Econometrica, Vol. 56, No. 5 (September, 1988), 1177–1190.

JUSTIFYING THE FIRST-ORDER APPROACH TO PRINCIPAL-AGENT PROBLEMS

BY IAN JEWITT[1]

The Mirrlees-Rogerson conditions for the first-order approach to principal-agent problems to be valid do not work if the principal can observe more than one relevant statistic. This is a problem since much of the literature since Holmstrom (1979) deals with precisely such a case. We also argue that the Mirrlees-Rogerson assumption that the distribution function of output is convex in the agents action is unsatisfactory even in the context of the basic model; it is too restrictive, being satisfied by very few of the distributions occurring in statistics textbooks. The present paper replaces this assumption and provides conditions which justify the first-order approach in the multi-statistic case.

KEYWORDS: Principal-agent problems, asymmetric information, optimal monitoring, convexity, monotone likelihood ratio, total positivity.

1. INTRODUCTION

SINCE MIRRLEES (1975) it has been clear that the so-called first-order approach to solving principal-agent problems is generally not valid. In spite of this, it seems that the convenience of the approach has often outweighed any reservations as to it's validity. Common practice in applications of the principal-agent model is to *assume* that the approach is valid even when specific assumptions are made which (for all we know) invalidate it. Some sufficient conditions for the validity of the first-order approach are known of course, but these only apply in the most basic model. These conditions, due to Mirrlees, are that the density of output with the agents action as parameter has a monotone likelihood ratio (MLR) and, in addition, that the distribution function of output as a function of the agents effort be convex at each level of output (CDF). The first condition has a fairly natural interpretation of more effort, more output, and also serves to imply (when the first-order approach is valid) that the agents payment be increasing in observed output. Milgrom (1981) gives a further very natural interpretation. The second condition is by no means as pleasing however. To see why, consider the following example. Output is subject to a simple additive disturbance ε with distribution function F, and effort, a, is measured in output terms. So, realized output is given by $x = a + \varepsilon$, and this has distribution function $F(x - a)$ which is only convex in effort if ε has an *increasing* density! Hence, an apparently natural case does not fit the condition. Another real drawback of the assumption is that most convenient examples do not fit this scheme: most of the distributions commonly occurring in statistics (and economics) do not have the CDF property. It would seem desirable therefore to derive some different conditions to justify the first-order approach, even in the context of the most basic model. The need is more acute for extensions of the basic model for which the Mirrlees-Rogerson

[1] This is a revised version of two earlier papers: "Justifying the First-Order Approach to Principal-Agent Problems" and "Justifying the First-Order Approach to Principal-Agent Problems. Part 2: Extensions of the Basic Model." I am grateful to the referees for some helpful suggestions.

conditions do not apply. In particular, they are not applicable to cases where the principal can observe more than a single variable. The often cited paper by Holmstrom (1979) is of this type as well as those by Baiman and Demski (1980a, 1980b).

All this suggests that the first order approach should perhaps be given up. The problem then is to replace it with something else. Grossman and Hart (1983) give a method but this has the major drawback for applied theorizing that it is less convenient. Indeed, to get results one often has to impose strong conditions, and these may sometimes be sufficient for the first order approach to have been valid in the first place (examples of this occur in Grossman and Hart, Propositions 8 and 9, and in Dye (1986)). The present paper derives a suite of conditions under which the first-order approach is valid, and which to an (arguably) acceptable extent avoid the drawbacks mentioned above. In particular, the CDF assumption is replaced by something else, and conditions are derived which are valid for problems with more than one variable.

Theorems 1 and 3 are the main results of the paper. Theorem 1 is applicable to the most basic model with a single variable observed by the principal. Corollary 1 gives a simple and useful class of densities which satisfy the conditions of the theorem. Theorem 3 requires somewhat stronger assumptions but is applicable to a wider class of problems including some occurring in Holmstrom (1979), and Baiman and Demski (1980a, 1980b). In the next section we outline the basic model and state and prove Theorem 1.

2. THE BASIC MODEL

The model is the usual one. The principal who is risk neutral[2] observes a realized output x, and pays the agent an amount $s(x)$. The agent on the basis of the agreed payment schedule $s(x)$ chooses an action a (effort say), and has a separable von Neumann-Morgenstern utility $u(s(x))-c(a)$. The usual assumptions are made concerning u and c, namely that u be increasing concave and c be increasing convex. In a different line of business the agent could receive expected utility R, so a constraint on the principal's choice of s is that the agent's maximized expected utility must not be less than R. The technology which is common knowledge is represented by the distribution function of output dependent on effort, $F(x, a)$. There are both discrete and continuous formulations of the problem in the literature, so to facilitate comparison with both, we shall assume that $F(x, a)$ is absolutely continuous with respect to the same nonnegative measure for each a. Hence, F has a density f and

$$(2.1) \qquad F(x, a)=\int_{-\infty}^{x} f(y, a)\, d\sigma(y).$$

[2] This simplifies the exposition in a not entirely trivial way. With a risk averse principal, there is the added difficulty in justifying the first order approach. This arises in signing a Lagrange multiplier. See footnote 3.

The principal's problem is to

$$\underset{s,\,a}{\text{maximize}} \int [x - s(x)]\, f(x, a)\, d\sigma(x)$$

subject to

$$a \in \arg\max \int u(s(x)) f(x, a)\, d\sigma(x) - c(a) \tag{2.2}$$

and

$$\int u(s(x)) f(x, a)\, d\sigma(x) - c(a) \geqslant R. \tag{2.3}$$

Constraint (2.2) is commonly referred to as the incentive compatibility constraint: the agent must have the correct incentives to go along with the principal's idea of how much effort should be provided. Constraint (2.3) has often been called the individual rationality constraint, but standard practice is likely to follow Arrow's (1986) suggestion, which is to call it the participation constraint.

The first-order approach replaces the incentive compatibility constraint by the condition that the agent's expected utility be stationary in effort. Under the assumption that we can differentiate through the integral, this condition can be written as

$$\int u(s(x)) f_a(x, a)\, d\sigma(x) - c'(a) = 0, \tag{2.4}$$

where of course the subscript a denotes a partial derivative. Since not all stationary points are global maxima, but under the assumption (which we now make) that effort is chosen from an open interval all global maxima are stationary points, the associated first-order problem gives the principal a strictly larger choice set than the original problem. So, in general, the two problems may have quite different solutions. If, on the other hand at the solution to the associated first-order problem, the agents expected utility turns out to be concave in effort, then it will necessarily be the case that the chosen level of effort will satisfy the original incentive compatibility condition. It will therefore follow that the solution to the first-order problem will solve the original problem. A necessary condition for s and a to solve the first-order problem is that there exist numbers λ and μ such that (2.5) and (2.6) hold:

$$\frac{1}{u'(s(x))} = \lambda + \mu \frac{f_a(x, a)}{f(x, a)}, \tag{2.5}$$

$$\int [x - s(x)]\, f_a(x, a)\, d\sigma(x) + \mu \left\{ \int u(s(x)) f_{aa}(x, a)\, d\sigma(x) - c'(a) \right\} = 0. \tag{2.6}$$

The question of the sign of μ in (2.5) is of interest.[3] The following simple Lemma gives a direct argument which shows that whenever the first-order problem *has a solution*, then $\mu > 0$. If in addition the first order approach *is valid*, then this has the further implication that given the sharing rule, the principal would like to see the agent increase their effort. This last statement follows directly from (2.6), and $\mu > 0$.

LEMMA 1: *If u is an increasing concave function and $c'(a) > 0$, then any μ satisfying (2.4) and (2.5) is positive.*

PROOF: Substituting (2.5) into (2.4) gives

$$\int_{-\infty}^{\infty} u(s(x))\left(\frac{1}{u'(s(x))} - \lambda\right) f(x, a)\, d\sigma(x) = \mu c'(a). \tag{2.7}$$

Using the fact that $Ef_a/f = 0$, (2.5) gives

$$\int_{-\infty}^{\infty}\left(\frac{1}{u'(s(x))}\right) f(x, a)\, d\sigma(x) = \lambda. \tag{2.8}$$

Hence, (2.7) states that the covariance of $u(s(x))$ and $1/u'(s(x))$ is equal to $\mu c'(a)$. Since u and $1/u'$ are both monotone in the same direction they have a nonnegative covariance, and since $c'(a)$ is positive by assumption, it follows that $\mu \geqslant 0$. We can rule out the $\mu = 0$ case for then $s(x)$ would be constant and this would violate (2.4). *Q.E.D.*

To get our results, we shall want to impose conditions on the shape of $u(s(x))$ at the optimum; to this end define

$$\omega(z) = u\left(u'^{-1}(1/z)\right), \quad z > 0. \tag{2.9}$$

THEOREM 1: *The first order approach is valid if*

(2.10a) $\int_{-\infty}^{y} F(x, a)\, dx$ *is nonincreasing convex in a for each value of y;*

(2.10b) $\int x dF(x, a)\, dx$ *is nondecreasing concave in a;*

(2.11) $f_a(x, a)/f(x, a)$ *is nondecreasing concave in x for each value of a;*

(2.12) $\omega(z)$ *is concave.*

[3] This question has been addressed in a number of places. Sometimes the conclusion is drawn on the assumption that the first-order approach is valid (e.g. Holmstrom (1979)). Clearly, this assumption is not a reasonable one to make in the present paper where the whole point is to establish the validity of the approach. An argument for $\mu > 0$, due to Lambert, which does not assume the validity of the first order-approach at the outset is given in Rogerson (1985, footnote 8). This argument however uses the (redundant) MLR condition. The question is significantly more complicated when the principal is risk averse rather than risk neutral; this case is the subject of Rogerson's paper. An earlier version of this paper (Jewitt (1986)) derives some further conditions.

PROOF: Let the payment schedule $s(x)$ solve the associated first-order problem. By Lemma 1, the number μ in the necessary condition (2.5) is positive. It follows directly from this and assumption (2.11) that

(2.13) $\quad 1/u'(s(x))$ is nondecreasing concave.

Assumption (2.12) amounts to the requirement that $u(s)$ is a concave transformation of the function $1/u'(s)$. Hence, since the class of nondecreasing concave functions is closed under composition, it follows that $u(s(x))$ is nondecreasing concave in x. It remains only to establish that the transformation $\varphi \to \varphi^*$ defined by

(2.14) $$\varphi^*(a) = \int \varphi(x)\, dF(x, a)$$

is concavity preserving, for then the solution to the first order problem will lie within the choice set available to the principal in the original problem. Condition (2.10) is necessary and sufficient for the transformation (2.14) to map the set of nondecreasing concave functions into itself (Jewitt and Kanbur (1986)). *Q.E.D.*

3. DISCUSSION

Condition (2.12) is not particularly stringent; neither is it particularly natural. It does however seem to have something of an economic interpretation which we shall come to in Section 5. It can be put somewhat differently. Making the change of variable $u'(x) = 1/z$ in (2.9), we have

$$\omega(1/u'(x)) = u(x), \quad \text{with } \omega \text{ concave.}$$

Differentiating this, one obtains

$$u'(x) = \omega'(1/u'(x)) \cdot \left(-u''(x)/[u'(x)]^2\right),$$

and since $\omega'(1/u'(x))$ is nonincreasing in x, it follows that

(3.1) $\quad -u''(s)/[u'(s)]^3$ is nondecreasing.

So, the coefficient of absolute risk aversion must not decline too quickly in comparison with the marginal utility of money. The condition is satisfied by any constant absolute risk averse utility, and any nondecreasing relative risk averse utility with coefficient of relative risk aversion bounded above one half.

Condition (2.11) is equivalent to the likelihood ratio being nondecreasing and concave. That is, for fixed $a_2 > a_1$, we have

(3.2) $$\frac{f(x, a_2)}{f(x, a_1)} \quad \text{is nondecreasing concave in } x.$$

The condition suggests that variations in output at higher levels are relatively less useful in providing "information" on the agents effort than they are at lower levels of output.

The condition (2.10) is the most natural of the three and should be regarded as entirely unobjectionable in a production context.[4] To see why consider the following situation. Output is produced with the aid of effort through a production process which is subject to random shocks. That is,

$$x = g(a, \varepsilon) \tag{3.3}$$

where ε is a random disturbance. A very natural assumption to make concerning this process is that the production function g should exhibit decreasing marginal returns in each state of nature, i.e. be concave. In this case, it will be true that for any nondecreasing concave function φ,

$$E\varphi(g(a, \varepsilon)) \quad \text{is also concave.} \tag{3.4}$$

This follows immediately from the fact that the class of nondecreasing concave functions is convex and closed under composition. Setting

$$F(x, a) = \text{prob}\{g(a, \varepsilon) \leqslant x\}, \tag{3.5}$$

it follows that

$$E\varphi(g(a, \varepsilon)) = \int \varphi(x)\, dF(x, a) \quad \text{is concave;} \tag{3.6}$$

in other words, the transformation $\varphi \to \int \varphi(x)\, dF(x, a)$ defined by (3.6) is concavity preserving. Hence, the assumption of a random production function with decreasing marginal returns in each state of nature is a more restrictive assumption than (2.10). Another interpretation of (2.10) is that the agent should never be able to improve the distribution of output, in the sense of second order stochastic dominance, simply by randomizing the amount of effort supplied.

Although condition (2.10) is very natural, it will often be easier to verify an over-sufficient condition rather than the necessary and sufficient ones directly. One such condition is given in the next section, another appears in Karlin (1963). This rests on the theory of total positivity.[5] A function $f(x, a)$ is said to be Totally Positive of degree n (TP_n) if for each $x_1 < x_2 < \cdots < x_n$; $a_1 < a_2 < \cdots < a_n$, we have

$$\begin{vmatrix} f(x_1, a_1) & \cdots & f(x_1, a_k) \\ \vdots & & \vdots \\ f(x_k, a_1) & \cdots & f(x_k, a_k) \end{vmatrix} \geq 0, \quad \text{for} \quad k = 1, \ldots, n. \tag{3.7}$$

If f is TP_n for all n, then it is said to be totally positive (TP). Note in particular that if a density is TP_2, then this is equivalent to the monotone likelihood ratio condition. Karlin (1963) shows that the transformation $\varphi \to \int \varphi(x)\, dF(x, a)$ is concavity preserving provided F has a TP_3 density, with expected output concave in effort a.

[4] The monotonicity part is actually implied by (2.11).
[5] An exposition of some of the relevant parts of this theory is given in Jewitt (1987).

Any exponential family is TP in an appropriate parameterization: i.e. any density which can be written in the form

$$f(x, a) = \theta(x)\psi(a)e^{\alpha(a)\beta(x)} \tag{3.8}$$

with α and β nondecreasing is TP. These densities are those possessing sufficient statistics and are, of course, a very important class. Since for such a density

$$f_a(x, a)/f(x, a) = \alpha'(a)\beta(x) + \psi'(a)/\psi(a), \tag{3.9}$$

the concavity of the likelihood ratio is determined by the concavity of β. These observations make it relatively straightforward to verify the required conditions for many common densities. A rather useful class is furnished by the following proposition.

COROLLARY 1: *Let the density of output be of the form* (3.8) (*i.e. admitting a sufficient statistic*) *with* $\beta(x)$ *concave; then the conditions* (2.10) *and* (2.11) *of Theorem* 1 *are satisfied provided only that expected output is concave in effort.*

We give a short list of some particular examples which satisfy the conditions of Theorem 1.

a. Gamma with mean κa:

$$f(x, a) = a^{-\kappa}x^{\kappa-1}e^{-x/a}/\Gamma(\kappa),$$
$$f_a/f = -\kappa a^{-1} + xa^{-2}.$$

b. Poisson with mean a:

$$f(x, a) = a^x e^{-a}/\Gamma(1 + x),$$
$$f_a/f = -1 + x/a.$$

c. Chi-squared with degree of freedom parameter a:

$$f(x, a) = k(a)\cdot x^{2a-1}\cdot e^{-x/2}, \qquad k(a) = \Gamma(2a)^{-1}\cdot 2^{-2a},$$
$$f_a/f = k'(a)/k(a) + 2\log x.$$

4. THE MANY VARIABLE CASE

Holmstrom (1979) asks the question: how should the agent's payment depend on other observables which may also be influenced by the agents effort? We shall make an assumption which is common in the literature, that at each level of effort a, the observable variables x and y are independently distributed with joint density $f(x, a)g(y, a)$.[6]

The principal's problem is to choose a payment $s(x, y)$ for each possible realization (x, y) of the random variables, and a level of effort a for the agent to

$$\text{maximize} \int\int [x - s(x, y)]\, dF(x, a)\, dG(y, a)$$

[6] It is possible to derive some conditions without this assumption, but we do not pursue the issue here.

subject to

$$a \in \arg\max \int\int u(s(x, y))\, dF(x, a)\, dG(y, a) - c(a) \tag{4.1}$$

and

$$\int\int u(s(x, y))\, dF(x, a)\, dG(y, a) - c(a) \geqslant R, \tag{4.2}$$

where $F(x, a)$ is the distribution function of output at effort level a as before, and $G(y, a)$ is the distribution function of the "other" observable random variable. The first order approach to analyzing this problem is to replace (4.1) with the weaker condition that the agents expected utility should be at a stationary point in effort. Necessary conditions for s to solve the first order version of the problem are given in Holmstrom (1979); they include

$$\frac{1}{u'(s(x, y))} = \lambda + \mu\left(\frac{f_a(x, a)}{f(x, a)} + \frac{g_a(y, a)}{g(y, a)}\right) \tag{4.3}$$

where a is the optimal level of effort. The same argument as in Lemma 1 establishes that $\mu > 0$. Equation (4.3) implies that the expected utility of the agent can be written as

$$\int\int \omega(r(x) + t(y))\, dF(x, a)\, dG(y, a) - c(a),$$

where

$$r(x) = f_a(x, a)/f(x, a), \tag{4.4}$$

$$t(y) = g_a(y, a)/g(y, a), \tag{4.5}$$

and

$$\omega(z) = u\big(u'^{-1}(1/(\lambda + \mu z))\big). \tag{4.6}$$

Note that the definition of ω has changed somewhat from (2.9); the reason is purely for notational convenience and no confusion should result. The reason the Mirrlees-Rogerson conditions work in the basic principal agent model is as follows. If $F(x, a)$ is convex in a for each x, then for *any nondecreasing* function φ, $\int \varphi(x)\, dF(x, a)$ is concave. The Mirrlees-Rogerson condition is not sufficient in the present problem, the reason being that the concavity preservation referred to above does not extend in the most obvious way to the multidimensional case. To see why not, consider the following simple two variable case with $F = G$:

$$\int\int \exp(x + y)\, dF(x, a)\, dF(y, a) = \left[\int \exp(x)\, dF(x, a)\right]^2 \tag{4.7}$$

if F satisfies the CDF and MLR conditions, then $\int \exp(x)\, dF(x, a)$ will be nondecreasing concave in a, but there is no reason to suppose that the increasing convex transformation (i.e., y^2) of it will be concave. The Mirrlees condition can be adapted. It turns out that what is required is the supplementary assumption on utility which we used in Theorem 1.

THEOREM 2: *The first order approach is valid for the above problem if* (a) *the distribution functions* $F(x, a)$ *and* $G(y, a)$ *both satisfy the MLR and CDF conditions, and* (b) *the agent's utility satisfies condition* (2.12) *from the statement of Theorem* 1.

PROOF: In light of the discussion, it only remains to prove that the conditions of the theorem imply that expected utility of the agent is concave in effort at the optimal contract. To simplify notation set

$$K(x, y) = \omega(r(x) + t(y)). \tag{4.8}$$

The problem is that of finding conditions on F, G, and K to ensure that

$$\int\int K(x, y)\, dF(x, a)\, dG(y, a) \quad \text{is concave.} \tag{4.9}$$

The conclusion follows from repeated application of the formula for integration by parts. Suppose for the sake of the argument that for each level of effort a, F and G both have their supports in the interval $[A, B]$:

$$\begin{aligned}
&\int\left\{\int K(x, y)\, dF(x, a)\right\} dG(y, a) \\
&= \int\left\{K(B, y) - \int F(x, a) K_x(x, y)\, dx\right\} dG(y, a) \\
&= \int K(B, y)\, dG(y, a) - \int F(x, a)\left\{\int K_x(x, y)\, dG(y, a)\right\} dx \\
&= K(B, B) + \left[-\int G(y, a) K_y(B, y)\, dy\right] + \left[-\int F(x, a) K_x(x, B)\, dx\right] \\
&\quad + \left[\int\int F(x, a) G(y, a) K_{xy}(x, y)\, dx\, dy\right].
\end{aligned} \tag{4.10}$$

Condition (2.12) implies that $K_{xy} \leqslant 0$. The CDF assumption implies that $F(x, a)G(y, a)$ is convex in a. Hence, the conditions of the theorem imply that each of the terms in square brackets are concave in a. *Q.E.D.*

There remains the objection to the CDF condition, and the conditions are not immediately applicable[7] for the next problem which we shall discuss. An alternative set of conditions can be obtained by strengthening the MLR condition to the concave likelihood ratio condition of last section. This assumption together with (2.12) above implies that $\omega(r(x) + t(y))$ is concave in (x, y). The problem then reduces to determining when this concavity property is preserved in (4.9). Our earlier conditions will not[8] suffice but something very similar will.

[7] Although one can show that for certain cases they are in fact applicable at the solution. One needs to verify that the inspection policy is of a form which makes them applicable.

[8] As far as I know, I have not constructed a counterexample.

LEMMA 2: *Let X_a be a random variable with distribution function $H(x, a)$. The following statements are equivalent*:

(4.11) *$H(x, a)$ is quasiconvex in (x, a)*,

(4.12) *X_a has the same distribution as $\varphi(a, \varepsilon)$*,

for some random ε, and some function $\varphi(a, \varepsilon)$ which is nondecreasing concave in a for each realization of ε, and nondecreasing in ε for each a.

PROOF: See Jewitt and Kanbur (1988). *Q.E.D.*

Condition (4.13) has a natural production function interpretation. There are decreasing marginal returns in each state of nature and moreover, states of nature can be ranked unambiguously from bad to good. Lemma 1 constitutes a reasonably flexible tool for this sort of problem. The reason is that we can use it to appeal to simple closure properties of the class of concave functions. This allows us to dispense with the problem of dimension: the class of nondecreasing concave functions is closed under composition regardless of dimensionality, and the same for their closure under positive linear combinations, i.e. expectations. This latter property allows us to forget to some extent about the domain of integration, a feature which will be useful in the next section.

THEOREM 3: *The first order approach is valid for the above problem if (a) the distribution functions $F(x, a)$ and $G(y, a)$ are quasiconvex in (x, a) and (y, a) respectively*, (b) *they both have the monotone concave likelihood ratio property* (2.11), *and* (c) *condition* (2.12) *on the agent's utility holds.*

PROOF: Since the μ in the necessary condition (4.3) is positive, it follows from the condition (2.11) and the assumption on the agents utility that the function $K(x, y)$ defined in (4.9) is nondecreasing concave. All that remains is to show that this and condition (a) of the theorem implies that

$$\int\int K(x, y)\, dF(x, a)\, dG(y, a) \quad \text{is concave in } a.$$

By Lemma 1, we have

$$\int\int K(x, y)\, dF(x, a)\, dG(y, a) = EK(\varphi_1(a, \varepsilon_1), \varphi_2(a, \varepsilon_2))$$

for some random ε_1, ε_2 and some pair of functions φ_1 and φ_2 each nondecreasing concave in a. Since, K is nondecreasing concave

$$\psi(a, \varepsilon_1, \varepsilon_2) = K(\varphi_1(a, \varepsilon_1), \varphi_2(a, \varepsilon_2))$$

is concave for each realization of $(\varepsilon_1, \varepsilon_2)$,

and concavity of the agents expected utility in effort now follows from the fact that the set of concave functions is closed convex. *Q.E.D.*

5. THE OPTIMAL MONITORING PROBLEM

In light of the importance of condition (2.12) in the above, it would be nice to have some sort of intuition as to it's meaning. Roughly speaking, and within the confines of the first-order approach, it seems that people with such utility functions are more easily motivated by "sticks" rather than "carrots." In the present section we investigate the optimal monitoring problem due to Baiman and Demski (1980a, 1980b);[9] this can be viewed as an extension of the model of Holmstrom (1979) to the case where the principal must pay a fixed fee K in order to observe y. The principal can observe x before deciding whether to pay for an investigation. In general, the principal may choose to make investigation random conditional on the observed output.

The principal's problem can be written as: choose reward functions $s(x, y)$ and $s(x)$, and probability of investigation $p(x)$ conditional on observed output x to

$$\text{maximize} \int\int \{[x - s(x, y) - K]\, p(x) + [x - s(x)][1 - p(x)]\}\, dF(x, a)\, dG(y, a)$$

subject to

$$(5.1) \qquad a \in \arg\max \int\int \{u(s(x, y))\, p(x) + u(s(x))(1 - p(x))\}\, dF(x, a)\, dG(y, a) - c(a)$$

and

$$(5.2) \qquad \int\int \{u(s(x, y))\, p(x) + u(s(x))(1 - p(x))\}\, dF(x, a)\, dG(y, a) - c(a) \geqslant R.$$

Note firstly that since this problem is linear in $p(x)$ we can restrict attention to p functions of the form

$$p(x) = 1 \quad \text{if } x \in S,\ S \text{ a subset of } \mathbb{R},$$
$$p(x) = 0 \quad \text{if } x \in \sim S,\ \sim S \text{ the complement of } S \text{ in } \mathbb{R}.$$

The problem now becomes

$$\text{maximize} \int\int_{x \in S} \{x - s(x, y) - K\}\, dF(x, a)\, dG(y, a) + \int_{x \in \sim S} \{x - s(x)\}\, dF(x, a)$$

subject to

$$(5.3) \qquad a \in \arg\max \int\int_{x \in S} u(s(x, y))\, dF(x, a)\, dG(y, a) + \int_{x \in \sim S} u(s(x))\, dF(x, a) - c(a)$$

[9] See also Young (1986), Lambert (1986), and Dye (1986).

and

$$\iint_{x\in S} u(s(x,y))\,dF(x,a)\,dG(y,a) + \int_{x\in \sim S} u(s(x))\,dF(x,a) - c(a) \geqslant R. \tag{5.4}$$

The first order approach to this problem is to replace (5.3) with its associated first order condition

$$\frac{d}{da}\left(\iint_{x\in S} u(s(x,y))\,dF(x,a)\,dG(y,a) + \int_{x\in \sim S} u(s(x))\,dF(x,a)\right) = c'(a).$$

The necessary conditions for the principal's problem include

$$\frac{1}{u'(s(x))} = \lambda + \mu\left(\frac{f_a(x,a)}{f(x,a)}\right) \quad \text{for } x \in \sim S \tag{5.5}$$

and

$$\frac{1}{u'(s(x,y))} = \lambda + \mu\left(\frac{f_a(x,a)}{f(x,a)} + \frac{g_a(y,a)}{g(y,a)}\right) \quad \text{for } x \in S. \tag{5.6}$$

Once again, the same argument as Lemma 1 establishes that $\mu > 0$. With r, t, and ω defined as in (4.4)–(4.6), the agents expected utility can be written as

$$\iint_{x\in S} \omega(r(x)+t(y))\,dF(x,a)\,dG(y,a) + \int_{x\in \sim S} \omega(r(x))\,dF(x,a) - c(a). \tag{5.7}$$

THEOREM 4: *The first order approach is valid for the optimal monitoring problem if the conditions of Theorem* 3 *hold.*

PROOF: As for Theorem 3. *Q.E.D.*

An Application

One aspect of this problem which is of interest is to describe the optimal investigation region S. An investigation policy is said to be *lower tailed* if investigation takes place only for realized values of output below some preassigned level, and *upper tailed* if it takes place only for values above some level. Within the confines of the first-order approach, it turns out[10] that investigation will be lower tailed if

$$\int u(s(x,y))\,dG(y,a) \leqslant u(s(x)) \quad \text{for all } x, \tag{5.8}$$

[10] See Baiman and Demski (1980b), and Lambert (1985).

where the s functions are defined as in (5.5) and (5.6) but with the extension to all $x \in R$, and upper tailed if the opposite inequality holds. This suggests the following very natural interpretation, due to Baiman and Demski (1980b): that the inspection policy will be lower tailed if inspection is bad for the agent in expected utility terms and that it will be upper tailed if it is good for the agent in expected utility terms. Thus we have criteria for whether it is optimal to use a "carrot" or a "stick" to motivate the agent. Indeed, interestingly, the distinction rests on properties of the agents utility function. Note that

$$(5.9) \qquad \int u(s(x, y))\, dG(y, a) = \int \omega(r(x) + t(y))\, dG(y, a);$$

hence, if $\omega(z)$ is concave, i.e. condition (2.12) holds,

$$(5.10) \qquad \int u(s(x, y))\, dG(y, a) \leqslant \omega\left(r(x) + \int t(y)\, dG(y, a)\right),$$

and since by the definition of t, and the fact that densities integrate to 1,

$$(5.11) \qquad \int t(y)\, dG(y, a) = 0,$$

we obtain

$$(5.12) \qquad \int u(s(x, y))\, dG(y, a) \leqslant \omega(r(x)) = u(s(x)).$$

This suggests the interpretation alluded to above, that people with utilities satisfying condition (2.9), are more cheaply motivated with a stick than a carrot. In a recent paper Dye (1986) derives some results without using the first order approach. Interestingly, the results he obtains (which are a condition for the inspection policy to be lower tailed) are precisely those in our Theorem 2 above. In fact given that the optimal inspection policy to the first order problem is lower tailed, it is possible to use the same argument as in the proof of Theorem 2 to establish the validity of the first order approach. This is indicative of the value of the approach; despite the apparently restrictive conditions under which it is valid, it seems that it is often difficult to do any better.

5. CONCLUSION

The results in this paper can be extended to the case of a risk averse principal, albeit with some further assumptions (Jewitt (1986)). Different results are also available in the present case; there are essentially three steps by validating the first order approach: (a) ensure that a certain Lagrange multiplier is positive, which may require some sort of stochastic dominance assumption together with assumptions on utilities, (b) make assumptions on the likelihood ratio and utilities to ensure that the utility of the agent as a function of observable variables falls within some known class, C say, (c) make assumptions on the distribution function(s) to ensure that this class is mapped into the class of concave functions. In this paper we have chosen C to be the class of concave increasing functions, but there are many other possibilities. By making C a larger set we get less restrictive conditions on the likelihood ratio's and utilities but have to accept a more stringent restriction elsewhere. The Mirrlees-Rogerson condition is to take C to be the class of nondecreasing functions. A "better" balance may be found somewhere between the two, but the choice will have to be made on the basis of tractability as much as anything else.

Department of Economics, University of Bristol, Bristol B58 1HY, England

Manuscript received September, 1986; final revision received November, 1987.

REFERENCES

ARROW, K. J. (1986): "Agency and the Market," in *Handbook of Mathematical Economics, Vol. III*, ed. by K. J. Arrow and M. D. Intrilligator. Amsterdam: North-Holland, pp. 1184–1195.

BAIMAN, S., AND J. DEMSKI (1980a): "Variance Analysis Procedures as Motivational Devices," *Management Science*, 26, 840–848.

——— (1980b): "Economically Optimal Performance Evaluation and Control Systems," *Journal of Accounting Research*, 18, 184–220.

DYE, R. A. (1986): "Optimal Monitoring Policies in Agencies," *The Rand Journal of Economics*, 17, 339–350.

GROSSMAN, S., AND O. HART (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7–45.

HOLMSTROM, B. (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10, 74–91.

JEWITT, I. (1986): "Justifying the First Order Approach to Principal Agent Problems," University of Bristol Discussion Paper 86/177.

——— (1987): "Risk Aversion and the Choice Between Risky Prospects: The Preservation of Comparative Statics Results," *Review of Economic Studies*, 54, 73–85.

JEWITT, I., AND S. M. KANBUR (1988): "Risk Aversion Risk, and Two-Stage Gambles," in preparation.

KARLIN, S. (1963): "Total Positivity and Convexity Preserving Transformations," in *Convexity, Proceedings of the Seventh Symposium in Pure Mathematics of the American Mathematical Society*, ed. by Victor L. Klee. Providence, Rhode Island: American Mathematical Society.

LAMBERT, R. A. (1985): "Variance Investigations in Agency Settings," *Journal of Accounting Research*, 22, 633–647.

MILGROM, PAUL (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380–391.

MIRRLEES, J. A., AND PAUL MILGROM (1975): "The Theory of Moral Hazard and Unobservable Behaviour," Nuffield College, Oxford.

ROGERSON, W. (1985): "The First Order Approach to Principal-Agent Problems," *Econometrica*, 53, 1357–1368.

YOUNG, R. A. (1986): "A Note on "Economically Optimal Performance Evaluation and Control Systems: The Optimality of Two-Tailed Investigations," *Journal of Accounting Research*, 24, 231–240.