MORAL HAZARD FROM COSTLESS HIDDEN ACTIONS\textsuperscript{1}

Martin Byford
Department of Economics and Finance
La Trobe University

June 2003

\textsuperscript{1}I am grateful to Chris Adam, Suren Basov, Ralph Bayer, Harry Clarke, Stephen King, Xiangkang Yin, and the participants in the 21\textsuperscript{st} Australian Economic Theory Workshop, and workshops at The University of Melbourne, La Trobe University and Adelaide University, for their comments on earlier versions of this paper. I would particularly like to thank Michael Harris and David Prentice for their guidance and support. All errors are the author’s.
Abstract

Principal-agent models typically rely on the assumption that the agent’s action has a positive and increasing marginal cost in order to explain the emergence of a moral hazard. This paper develops a model in which an agent can manipulate a project’s type, and in particular the project’s risk, through a costless hidden action. It is shown that even though the action is costless, the agent’s career concerns may give rise to preferences over the type space that deviate from those of the principal. With the agent’s action hidden, these preferences create a moral hazard.
For more than two decades, the development of moral hazard theory has been limited by the use of three confining assumptions. The primary purpose of these assumptions has been to render moral hazard models concave, thus validating the use of the analytically more convenient first-order approach. However, while the Mirrlees-Rogerson conditions\(^1\) — the twin assumptions that a production technology possesses both a Monotone Likelihood Ratio (MLR) and a Convex Distribution Function (CDF) — in conjunction with the assumption that the marginal cost of the agent’s action is positive and increasing, have been shown to be sufficient to ensure this concavity, none of the three is necessary. Nevertheless, the assumptions remain a standard feature of moral hazard models because they fit well with the common characterization of moral hazard as a problem of motivating a risk-averse agent to exert costly effort, when the amount of effort that the agent exerts is hidden. When the agent’s action is taken to represent the effort that he supplies, the Mirrlees-Rogerson conditions imply that increasing effort will everywhere have a beneficial effect on a project’s outcome, although at a decreasing rate, while the costly effort assumption can be taken to represent the agent’s effort aversion.

Yet, convenient as it may be to restrict the discussion of moral hazard to a costly effort paradigm, there exist other classes of hidden actions available to an agent, where the assumptions associated with this paradigm are not reasonable. One such class of problem, particularly relevant to discussions of executive compensation, is where the action required of the agent is to select a project’s type. It is plausible, if not probable, that the private cost

\(^1\)See Mirrlees [13] and Rogerson [14]. Also Jewitt [10].
that an agent incurs when selecting a project’s type will be independent of the type that is chosen. On its own, a costless hidden action is not sufficient to create a moral hazard. However, if the choice of type will affect not only the expected outcome of a project, but also the risk associated with that outcome, then a risk-averse agent who is exposed to the risk might find his desire to improve the expected outcome conflicting with his preferences regarding the risk. In turn, this may lead the agent to take an action that deviates from the optimum.

It is easy to understand why costless hidden actions have not been incorporated into static moral hazard models. In a static model, the optimal contract would involve the agent receiving a fixed transfer, and insulated from risk, the agent would weakly prefer to select the project’s type to maximize the expected outcome. In a dynamic context the situation is more complicated as the agent is exposed to additional sources of incentives. Holmstrom and Ricart i Costa [9] have shown how an agent’s career concerns can create incentives over risk, even when the agent receives a fixed wage in each period. Following the standard assumptions of the learning model literature, they assume that the market learns about an agent’s ability by observing the outcomes that he produces in each period. As a consequence, the agent’s future earnings become a function of these outcomes, exposing the agent to the risk associated with an outcome regardless of the structure of his wage. Yet despite their ability to capture the moral hazard that may emerge when the agent’s action is costless but hidden, dynamic models, such as those proposed by Holmstrom [8], and Dewatripont, Jewitt and Tirole [2], continue to rely on the assumption that the agent’s action has a positive and increasing
marginal cost, to explain the emergence of the moral hazard. Indeed, the technology in Holmstrom’s [8] dynamic model of moral hazard is such that the agent is explicitly prevented from using his action to manipulate the risk associated with a project.

This paper presents a dynamic model in which the action required of the agent is to select a project’s type. The choice of type is hidden but costless, however the agent’s career concerns are such that they may create incentives over risk, potentially distorting the action. Where career concerns do lead to perverse behavior, it is shown that an explicit incentive contract can be used to improve the expected outcome, although such a contract will not generally be first-best. Moreover, to accommodate the possibility that the expected outcome maximizing type may be internal to the type space, it is not assumed that the production technology displays a MLR. Instead, an alternative method for ensuring the concavity of the problem, based in part on the practicalities of contracting, is proposed.

This model is presented as a baseline. By showing that within a dynamic context, a costless hidden action can give rise to a moral hazard, it becomes clear that moral hazard models can be applied to broader classes of problems that lie beyond the traditional focus of the costly effort paradigm. This model can be used to explain why managers in some industries seem inclined to reduce the risk associated with their projects, even where this intervention reduces expected outcomes. It is also possible to explain how the career structures in other industries may lead risk-averse managers to recklessly pursue risky projects. Unlike some models that presuppose a particular contract structure in order to create distortions in an agent’s action,
it is shown here that these perverse behaviors can remain under optimal contracting. Furthermore, any system of preferences that the agent may have over the type space — including, but not restricted to effort aversion — can be placed over the baseline incentives that arise from career concerns, expanding the possible applications of this model still further.

The paper proceeds in four parts. Section 1 presents a general model of the agent’s career concerns, showing how these concerns can create incentives over the risk involved in a project. In any period the incentives are equivalent to static preferences over the type space, reducing the dynamic principal-agent problem to a familiar static form. The equivalent static program is then analyzed in section 2. Section 3 presents an example that allows the contract to take a particular form. While this example is specific, the results are typical of an important class of problem. The paper concludes with a discussion of contract design and concavity.

1 Career Concerns and Risk

A risk-neutral principal employs a risk-averse agent to deliver a project. The agent is responsible for selecting the project’s type ex-ante. The agent does not have any endogenous preferences over the choice of type, and neither the principal nor the market are able to observe the agent’s choice. The project’s type is denoted by a vector of n orthogonal parameters; \( a^T = \{a_1, a_2, \ldots, a_n\} \). The type is selected from the choice set \( A \) which is in turn a subset of \( R^n \). The outcome of the project \((x)\) is stochastic and contingent upon both the agent’s choice of \( a \) and the agent’s ability \( \theta \). The distribution of \( x \) is given by the function \( F(x|a, \theta) \).
Assumption 1 \( F(x|a, \theta) \) is continuous and twice differentiable over the outcome space \( X \subset \mathbb{R} \), the ability space \( \Theta \subset \mathbb{R} \), and \( A \). The corresponding density function \( f(x|a, \theta) \) is likewise twice differentiable and continuous in all three domains. Finally \( f(x|a, \theta) > 0 \) for all \( x, a \) and \( \theta \).

Beyond the conditions stated in assumption 1, no further assumptions concerning the interaction of \( a \) and \( \theta \) in the production technology need be made. Importantly, the possibility that \( a \) and \( \theta \) are not independent, and as such the cross derivative \( f_{a\theta} \) is not equal to zero, can be accommodated.

1.1 Learning

Neither the principal nor the market can directly observe the agent’s true ability. Instead, they rely upon the outcomes that an agent produces in each period, to act as signals of that ability. Consider an agent who will live and work for \( T \) periods. Assume that in the first period, the market holds the prior belief \( \theta_1 \). Following the first period, the market uses the outcome \( x_1 \), produced by the agent, to update their collective belief about the agent’s ability. In this way, the market’s second period prior belief \( \theta_2 \), is a function of both the first period prior \( \theta_1 \), and the first period outcome \( x_1 \). In any subsequent period \( t \), the prior belief about the agent’s ability is a function of the first period prior and the history of outcomes that the agent has produced: \( \theta_t(\theta_1, x_1, \ldots, x_{t-1}) \). It is not necessary to assume either a particular production technology, or method of updating, at this point. We leave open the possibility that the agent’s ability may improve over time such as is the case in Gibbons and Waldman’s [4] learning model of promotion and wage dynamics. There is, however, one assumption that is imposed.
**Assumption 2** The likelihood ratio $f_\theta/f$ is everywhere increasing in $x$, and as such higher ability is everywhere beneficial to any project’s outcome in the sense of first-order stochastic dominance.

With the likelihood ratio $f_\theta/f$ monotone in $x$, higher outcomes always imply higher ability, and thus we can conclude that rational updating of beliefs will leave $\theta_t$ increasing in every $x_i$ for $i < t$. Moreover, with ability always having a beneficial effect on productivity, the market value $\bar{s}_t$, of an agent at time $t$, in a sufficiently competitive labour market, will be increasing in that agent’s perceived ability. It follows that $\bar{s}_t$ can also be written as a function of the first period prior and the history of outcomes, and expressed as such $\bar{s}_t(\theta_1, x_1, \ldots, x_{t-1})$ will likewise be increasing in all $x_i$.

Looking forward from period $t$, the market’s beliefs about the agent’s ability in all subsequent periods — and thus the agent’s market value — will be an increasing function of the outcome $x_t$. With the outcome the product of a stochastic process, an agent who is forced to sell his labour at market value in each period will still carry risk because of the link between the stochastic outcome and his future market value.

**Assumption 3** The agent is risk-averse hence $u' > 0$ and $u'' < 0$.

If the agent is prevented from saving or borrowing between periods\(^2\) then the utility that the agent receives in period $t$ will be the utility that he

---

\(^2\)The assumption that financial markets are unwilling to advance an agent unsecured loans against his future income is strict, but not particularly unrealistic. More problematic is the assumption that the agent may not save for consumption in future periods. While this assumption has been imposed in the interests of simplifying the analysis, if the construction of the model is such that the agent always expects his wage to increase in the next period then the agent will not generally have an incentive to save.
extracts from his share \(s_t(x_t)\) of the outcome that he produces in that period. With the principal risk-neutral, first-best risk sharing clearly requires the principal to accept all the risk associated with the outcomes produced by the agent. The problem is that even if the principal offers the agent a fixed wage, equal to his market wage, in some period, the agent is still affected by the outcome of the project, as the agent’s market value in all future periods is a function of the outcome that he produces. To completely insulate the agent from risk, the principal and the agent must be able to enter into a binding agreement in which the principal commits to employing the agent, at a pre-agreed fixed wage, in each remaining period of the agent’s working life.

In their treatment of a basic learning model, Harris and Holmstrom [6] assumed that agents are not able to sell themselves to firms. Where indentured servitude is not possible, an agent is free to accept higher offers from the market whenever they arise. On the other hand, Harris and Holmstrom assumed that firms will be able to provide agents with a guarantee that their wage will never fall from one period to the next. Their conclusion, that such a guarantee is optimal, does not fit well with the available evidence on the contracting practices of firms. For example, in their extensive study of the wage policy of a firm, Baker, Gibbs and Holmstrom [1] found that after working for the firm for ten years, 15% of the firm’s managers had experienced real wage declines. In more recent learning models, Farber and Gibbons [3], and Gibbons and Waldman [4], have assumed that the agent receives his market value in each period. However, these models do not address the problem of moral hazard, and as such they do not need to incorporate explicit incentives
as a means of influencing the agent’s behavior.

We will proceed from the assumption that the principal cannot credibly make commitments for future periods. This may be because the principal can act opportunistically to reduce the agent’s wage if it becomes apparent that the agent is less able than was originally thought, or because the firm is not infinitely lived, vulnerable to takeovers or bankruptcy in which existing contracts may be voided. We will further assume that in any future period the agent expects to receive utility equal to the utility that he would derive from receiving his market value with certainty; it is shown below that this will always be true. Under these assumptions, an agent’s ex-ante expectation of the utility that he will receive from participating in a project is,

\[
\int_X u(s_t(x_t)) f(x_t|a, \theta_t) \, dx_t + \sum_{i=t+1}^T \beta^{i-t} \int_X u(\bar{s}_i(x_t)) f(x_t|a, \theta_t) \, dx_t - \sum_{i=t+1}^T \beta^{i-t} u(E[\bar{s}_i|\theta_t]),
\]

where \( \beta < 1 \) is the rate at which the agent discounts the future. The first term in (1) is the expectation of the utility that the agent will derive from his share of the outcome \( x_t \). The second and third terms represent the deviation of the agent’s expected utility in future periods, from the prior expectations of that utility given the prior belief of the agents ability; the agent’s career concerns. It follows from assumption 2 that the second term is unambiguously increasing in \( x_t \). Furthermore, the third term can be treated as a constant, as it is a function of prior expectations.
1.2 Incentives Over Risk

Conceptually (1) can be divided into two parts. The first term is a function of an explicit incentive contract, where the form of the sharing rule $s_t(x_t)$ is determined by the principal ex-ante. The principal may not need to include explicit incentives in the sharing rule, choosing instead to offer the agent a fixed wage contract. Explicit incentives will only be required if the second and third terms in (1) would cause the agent’s action to deviate from the optimum. The second and third terms represent the agent’s career concerns. With contracts limited to commitments in the current period, the principal does not have any control over the second and third terms. The career concerns are derived entirely from the technology $F(x|a, \theta)$, and the structure of the labour market.

As higher outcomes always implying higher ability, the career concerns encapsulated in (1) create implicit incentives for the agent to maximize the expected outcome of the project. However, the agent’s career concerns may also produce preferences over risk if the $u(s_i(x_t))$ functions are non-linear. If the $u(s_i(x_t))$ terms are all concave transformations of $x_t$, then risk will be costly for the agent. Consequently, ceteris paribus the agent can improve the expectation of his future utility by acting to reduce the risk. Conversely, if the $u(s_i(x_t))$ terms are convex transformations of $x_t$, then increasing risk will increase the agent’s expected future utility.

Determining the behaviour of each $u(s_i(\theta_i(x_t)))$ is not straightforward because every term comprises a series of three consecutive transformations of $x_t$. Working from the outside in, $u$ is an increasing concave transformation
that increases the cost of risk to the agent, leaving the agent less partial to risk than if we were considering $\delta_i(\theta_i(x_t))$ alone. As a function of perceived ability, $\delta_i(\theta_i(x_t))$ reflects the improvement in expected outcomes that is associated with employing agents with higher apparent ability.\footnote{In a perfectly competitive labour market we could go further and say that $\delta_i(\theta_i) = E[x_t|\theta_i]$.} While $\delta_i$ will be, at the very least, non-decreasing in its arguments, whether the various $\delta_i$ terms will be concave, convex or some combination of both, is a function of both the production technology and the structure of the labour market. Finally we consider $\theta_i(x_t)$. The dependency of $\theta_i$ on $x_t$ is a product of the way in which the market updates its beliefs about the agent’s abilities, and may take many forms. Subject to our previously stated assumptions, $\theta_i$ will be increasing in $x_t$. Moreover, as any particular outcome becomes less important to the updating of the market’s perceptions as time goes on, we might reasonably assume that $\theta_i$ becomes less sensitive to $x_t$ with $i$. All in all, a series of confining assumptions must be made about each $u(\delta_i(\theta_i(x_t)))$ term in order to assume the nature of the preferences created by the agent’s career concerns. However, with $u$ always concave, it seems more likely that the second term in (1) will leave an agent averse to a project’s risk. This fits well with the prominence of anecdotes from many industries, about managers who manipulate the type of projects in order to reduce the associated risk.

The fact that career concerns can influence an agent’s attitudes toward a project’s risk, is not, on its own, sufficient to distort that agent’s action. In two cases in particular, the choice of type will be trivial. If $a$ influences the mean of the outcome, while leaving the higher order moments unchanged,
then the agent will select the type to maximize the expected outcome as $a$ has no control over the project’s risk. Similarly, if the higher order moments are functions of $a$, and the mean is independent of the type, the agent can select $a$ to manipulate the risk without interfering with the principal’s interests. A moral hazard will only exist if the choice of type presents the agent with a tradeoff in which he has to reduce the expected outcome in order to manipulate a project’s risk.

In order to simplify the discussion, the career concerns in (1) can be reduced to a set of static preferences over the choice set. This is possible because of the assumption that the principal cannot make commitments concerning the form of agent’s contract in subsequent periods. By carrying through the integral in the second term of (1) we can eliminate $x_t$ and express the agent’s career concerns as a function of $a$ alone. The resulting transformation can be expressed as,

$$\psi(a) = \sum_{i=t+1}^{T} \beta^{i-t} \int x u(\bar{s}_i(x_t)) f(x_t|a, \theta_t) \, dx_t - \sum_{i=t+1}^{T} \beta^{i-t} u(E[\bar{s}_i|\theta_t]). \tag{2}$$

Note that because the sharing rule $s_t(x_t)$ does not appear in (2), the only choice variable in $\psi$ is $a$. The preferences brought about by the agent’s career concerns are therefore independent of any action that is available to the principal. If $u$, $\bar{s}_i$ and $\theta_i$ are all continuous in their arguments, and the choice set $A$ is closed, then there will exist at least one $a$ in $A$, for which $\psi(a)$ is maximized. The necessary and sufficient condition for the emergence of a moral hazard, is that the type $a^\psi$, that maximizes $\psi(a)$, not equal the type $a^\ast$, that maximizes the expected outcome alone. In this model both $a^\psi$ and $a^\ast$ may be strictly internal to the choice set.
2 The Program

In the costly effort paradigm, the cost of providing effort is endogenous to the agent, representing his inherent effort-aversion. In contrast, the preferences described by the $\psi$ function are the result of exogenous career concerns. Nevertheless, $\psi(a)$ is analogous to a cost function from static costly effort models, allowing a static form of the program to be used here.\(^4\) The standard form of this program is due to Mirrlees [12], while the analysis employed follows Holmstrom [7] and Jewitt [10]. Time indices, and the ability terms, have been suppressed for notation simplicity.

In some period $t$, the principal is seeking to,

$$\max_{a,s} \int_X (x-s(x))f(x|a)\,dx,$$  \hspace{1cm} (3)

subject to the participation constraint,

$$\int_X u(s(x))f(x|a)\,dx + \psi(a) \geq u(\bar{s}),$$  \hspace{1cm} (4)

and the incentive compatibility constraints,

$$\int_X u(s(x))f_{a_i}(x|a)\,dx + \psi_{a_i}(a) = 0, \quad \forall i \in \{1, \ldots, n\}. \hspace{1cm} (5)$$

The sign on the $\psi$ function is positive, instead of the negative that is more familiar from costly effort models. This is because $\psi(a)$ captures the benefits to the agent of selecting the type $a$ — in terms of the change to his expected future utility — rather than a cost, as is usual in the costly effort paradigm. The LHS of the participation constraint (4), is identical to (1), and as such

\(^4\)The problem remains a dynamic one, it is just that all the dynamic elements are suppressed by the transformation in (2).
(4) states that the agent will not accept a contract unless its expected value can at least match that of the contracts that the agent is being offered in the market. For the purposes of this model, the possibility that the agent would prefer leisure to some contracts, is not considered. The incentive compatibility constraints state that the agent will select $a$ to maximise his expected utility. As we have seen, the characterization of the incentive compatibility constraint in (5), as a set of $n$ first-order conditions, is only valid if the program is concave.

2.1 A Digression Concerning Concavity

Jewitt [10] has shown that the expectation of a non-decreasing function of $x$, will be concave in $a$, if the production technology displays a CDF. For an $n$ dimensional type vector, the distribution function $F$ will be convex if the Hessian matrix,

$$
\begin{bmatrix}
F_{a_1a_1} & F_{a_1a_2} & \cdots & F_{a_1a_n} \\
F_{a_2a_1} & F_{a_2a_2} & \cdots & F_{a_2a_n} \\
\vdots & \vdots & \ddots & \vdots \\
F_{a_na_1} & F_{a_na_2} & \cdots & F_{a_na_n}
\end{bmatrix},
$$

is positive definite.\footnote{Jewitt [10] proposes an alternative to a CDF that would also be valid here.} With $\sum \beta^{i-t} \int u(s_i(x_t))f \, dx_t$ non-decreasing in $x_t$, a CDF will be sufficient to render $\psi(a)$ concave in $a$. Likewise, a CDF will render the expected outcome $E[x]$, a concave function of $a$. Given this fact, a CDF has an intuitively appealing interpretation: It implies that moving the type toward the expected outcome maximizing type $a^x$, will always increase the expected outcome, although at a rate that decreases with proximity to $a^x$. 
The remainder of the agent’s objective function (1), will be concave if the sharing rule is non-decreasing in \( x \). In costly effort models it is typically assumed that the likelihood ratio \( f_a / f \) is monotone non-decreasing in \( x \). Through optimization, this monotonicity translates to the optimal sharing rule. However, if \( a^x \) lies in the interior of the choice set, then in some region increasing \( a \) will reduce \( E[x] \), which is inconsistent with the likelihood ratio being monotone non-decreasing in that region.\(^6\) One possible alternative to a MLR, that would render the sharing rule non-decreasing in \( x \), is proposed in the final section. However, any set of conditions that are sufficient to render the problem concave, will validate the analysis employed in this paper.

### 2.2 Optimization

It is apparent from (5), that in the absence of an incentive contract, the agent will select \( a \) equal to \( a^\psi \). If \( a^\psi \) is not equal to \( a^x \), then it is only by creating additional incentives over the choice set that the principal can begin to align the agent’s preferences with his own. Given the restrictions on contracting, this means offering the agent an explicit incentive contract. The optimal incentive contract can be found by solving the program (3) – (5).\(^7\) The Lagrangian for this program is,

\[
\mathcal{L} = \int_X (x - s) f \, dx + \lambda \int_X u(s) f \, dx + \sum_{i=1}^{n} \mu_i \int_X u(s) f_{a_i} \, dx, \tag{7}
\]

where \( \lambda \) is the Lagrange multiplier on (4) and \( \mu^T = \{ \mu_1, \ldots, \mu_n \} \) is the vector of multipliers on the \( n \) constraints in (5). Pointwise optimization in \( s \) then

---

\(^6\)Rogerson [14] has proved that where a production technology has a MLR, an increase in \( a \) will improve the outcome in the sense of first-order stochastic dominance. Milgrom [11] offers a further extensive discussion of the economic significance of a MLR.

\(^7\)With the agent’s action hidden, this contract will be second best.
yields the identity,
\[
\frac{1}{w'(s(x))} = \lambda + \sum_{i=1}^{n} \mu_i \frac{f_{a_i}(x|a)}{f(x|a)}.
\] (8)

Our first concern is to determine whether the participation constraint is binding.

**Proposition 1** At the optimum, the participation constraint (4) will always be satisfied with equality.

**Proof** Jewitt [10] has shown \( \lambda \) to be strictly positive. It follows from the Kuhn-Tucker conditions that the participation constraint (4) must be binding.\(^8\) QED

From proposition 1 we see that in any period \( t \) the ex-ante expectation of the agent’s utility will be equal to the utility that the agent would gain from receiving his market value with certainty. This result does not rely on the previously stated assumption that such a constraint will likewise be binding in all subsequent periods, as \( \psi(a) \) does not appear in (8). It can therefore be concluded that in any period of the agent’s lifetime, the agent’s expected utility will be determined by a binding participation constraint.

The behavior of the vector of multipliers \( \mu \) is also of interest due to its role in determining the direction in which \( s \) depends on \( x \). Taking partial derivatives of (7) in each \( a_i \) yields the system of equations,

\[
\begin{bmatrix}
\mu_1 \\
\mu_2 \\
\vdots \\
\mu_n
\end{bmatrix} = -
\begin{bmatrix}
v_{11} & v_{12} & \ldots & v_{1n} \\
v_{21} & v_{22} & \ldots & v_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
v_{n1} & v_{n2} & \ldots & v_{nn}
\end{bmatrix}^{-1}
\begin{bmatrix}
\int (x - s(x)) f_{a_1} \, dx \\
\int (x - s(x)) f_{a_2} \, dx \\
\vdots \\
\int (x - s(x)) f_{a_n} \, dx
\end{bmatrix},
\] (9)

\(^8\)Grossman and Hart [5] offer a different proof, in the context of costly effort, that is equally valid here.
where \( v_{ij} = \int u(s)f_{a_ia_j} \, dx + \psi_{a_ia_j} \). The Lagrange multiplier \( \lambda \) does not appear in (9) as each \( \int u(s)f_{a_i} \, dx + \psi_{a_i} = 0 \) by (5). Note that the matrix \([v]\) is in fact the Hessian of the agent’s objective function. A convenient feature of having forced the program to be concave is that this Hessian will be negative definite, and therefore \( \text{det}[v] \) must be non-zero.\(^9\) With the matrix \([v]\) invertible, none of the multipliers \( \mu_i \) will be undefined.

The \( \mu \) terms are central to the form of the optimal sharing rule. With \( u \) concave, the LHS of (8) is increasing in \( s \). The sharing rule will therefore be increasing in \( x \) if, and only if, the weighted sum of likelihood ratios \( \sum \mu_i f_{a_i}/f \), is increasing in \( x \). Each likelihood ratio \( f_{a_i}/f \) indicates the extent to which an outcome is indicative of the agent having selected a high value of \( a_i \). Where \( \mu_i \) is positive, the optimal sharing rule rewards the agent when it appears that he has selected a high value of \( a_i \). This in turn provides an incentive for the agent to select \( a_i \) greater than \( a_i^\psi \), his preferred action in the absence of an explicit incentive contract. In contrast, where \( \mu_i \) is negative, the optimal contract penalizes the agent when the outcome implies a high value of \( a_i \), leading the agent to select \( a_i < a_i^\psi \).

Interestingly enough, there is no generally applicable result that prevents a term \( \mu_i \) from being negative. This contrasts sharply with static costly effort models where \( \mu \) is unambiguously positive. In the costly effort paradigm each parameter always has a positive and increasing marginal cost. Therefore, in the absence of countervailing incentives the agent will want to minimize each parameter, and the type \( a^\psi \) that minimizes the agent’s aggregate cost function must lie on the boundary of the choice set \( A \), as the cost of any type

\(^9\)In fact \( \text{det}[v] \) will be strictly negative for \( n \) odd, and strictly positive for \( n \) even.
that is internal to the choice set can always be further reduced by reducing
the value of one or more of the parameters. Likewise, with effort always
beneficial to the outcome, the type $a^x$ that maximizes the expectation of
that outcome must also lie on the boundary of $A$, although typically on the
opposite side of the choice set to $a^\psi$. The principal therefore benefits, in terms
of the outcome, from encouraging the agent to increase his choice on each
parameter. It follows that subject to costly effort paradigm assumptions,
each term in $\mu$ will be positive.

The model presented in this paper is considerably more flexible. The
types $a^\psi$ and $a^x$ may lie on the boundary of $A$, or they may be internal to
the choice set. Subject to the behavior of $f(x|a)$ as it varies with each pa-
rameter, and the agent’s attitude to risk, it is possible that $a^\psi_i$ may be greater
than $a^x_i$ for some $i$, and less than $a^x_i$ for others. Moreover, depending upon
the location of the optimal type $a^{OP}$,\footnote{The optimal type $a^{OP}$ is defined as the type that solves the optimization program (3) — (5).} relative to $a^\psi$, it may be necessary
to reduce the agent’s preference for some parameter. In this case the cor-
responding multiplier $\mu_i$ will be negative, and the optimal sharing rule will
punish outcomes that indicate a high choice of $a_i$, while rewarding outcomes
that suggest that $a_i$ was small.

2.3 Discussion

We would like to be able to develop some general rule that would locate the
optimal type within $\mathbb{R}^n$. It would be particularly convenient if this rule could
be stated in terms of the two existing points of reference, $a^\psi$ and $a^x$. Such
a rule is developed in the example below, for the case where $n = 1$. Unfor-
fortunately, when \( n > 1 \), knowing where \( a^\psi \) and \( a^x \) lie, is not even sufficient to confine the location of \( a^{OP} \) within a bounded region. Indeed, in the absence of further very strict assumptions concerning the distribution \( F(x|a) \), it remains possible for \( a^{OP} \) to lie almost anywhere in \( \mathbb{R}^n \). The uncertainty regarding the location of \( a^{OP} \) is problematic for the analysis employed in this paper, as if the stationary point does not occur within the choice set, the first-order approach would not be valid. This problem can be addressed by assuming that a stationary point exists in \( \mathbb{R}^n \), and that \( A \) is of sufficient size to contain it, although these assumptions are not particularly natural.

At first glance, using a single outcome as a signal for the agent’s choice of the \( n \) orthogonal parameters, may also seem problematic. Some outcomes may imply that the agent selected one parameter \( a_i \) in accordance with the principal’s wishes, while at the same time indicating that the choice of a second parameter \( a_j \) was undesirable. For that \( x \), the terms \( \mu_i f_{a_i}/f \) and \( \mu_j f_{a_j}/f \) would act in opposite directions, reducing the magnitude of the associated incentives. However, it is worth keeping in mind that the \( n \) constraints in (5) ensure that there will only ever be one choice of type that is incentive compatible.\(^{11}\) Moreover, unless \( \sum \mu_i f_{a_i}/f = 0 \) for all \( x \), the optimal sharing rule will include explicit incentives and \( a^{OP} \) will not be equal to \( a^\psi \).

Many of the general results developed in the context of the costly effort paradigm remain valid in the model presented in this paper. Among such results are Holmstrom’s [7] Sufficient Statistic Theorem — that states that an optimal incentive scheme incorporating an informative signal will Pareto dominate an optimal incentive scheme that is not contingent on that signal

\(^{11}\) Always assuming that the program is concave with a stationary point.
and Holmstrom’s [7] core result that where the optimal sharing rule incorporates explicit incentives, the contract will be second best. As always, exposing a risk-averse agent to a project’s risk imposes a cost upon that agent. Because of the binding participation constraint, whenever the principal incorporates explicit incentives into the sharing rule, the principal must compensate the agent for the cost of bearing the associated risk by including a risk premium in the contract.

A further extension of the model, also compatible with the results mentioned above, comes from considering an agent’s endogenous preferences over a project’s type. The most obvious arise when some parameters are analogous to effort — and consequently are associated with a positive and increasing marginal cost — but other factors may also affect the agent’s action. Preferences over the type space might come from ideological or moral concerns, or even from some private benefit that the agent gains from working on a project that is in line with his intellectual interests. Let the agent’s endogenous preferences be represented by the function \( \phi(a) \); these preferences can be easily incorporated into the model by adding \( \phi(a) \) to the LHS of the participation constraint (4), and its partial derivatives to the \( n \) incentive compatibility constraints (5). Moreover, if \( \phi(a) \) is concave then the first-order approach remains valid.

3 Example

From (8) we can see that the structure of the optimal contract is heavily dependent upon on the likelihood ratios \( f_a/f \). Even given our previous assumptions concerning the distribution of \( F(x|a, \theta) \), there remain an almost
limitless number of possible structures for these ratios. In order to proceed further, and characterize the structure of the optimal sharing rule, we must make particular assumptions about the distribution of $x$. The example presented here has a simple construction but illustrates features of the optimal sharing rule that are common to a broad class of plausible distributions.

Consider the case of a manager facing a simple investment decision. The action required of the manager is to determine how much to invest in the project, with the one-dimensional type $a$ representing the magnitude of the investment that is chosen. Assume that the investment displays diminishing marginal expected returns to scale: Specifically, that the marginal expected return is $2a^2 - 2a$. Integrating, we find that the expected return can be expressed as $E[x] = 2a^2a - a^2$, a concave quadratic function with a stationary point at $a = a^x$. Furthermore, assume that the variance of the investment’s outcome is proportional to the amount invested such that $\text{var}[x] = a$. And finally, assume that the outcome $x$ is normally distributed.

The behavior of $f(x|a)$ creates a dynamic in which the project’s type influences both the mean and the variance of the outcome, simultaneously. Moreover, with the mean and the variance moving together in the region $a < a^x$, and in opposite directions elsewhere, an agent with incentives over risk will face a tradeoff in which his preferences lead him to select $a$ other than equal to the expected outcome maximizing value $a^x$. Figure 1 illustrates the behavior of $E[x]$ and $\text{var}[x]$ as they vary with $a$. If an agent’s career concerns, as defined by the RHS of (2), are concave in $x$, there will exist

---

12 It is assumed that there is no fixed cost or benefit that can be derived simply from participating in the project.
some $a < a^x$ where the marginal benefit, in terms of these career concerns, of increasing the expected return, will equal the marginal cost of increasing the variance. Therefore, for an agent whose career concerns leave him averse to risk, $a^\psi$ will be strictly less than $a^x$. This case is shown in panel (a) of figure 1. Conversely, when the agent’s career concerns are convex, and risk is therefore valuable in terms of expected future earnings, $a^\psi$ will be strictly greater than $a^x$. This case is shown in panel (b).

**Proposition 2** The optimal sharing rule will lead the agent to select the project’s type from the interval $[a^\psi, a^x]$.

**Proof** Begin by defining $a^P$ as the value of $a$ that maximises the principal’s return, as defined by (3), without constraints. From (9) we can see that $\mu$ will be positive when the optimal action $a^{OP} < a^P$, and negative when $a^{OP} > a^P$. From Jewitt [10] we get an alternative characterization of $\mu$,

$$\text{cov} \left[ \frac{1}{u'(s(x))}, u(s(x)) \right] = -\mu \psi'(a). \tag{10}$$

The LHS of (10) is unambiguously positive as both $u$ and $u'$ are monotone increasing in their common arguments. This leaves $\mu$ positive when $a^{OP} > a^\psi$, and negative when $a^{OP} < a^\psi$. To satisfy (9) and (10) simultaneously, $a^{OP}$ must lie in the interval $[a^\psi, a^P]$. This holds regardless of where $a^\psi$ lies in relation to $a^x$. Of course $a^P$ need not be equal $a^x$. If the optimal sharing rule is not linear in $x$ then $a^P$ may be either greater or less than $a^x$, depending upon whether $s$ is concave or convex. There are three cases to consider, corresponding to the three possible locations of $a^P$.

**Case 1** $a^P \in [a^\psi, a^x]$; If $a^P$ is in the interval $[a^\psi, a^x]$ then $a^{OP}$ must likewise lie in $[a^\psi, a^x]$.
Case 2 — \(a^\psi \in (a^P, a^x)\): This case can be eliminated as a possibility, as if \(a^\psi\) were to lie strictly between \(a^P\) and \(a^x\), the principal could strictly improve both \(E[x]\) and \(\psi(a)\) by paying the agent a fixed wage, leading the agent to select \(a = a^\psi\). By doing so the principal would further gain by no longer having to compensate the agent for carrying risk. With \(s'(x) = 0\) for all \(x\), \(a^P = a^x\) and therefore \((a^P, a^x) = \emptyset\).

Case 3 — \(a^x \in [a^\psi, a^P)\): The case where \(a^x\) lies within \([a^\psi, a^P)\) under the optimal contract, cannot be excluded as a possibility. However, were this the case, \(a^{OP}\) could not lie in \((a^x, a^P]\) as the principal could both increase \(E[x]\) and \(\psi(a)\), and reduce the compensation that the agent would require for carrying the risk, by reducing the incentive power of the contract,\(^\text{13}\) thus bringing the agent’s action into the interval \([a^\psi, a^x]\). QED

Further structure for the optimal contract can be derived by considering the likelihood ratio for the example. From the previously stated assumptions about the distribution of \(x\) — normal with \(E[x] = 2a^x a - a^2\) and \(\text{var}[x] = a\) — we get the density function,

\[
f(x|a) = \frac{1}{\sqrt{2\pi}a} \exp \left\{ -\frac{1}{2} \left( \frac{x - 2a^xa + a^2}{a} \right)^2 \right\}. \tag{11}
\]

The likelihood ratio is then,

\[
\frac{f_a(x|a)}{f(x|a)} = x^2 \frac{1}{2a^2} - x - 2(a^x)^2 + 4a^x a - \frac{3}{2} a^2 - \frac{1}{2a},
\]

which is a convex parabola in \(x\). It follows from (8) that the optimal sharing rule will be quasi-convex with a minima at \(x = (a^{OP})^2\). Figure 2 shows two possible sharing rules for this example. Panel (a) illustrates the form of the

\(^\text{13}\)By reducing \(|\mu|\) in (8).
optimal sharing rule when $\mu$ is positive, while panel (b) shows the case where $\mu$ is negative.

With the agent’s career concerns concave in $x$ — and therefore $a^{\psi}$ less than $a^x$, and $\mu$ positive — the principal encourages the agent to increase the value of $a$ that he selects by rewarding both high and low outcomes, while penalising median results. To see why this is so, consider figure 3. Panel (a) of figure 3 shows how the mean and variance of $x$ vary with the agent’s choice of action. Note the position of the action $a'$, less than the expected outcome maximizing action $a^x$. Panel (b) shows the density functions when the agent chooses either $a'$ or $a^x$ as his action. We see that as the action moves from $a'$ to $a^x$, the mean of the distribution increases. At the same time, the variance of $x$ is also increasing, redistributing mass from the centre to the tails of the distribution. If we were to increase the agent’s action further, the mean would begin to decline, yet the redistribution of mass would continue. Because increasing the action always increases risk, to encourage the agent to increase his action the principal must reward those outcomes that indicate risk taking; in this example both high and low outcomes. On the other hand, if $a^{\psi}$ is greater than $a^x$, and consequently the principal wants to discourage risk taking, then the optimal sharing rule penalises extreme outcomes, both high and low, and rewards median outcomes as indicators of a conservative attitude to risk.

Specific assumptions about the behavior of $f(x|a)$ have been made in this example, but as can be surmised from figure 3, many distributions that have a mean that is concave in $a$ with a stationary point internal to the choice set, and a variance that is monotone increasing in $a$, will display similar
properties. In cases where the increasing variance is constantly redistributing mass into the tails of a density function with a single central mode, the likelihood ratio will be convex with a turning point, and consequently the sharing rule will be quasi-convex with a turning point internal to the outcome space.

4 Implications for Contracting

An optimal sharing rule that is decreasing in $x$ in some regions, stands in stark contrast to the results of standard static costly effort models. By assuming that the distribution function has a MLR, the standard treatments of moral hazard with costly effort conclude that sharing rules should be non-decreasing in outcomes. Such results correlate much better with observed contractual practices than do models that suggest that the sharing rule should decrease in outcomes, in some region. That is not to say that the conclusion of this example is not intuitively pleasing. If, by increasing the risk associated with an outcome, the agent makes both high and low outcomes more likely, it makes sense that to encourage the agent to take more risks both high and low outcomes should be rewarded.

Problems emerge when moving beyond the limiting assumptions of the model. It is possible that the agent, as a hidden action, may have the capacity to sabotage the project on which he is working. With a sharing rule that is non-decreasing in the outcome, the agent would never want to do this. But if the agent is rewarded for low outcomes, and can ensure a low outcome though some act of sabotage, then there exists a powerful incentive for the agent to do just that. It is not unreasonable to assume that in order to
prevent the occurrence of perverse incentives, firms may prefer to construct sharing rules that are non-decreasing in outcomes. That is, firms may be preventing sabotage by refusing to reward agents for failures. This policy can be captured by incorporating the constraint,

$$s'(x) \geq 0,$$  \hspace{1cm} (12)

into the program. This constraint will only be binding in regions where (8) would leave the sharing rule decreasing in $x$. Elsewhere the non-decreasing sharing rule is completely defined by the identity (8), however this does not mean that the constraint (12) will not influence the form of the non-decreasing sharing rule in these regions. Recall from (9), that $\mu$ is defined by a series of integrals of functions of $s(x)$, where the integrals are taken across the entire outcome space. If the behavior of $s(x)$ changes in some part of that space then it may change the magnitude of the various $\mu_i$ terms. In turn, this means that the non-decreasing sharing rule $\hat{s}(x)$ will be an increasing transformation of the optimal sharing rule where $s'(x) > 0$. In the regions where $s'(x) < 0$, and the constraint (12) is binding, we of course get $\hat{s}'(x) = 0$.

The optimal sharing rule uses all of the information contained in the outcome to draw inferences about the action taken by the agent. If the non-decreasing sharing rule constraint is imposed, and is binding in some region of the outcome space, then the information carried by the outcomes in that region is ignored by the non-decreasing sharing rule. For this reason, a contract based on the optimal sharing rule derived from (8) will Pareto dominate a contract based on $\hat{s}(x)$.

A second important issue of construction concerns the concavity of the
problem. The incentive compatibility constraint for our problem is more correctly stated as,

\[
\arg \max_{a' \in A} \int_X u(s(x)) f(x|a) \, dx + \psi(a),
\]

which is, in fact, an infinity of constraints in the form,

\[
\int_X u(s(x)) f(x|a') \, dx + \psi(a') \geq \int_X u(s(x)) f(x|a) \, dx + \psi(a), \quad \forall a \in A.
\]

It is only when the problem is concave that we can replace (13) with the first-order conditions in (5). As mentioned above, the problem will be concave if the sharing rule is non-decreasing in \(x\). Therefore, concavity will be assured if the constraint \(s'(x) \geq 0\) has been implemented in order to sabotage proof the project. However, having a non-decreasing sharing rule is not a necessary condition. The necessary condition is,

\[
\int_X s'(x)u'(s(x)) F_{aa}(x|a) \, dx > 0.
\]

\(F\) and \(u'\) are always positive by assumption. Yet even if \(s'\) is negative in some regions, the LHS of (15) can remain positive if the regions over which \(s'\) is positive are sufficiently large. Unfortunately, (15) does not translate well into a simple and intuitively pleasing condition that can be imposed upon the distribution.

This paper has presented a very basic moral hazard model. Numerous extensions are possible, tying this model to both the costly effort and learning model literatures. Such is the flexibility of the model that it should prove to be compatible with all learning models in which the agent receives his
market value in each period. Likewise, this model is consistent with some of
the key results of the moral hazard literature such as the value of additional
information (Holmstrom, [7]). It is worth considering how other incentives
might impact upon the agent’s choice of type. For example, an executive may
derive an endogenous benefit from improving his public profile, or he might
be seeking ancillary opportunities such as a political career. Depending on
the structure of these payoffs, they may also affect the agent’s attitude to
risk.

References


cerns, Part I: Comparing Information Structures, Review of Economic

[3] H. Farber and R. Gibbons, Learning and Wage Dynamics, Quarterly

namics Inside Firms, Quarterly Journal of Economics 114 (1999), 1321-
1358.


Figure 1: Behavior of Mean and Variance of $x$
Figure 2: The Optimal Sharing Rule

(a) $a^\psi < a^x$

(b) $a^\psi > a^x$
Figure 3: Change in $f(x|a)$ with $a$