica*, 70, 1139–1166.

Holmstrom, B., and P. Milgrom, 1987, "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303–328.

Holmstrom, B., and J. Ricart i Costa, 1986, "Managerial Incentives and Capital Management," *Quarterly Journal of Economics*, 101, 835–860.

Jewitt, I., 1988, "Justifying the First-order Approach to Principal-agent Problems," *Econometrica*, 56, 1177–1190.

Karatzas, I., and S. E. Shreve, 1987, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

Kaplan, R., 1984, "The Evolution of Management Accounting," *The Accounting Review*, LIX, 390–418.

Laffont, J.-J., and J. Tirole, 1986, "Using Cost Observation to Regulate Firms," *Journal of Political Economy*," 94(3), 614–641.

Lambert, R. A., 1986, "Executive Effort and Selection of Risky Projects," *Rand Journal of Economics*, 17, 77–88.

Maskin, E., and J. Riley, 1984, "Optimal Auctions with Risk Averse Buyers," *Econometrica*, 52, 1473–1518.

Matthews, S. A., 1983, "Selling to Risk Averse Buyers with Unobservable Tastes," *Journal of Economic Theory*, 30, 370–400.

McAfee, R. P., and J. McMillan, 1986, "Bidding for Contracts: A Principal-Agent Analysis," *Rand Journal of Economics*, 17, 326–338.

Mirrlees, J. A., 1999, "The Theory of Moral Hazard and Unobservable Behavior: Part 1," *Review of Economic Studies*, 66, 3–21.

Müller, H., 1998, "The First-Best Sharing Rule in the Continuous-Time Principal-Agent Problem with Exponential Utility," *Journal of Economic Theory*, 79, 276–280.

Müller, H., 2000, "Asymptotic Efficiency Dynamic Principal-Agent Problems," *Journal of Economic Theory*, 91, 292–301.

Ou-Yang, H., 2003, "Optimal Contracts in a Continuous-Time Delegated Portfolio Management Problem," *Review of Financial Studies*, 16, 173–208.

Picard, P., 1987, "On the Design of Incentive Schemes under Moral Hazard and Adverse Selection," *Journal of Public Economics*, 33, 305–331.

Revuz, D., and M. Yor, 1999, *Continuous Martingale and Brownian Motion*, Springer, New York.

Rey, P., and B. Salanié, 1996, "On the Value of Commitment with Asymmetric Information," *Econometrica*, 64, 1395–1414.

Rogerson, W., 1985, "The First-Order Approach to the Principal-Agent Problems," *Econometrica*, 53, 1357–1367.

Salanié, B., 1990, "Sélection Adverse et Aversion pour le Risque," *Annales d'Economie et de Statistque*, 18, 131–149.

Schättler, H., and J. Sung, 1993, "The First-Order Approach to Continuous-Time Principal-Agent Problem with Exponential Utility," *Journal of Economic Theory*, 61, 331–371.

Schättler, H., and J. Sung, 1997, "On Optimal Sharing Rules in Discrete- and Continuous-Times Principal-Agent Problems with Exponential Utility," *Journal of Economic Dynamics and Control*, 21, 551–574.

Stole, L. A., 2001, *Lectures on the Theory of Contracts and Organizations*, http://gsblas.uchicago.edu/papers/lectures.pdf.

Sung, J., 1991, "Managerial Compensation and Corporate Management Under Moral Hazard," (Ph.D. thesis), Washington University, St. Louis, MO.

Sung, J., 1992, *Corporate Insurance and Optimal Contract*, mimeo, UIC, Chicago, IL.

Sung, J., 1995, "Linearity with Project Selection and Controllable Diffusion Rate in Continuous-Time Principal-Agent Problems," *Rand Journal of Economics*, 26, 720–743.

Sung, J., 1997, "Corporate Insurance and Managerial Incentives," *Journal of Economic Theory*, 74, 297–332.

Copyright of Review of Financial Studies is the property of Society for Financial Studies and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.