Ranking Contingent Monitoring Systems*

Marie-Cécile Fagart and Bernard Sinclair-Desgagné

December 2006

This paper seeks to provide a ranking of information systems in a setting of contingent monitoring. Control strategies that make the acquisition of additional information conditional on observing certain outcomes largely elude the existing ranking criteria. We show that this happens because contingent monitoring involves more than the classical tradeoff between risk-sharing and incentives; it also requires a balancing of incentives and *downside* risk. We then develop a refinement of the most common information system orderings that conveys this feature. This allows us to re-interpret and generalize some of the literature's key results concerning, for instance, auditing policies with independent or with correlated signals and monitoring systems where the precision of an added signal is endogenous.

KEYWORDS: Principal-agent, moral hazard, value of information, conditional monitoring, optimal audits, downside risk aversion.

---

* Fagart: LIRAES-Paris V and Université de Rouen, France; marie-cecile.fagart@univ-rouen.fr. Sinclair-Desgagné (corresponding author): HEC Montréal, CIRANO, and École polytechnique (Paris); bsd@hec.ca. Special thanks are due to Joel Demski, Louis Eeckhoudt, Steven Huddart, two anonymous reviewers and an Associate Editor for encouragements and key suggestions. We are also grateful to Anil Arya, John Conlon, Glenn Ellison, Claude Fluet, Christian Gollier, Son Ku Kim, Jacques Lawarrée, Philippe Marcoul, Anne Perrot, Jean-Pierre Ponssard, Patrick Rey, and François Salanié for valuable comments. This paper benefitted from questions and observations by seminar participants at the University of Nottingham, IDEI-Toulouse, HEC Montréal, Iowa State University, CORE, the University of Paris I (ETAPE), the 2003 Canadian Economic Theory Conference, and the 2004 European and North American Econometric Society Meetings.

1. INTRODUCTION

Business firms and public regulators are confronted on a daily basis with the comparison of the information generated by various management controls. A large number (probably the majority) of these schemes involve some form of contingent monitoring (through, for instance, internal audits, ad hoc personalized evaluations or fast tracks). Yet, there is currently no theoretical mean to rank the information systems induced by cost-equivalent contingent monitoring policies.

Early research on ranking information systems centered on developing orderings based on statistical decision theory - notably on Blackwell's theorem and its fineness criterion. In the principal-agent (hence, game-theoretic) framework, the *informativeness criterion* was next introduced by Holmström (1979). This criterion conveys the intuitive requirement that, to be valuable, an amendment must bring information about the manager's actions beyond what can already be gathered with the current information system. One shortcoming, however, is that it does not allow a comparison of the relative contribution of different amendments. Developed more recently by Kim (1995), the *MPS criterion* - which ranks information systems according to the mean-preserving spread relation between their respective likelihood ratio distributions - is so far the ordering that best deals with this issue.[1] This criterion renders the intuition that an information system is better when the firm can infer the manager's actions with greater accuracy. While it embodies the previous classifications, it allows the comparison of systems that are not nested.

[1] Other criteria have also been proposed by Jewitt (1997) and Demougin and Fluet (2000). As these authors show, such alternative rankings are mathematically equivalent to the MPS criterion.

Among the possible information system amendments, however, one type still eludes the MPS criterion: those that make the acquisition of further information conditional upon observing certain results. As Holmstrom (1979, p. 87) and others pointed out, such strategies of contingent monitoring should nonetheless remain significant in practice, because obtaining more data often involves unsunk additional expenses. This paper's objective is now to improve the comparison of information systems in this setting.

The text unfolds as follows. The next section lays out a standard principal-agent model with contingent monitoring, linear monitoring cost, and conditionally independent signals. Unlike several works on this topic that build on Baiman and Demski (1980)'s seminal article, we do not require that the agent's utility function belong to the HARA (hyperbolic absolute risk aversion) family.[2] The section ends with a reformulation of the principal-agent model as a one-person decision problem *à la* Grossman and Hart (1983). Using this, Section 3 then presents the intuition why the MPS criterion does not provide a useful ranking of information systems in this context. Basically, implementing a new contingent monitoring policy while keeping the ex ante probability of investigation constant amounts to making mean *and* variance-preserving transformations of specific lotteries. This suggests (see Menezes et al. 1980) that the information systems generated by contingent monitoring must now be sorted, not according to second-order stochastic dominance (which is the same here as the mean-preserving-spread ordering), but via *third-order* stochastic dominance. This is shown formally in Section 4, Theorem 1. Balancing

[2] A notable exception is Dye (1986), who studies imperfect audits (but also, and mostly, perfect audits) under more general utility functions, but who assumes that the agent's actions belong to a discrete set.

the pros and cons of various contingent monitoring policies thus involves choices with respect to added *local* risk, a behavior that is actually captured by the concept of *prudence* (Eeckhoudt and Schlesinger 2005). Theorem 2 of Section 5 provides an extension of Kim (1995)'s criterion which conveys this notion and so deals with mean and variance-preserving amendments of information systems. This yields a generalization (Proposition 1) of Baiman and Demski (1980)'s characterization of optimal audits. Further implications and generalizations when signals are correlated (Proposition 2), as in Lambert (1985), and where the precision of additional data is endogenous (Proposition 3), as in Kim and Suh (1992), are also derived in Section 6. Section 7 concludes the paper.

2. THE MODEL

Consider a one-period relationship between a principal and an agent. An amount of effort $a \in [0, \infty)$ is expected from the latter. This effort, however, is only imperfectly observable through some random variables, X and Y. We assume (until Section 6) that X and Y are conditionally independent, so for a given effort a the realizations x and y of the random variables obey the conditional distributions $F(x, a)$ and $G(y, a)$ respectively. These distributions have respective densities $f(x, a)$ and $g(y, a)$ that exhibit constant bounded supports Γ_X (with bounds $\underline{x}$ and $\overline{x}$) and Γ_Y, and are twice differentiable in a.

The likelihood ratios associated with X and Y will now be respectively denoted $L_X(x, a) = \dfrac{f_a(x, a)}{f(x, a)}$ and $L_Y(y, a) = \dfrac{g_a(y, a)}{g(y, a)}$.[3] A standard assumption is that these

[3] Throughout this paper the subscript $_a$ refers to the partial derivative with respect to a.

ratios satisfy the *Monotone Likelihood Ratio Property* (MLRP), that is: $L_X(x,a)$ and $L_Y(y,a)$ increase in x and y respectively, for every a. Clearly, L_X and L_Y are themselves random variables, and their respective distribution - called a *likelihood ratio distribution* - constitutes a formal representation of an accounting *information system*.[4] It is well known that all likelihood ratio distributions have the same mean $E[L_X] = 0$. The variance of, say, L_X is then given by $Var(L_X) = E[(L_X)^2]$; it is often denoted I_X and called the "Fisher information index" associated with X.[5]

The risk neutral principal routinely observes the value of X. Based on this, she may either compensate the agent immediately according to a wage schedule $w(X)$, or she may monitor the agent further at a constant cost K - thereby also gathering signal Y - and pay him according to a sharing rule $s(X,Y)$. We suppose that the principal can commit to a probability $m(x)$ of further monitoring after observing $X = x$. Her expected cost when the agent delivers effort a is therefore given by

$$EC = \int_{\Gamma_X} \int_{\Gamma_Y} \{(1 - m(x))w(x) + m(x)s(x,y)\} dF(x,a)dG(y,a) \qquad (1)$$
$$+ K \int_{\Gamma_X} m(x)dF(x,a).$$

The latter integral $M(a) = \int_{\Gamma_X} m(x)dF(x,a)$ gives the expected monitoring probability (which we also call the monitoring *intensity*) under a contingent monitoring *policy* $m(X)$.

[4] Actually, it is the density functions f and g themselves which are usually interpreted as information systems. But since there is a one-one relationship between these and their associated likelihood ratio distributions, calling the latter an information system will not create confusion.

[5] The Fisher information index is well-known to statisticians and econometricians. Since $E_X[(L_X)^2] = E_X[-\frac{\partial L_X}{\partial a}]$, this index measures the sensitivity of the likelihood ratio with respect to a (or the *informational content* of X about a). For a compelling illustration of the usefulness of this index in principal-agent analyses, see Dewatripont et al. (1999).

The agent's preferences are assumed to be additively separable in effort and wealth. The cost of effort is scaled so that its first-order derivative is equal to 1. The agent's attitude with respect to uncertain variations of his wealth exhibits risk aversion and is represented by a positive, strictly concave and three-times continuously differentiable Von Neumann-Morgenstern (VNM) utility index $u(\cdot)$. The agent's expected utility after exerting an effort a under a contract $[w, s, m]$ is then given by

$$EU = \int_{\Gamma_X} \int_{\Gamma_Y} \{(1 - m(x))u(w(x)) + m(x)u(s(x, y))\} dF(x, a) dG(y, a) - a. \tag{2}$$

The principal's problem is now to select a contingent monitoring policy $m(X)$ and wage schedules $w(X)$ and $s(X, Y)$ that implement some desired effort level a at a minimal cost, provided the agent thereby achieves his reservation utility level $\underline{U}$ and is also willing to deliver the expected effort. Formally, this amounts to minimizing (1), subject to the participation and incentive compatibility constraints given respectively by

$$EU = \int_{\Gamma_X} \int_{\Gamma_Y} \{(1 - m)u(w) + mu(s)\} dF dG - a \geq \underline{U}, \text{ and} \tag{3}$$

$$a = \arg\max_e \int_{\Gamma_X} \int_{\Gamma_Y} \{(1 - m)u(w) + mu(s)\} dF(x, e) dG(y, e) - e. \tag{4}$$

The latter constraint involves a continuum of inequalities and is thus not generally tractable. In what follows, we replace it by a friendlier one which requires that the effort level a be an interior stationary point of the agent's expected utility function, that is:

$$\int_{\Gamma_X} \int_{\Gamma_Y} [f_a(x, a)g + fg_a(y, a)]\{(1 - m(x))u(w(x)) + m(x)u(s(x, y))\} dx dy - 1 \geq 0. \tag{5}$$

We assume this "first-order approach" always yields a solution to the principal's problem.[6]

We shall now rewrite the principal's problem in terms of utility levels instead of wages, following Grossman and Hart (1983)'s approach. Let

$$\begin{aligned} u(w(x)) &= u_N(x) \text{ and } E_Y[u(s(x,Y))] = u_A(x), \\ \text{so } u(s(x,y)) &= u_A(x) + \omega(x,y) \text{ with } E_Y[\omega(x,Y)] = 0. \end{aligned}$$

The term $\omega(x,Y)$ represents therefore the contingent fair "lottery" which occurs when the observation of x triggers additional monitoring; the "prizes" of this lottery are of course given in utility units. Let also $\varphi = u^{-1}$ denote the inverse of $u(\cdot)$. Expressions (1), (3) and (5) are now equivalent to the following formulae (6), (7) and (8) respectively:

$$EC = \int_{\Gamma_X}\int_{\Gamma_Y} \{(1-m)\varphi(u_N) + m[\varphi(u_A + \omega(x,y))]\} dF dG + K\int_{\Gamma_X} m dF \quad (6)$$

$$EU = \int_{\Gamma_X} \{(1-m)u_N + m u_A\} dF - a \geq \underline{U} \quad (7)$$

$$EU_a = \int_{\Gamma_X} \{(1-m)u_N + m u_A\} dF_a + \int_{\Gamma_X}\int_{\Gamma_Y} m\omega dF dG_a - 1 \geq 0. \quad (8)$$

The principal's problem is then that of a surrogate VNM decision-maker with utility index $-\varphi(\cdot)$, who selects contingent lotteries $\omega(x,Y)$ and their respective probabilities $m(x)$ together with the "prizes" $u_N(x)$ and $u_A(x)$. This formulation is key for what follows.

[6] Sufficient assumptions to guarantee the validity of the first-order approach in a multi-signal context are the *MLRP* and *Generalized CDFC* conditions defined in Sinclair-Desgagné (1994), or the more general *Concave Increasing-Set Probability* (CISP) condition recently proposed by Conlon (2006).

3. WHAT IS CONTINGENT MONITORING?

This section will now draw attention to some characteristic features of contingent monitoring, in order to fix intuition for the upcoming formal results.

First, imagine a decision-maker who has to choose between two lotteries, L1 and L2. Both lotteries propose a pair of equiprobable monetary prizes, which are either q or "q^* plus the outcome of another fair lottery ξ" for Lottery L1, and either q^* or "q plus the outcome of ξ" for Lottery L2, where $q < q^*$. Suppose the decision-maker's preferences can be represented by a VNM utility index $V(\cdot)$. In this case, since L1 and L2 have the same mean and variance, knowing that the decision-maker is risk averse, so $V''(\cdot) \leq 0$, is not sufficient to predict which lottery will be chosen. If $V''' > 0$, however, the derivative

$$\Pi'(z) = V'(z) - EV'(z + \xi) \tag{9}$$

is negative so we have that

$$V(q^*) - E[V(q^* + \xi)] < V(q) - E[V(q + \xi)],$$

and Lottery L1 is picked. The reverse occurs of course if $V''' < 0$. After Kimball (1990), a VNM decision-maker whose utility $V(\cdot)$ has a positive third derivative is called *prudent*.[7]

A prudent individual prefers to be facing an additional gamble ξ when the better outcome has occurred (Eeckhoudt and Schlesinger 2005). This behavior is often seen as a form of decreasing risk aversion: the function $\Pi(\cdot)$ defined in (9) is indeed a measure, in

[7] Kimball (1990) builds on the literature on precautionary savings. Consider a risk averse household facing an uninsurable future risk ξ, where $E(\xi) = 0$. If it decides to save now an amount s of its daily income θ, it gets $\theta - s$ today but $\theta + s + \xi$ tomorrow. Assume that the household's utility function in each period is $V(\cdot)$, with $V' > 0$ and $V'' < 0$, so its objective is to maximize $V(\theta - s) + EV(\theta + s + \xi)$ with respect to s. It can be verified that positive savings $s > 0$ occur if the marginal utility of income is convex, that is: $V'''(\cdot) > 0$.

utility units, of the decision-maker's risk premium, and $V''' > 0$ means that this premium decreases with respect to utility. As far as the more usual coefficient of absolute risk aversion $R_V(z) = -\frac{V''(z)}{V'(z)}$ is concerned, however, we have that

$$R_V'(z) = R_V(z)[R_V(z) - P_V(z)],$$

where $P_V(z) = -\frac{V'''(z)}{V''(z)}$ is the coefficient of absolute prudence defined by Kimball (1990). Having $V''' > 0$ is thus only a necessary condition to observe $R_V'(z) < 0$. The latter is equivalent to saying that $P_V(z) > R_V(z)$.

Let us now return to the principal's problem, as it was formulated at the end of Section 2. To simplify, let signal X take only two values x and x'. Suppose the VNM decision maker with utility index $-\varphi(\cdot)$ is then contemplating the following lottery with prizes $u(x) < u(x')$ and contingent gamble $\omega(x, Y)$:

Lottery Ł1: Receive $u(x) + \omega(x, Y)$ with probability $f(x, a)$

or $u(x')$ with probability $f(x', a)$.

An alternative to this lottery would be the following one.

Lottery Ł2: Receive either $u(x) + \omega(x, Y)$ with probability $m(x)f(x, a)$

or $u(x)$ with probability $(1 - m(x))f(x, a)$

or $u(x') + \omega(x, Y)$ with probability $m(x')f(x', a)$

or $u(x')$ with probability $(1 - m(x'))f(x', a)$.

As long as

$$m(x')f(x', a) + m(x)f(x, a) = f(x, a), \tag{10}$$

Lottery Ł2 is transferring part of the risk attached with the lower prize $u(x)$ to the larger one $u(x')$ while keeping the same mean and variance as Ł1. It is therefore also feasible and it has the same cost. To minimize (6), the decision maker favors Ł2 if

$$f(x,a)\{m(x)E_Y[\varphi(u(x)+\omega(x,Y)]+(1-m(x))\varphi(u(x)\}$$
$$+f(x',a)\{m(x')E_Y[\varphi(u(x')+\omega(x,Y)]+(1-m(x'))\varphi(u(x')\}<$$
$$f(x,a)E_Y[\varphi(u(x)+\omega(x,Y)]+f(x',a)\varphi(u(x')).$$

By equation (10), this inequality reduces to

$$E_Y[\varphi(u(x')+\omega(x,Y)]-\varphi(u(x'))<E_Y[\varphi(u(x)+\omega(x,Y)]-\varphi(u(x)),$$

which is verified when $\varphi''' < 0$, i.e. when the VNM utility index $-\varphi(\cdot)$ is precisely that of a prudent decision-maker.[8]

This example highlights that contingent monitoring fundamentally involves choices between mean and variance-preserving adjustments in the distribution of the agent's utility levels. The principal's decision is now driven by the sign of $\varphi'''(\cdot)$. But φ''' is negative (positive) when $3R_u - P_u < (>)\ 0$, where $R_u = -\frac{u''}{u'}$ and $P_u = -\frac{u'''}{u''}$ are the agent's indices of absolute risk aversion and absolute prudence respectively. The difference $P_u - 3R_u$ is in fact an indicator of the agent's *downside* risk aversion.[9] If it is positive (negative), the

[8]This conclusion is valid as well when the gamble $\omega(x,Y)$ in Lottery Ł1 is attached to $u(x)$ and to $u(x')$ with probabilities strictly between 0 and 1. The supporting argument is analogous but slightly more complicated. It can be obtained from the authors upon request.

[9]The product $R_u(P_u - 3R_u)$ is actually a lower bound for the *index of downside risk aversion* $S = R_u(P_u - \frac{3}{2}R_u)$ introduced by Keenan and Snow (2002).

agent's prudence is high (low) enough so the principal economize in fostering action a by making the agent bear more risk when he expects higher (lower) payoffs.

This discussion completes and clarifies some important comments that can be found in the literature. The typical rationale for the principal's choice of a particular contingent monitoring scheme runs in terms of the classical tradeoff between risk sharing and incentives. To illustrate their results, people often use the constant relative risk aversion (CRRA) utility function $v(z) = \frac{z^{1-\alpha}}{1-\alpha}$ $(0 < \alpha < 1)$, observing that attaching extra monitoring with the agent's higher (lower) utility is optimal when his coefficient of relative risk aversion α is smaller (larger) than $\frac{1}{2}$. Since $\alpha < \frac{1}{2}$ if and only if $\varphi''' < 0$, the latter sentence is truly equivalent to saying that the decision maker in the principal-agent problem written according to the agent's utility levels is prudent.

These remarks will now steer our analysis of information systems.

4. SORTING INFORMATION SYSTEMS

To distinguish among contingent monitoring policies, several features come to mind. First, a policy exhibits an overall intensity M, which determines its total cost $K \cdot M$. Second, the function $m(\cdot)$ can take various forms. It is totally random if $m(x)$ is equal to some constant fraction, lower-tailed when its value decreases with x, upper-tailed when it increases as x gets larger, and two-tailed if $m(x)$ is smaller than $m(x')$ and $m(x'')$ where $\underline{x} \leq x' \leq x^* < x < x^{**} \leq x'' \leq \overline{x}$ for some bounds x^* and x^{**}. The question we now ask is how these features translate into specific attributes of the induced information systems.

Clearly, the likelihood ratio $L_m(X, Y)$ associated with a policy m is given by

$$L_m(X,Y) = \begin{cases} L_X(X,a) & \text{and there is no further monitoring, or} \\ L_X(X,a) + L_Y(Y,a) & \text{and additional monitoring occurs .} \end{cases}$$

The first and second moments of its distribution are then respectively

$$E(L_m) = 0 \;\text{ and }\; Var(L_m) = I_X + MI_Y \; . \tag{11}$$

An important implication of (11) is that two contingent monitoring policies - be they upper-tailed, lower-tailed, two-tailed, or totally random - that share the same intensity M (so the same cost) will generate likelihood ratio distributions having the same variance. Hence, it is not possible for one of these distributions to yield the other via some mean-preserving spread of probability mass. This entails, in particular, that Kim (1995)'s MPS criterion cannot distinguish between the information systems generated by contingent monitoring schemes which are equally costly.

The discussion of the previous section suggests, however, that amending a given contingent monitoring policy without affecting its cost amounts to making mean *and* variance-preserving alterations of the associated information system. The following theorem, which is proved in the Appendix, makes this intuition rigorous. Before stating this result, recall that a given probability distribution has *more downside risk* than another if it can be obtained from the latter by a sequence of probability mass transfers which keep the mean and variance unchanged while shifting dispersion from the right to the left. Let us also call a policy m_A *relatively more upper-tailed* than another policy m_B when the latter displays larger cumulated monitoring frequencies (or larger *downside intensities*) than the former, that is when $\int_{\underline{x}}^{x} m_B dF \geq \int_{\underline{x}}^{x} m_A dF$ for all x.[10]

THEOREM 1. *Let m_A and m_B be two contingent monitoring policies having the same intensity. If m_A is relatively more upper-tailed than m_B, then the distribution of L_{m_B} has more downside risk than the distribution of L_{m_A}.*

This statement clarifies what amending a contingent monitoring policy actually does to the corresponding information system. The conditional expectation and variance of the likelihood ratio distribution associated with policy $m(X)$ are respectively

$$E(L_m \mid X = x) = L_X(x, a) \quad \text{so} \quad Var(L_m \mid X = x) = m(x) I_Y(a) \ . \tag{12}$$

Raising the probability $m(x)$ of further monitoring amounts therefore to increasing the local variance of L_m at $X = x$. Conversely, decreasing $m(x')$ contracts the distribution of L_m at $X = x'$. A combination of both transformations thereby involves a reallocation of *local* variance - leftward if $x < x'$ or rightward if $x' > x$ - within the information system.

According to Menezes et al. (1980, proposition 3), the above theorem is equivalent to saying that, when a contingent monitoring policy m_A is relatively more upper-tailed than a policy m_B, the likelihood ratio distribution associated with m_A dominates the one corresponding to m_B in the sense of *third-order stochastic dominance*. Hereafter, we shall denote this as $L_{m_A} \succsim_3 L_{m_B}$.

Consider, finally, the following same-intensity, bang-bang, upper-tailed and lower-tailed policies $m_{UT}(X)$ and $m_{LT}(X)$ defined as

[10] Clearly, not all contingent monitoring policies can be compared using the "relatively more upper-tailed" criterion. This ranking is therefore an incomplete one (like all the other orderings currently offered in the literature).

$$m_{UT}(x) = 1 \text{ if } x \geq x^{**} \text{ and } m_{UT}(x) = 0 \text{ if } x < x^{**}$$

$$m_{LT}(x) = 0 \text{ if } x > x^{*} \text{ and } m_{LT}(x) = 1 \text{ if } x \leq x^{*}$$

for some thresholds x^* and x^{**}. Any policy $m(X)$ that shares the same monitoring frequency will satisfy (see the Appendix)

$$\int_{\underline{x}}^{x} m_{LT} dF \geq \int_{\underline{x}}^{x} m dF \geq \int_{\underline{x}}^{x} m_{UT} dF \tag{13}$$

for all x, so $L_{m_{UT}} \gtrsim_3 L_m \gtrsim_3 L_{m_{LT}}$ by Theorem 1. Relative to the third-order stochastic dominance ordering, any set of information systems generated by cost-equivalent contingent monitoring policies is thus bounded above and below by the systems corresponding respectively to bang-bang upper and lower-tailed contingent monitoring policies.

Now that we can classify the information systems induced by contingent monitoring, the upcoming section will turn to decision-making and the principal's choice of policy.

5. CHOOSING AN INFORMATION SYSTEM

At an optimum, it is intuitive and can easily be shown that the intensity of contingent monitoring should decrease with respect to the unit cost K.[11] To discriminate among *same-cost* information systems, we shall now establish a strict extension of Kim (1995)'s ranking and apply it to the classification developed in the preceding section. Beforehand, let us introduce the function $\Delta(u, \sigma) = u\sigma - \varphi(u)$, and let $\Delta^*(\sigma) = Max_{u \in W}\{\Delta(u, \sigma)\}$. This allows to state the following general criterion.

[11] This assertion somewhat generalizes Holmström (1979)'s "sufficient statistic" result: it says that any informative signal about the agent's effort has positive value for the principal, *even* when gathering such a signal is an endogenous (i.e., strategic) decision.

Theorem 2. *The principal favors performance measure Z over performance measure $\check{Z}$ to implement action a if $E_Z[\Delta^*(\lambda_Z+\mu_Z L_Z)] \geq E_{\check{Z}}[\Delta^*(\lambda_Z+\mu_Z L_{\check{Z}})]$, where λ_Z and μ_Z are the multipliers of the participation and incentive constraints in the principal's problem with performance measure Z.*

Since $\Delta(u,\sigma)$ is a linear function of σ, $\Delta^*(\cdot)$ is convex. It follows that *Theorem 2 encompasses the MPS criterion*: if the distribution of L_Z is a mean-preserving spread of that of $L_{\check{Z}}$, then $E_Z[\Delta^*(\lambda_Z + \mu_Z L_Z)] \geq E_{\check{Z}}[\Delta^*(\lambda_Z + \mu_Z L_{\check{Z}})]$ and the principal prefers indeed signal Z to signal $\check{Z}$.

By the envelope theorem, moreover, $\frac{d}{d\sigma}\Delta^*(\sigma) = u(\sigma)$ where $u(\sigma)$ satisfies $\sigma = \varphi'(u(\sigma))$. Hence, $\frac{d}{d\sigma}\Delta^*(\sigma) = \varphi'^{-1}(\sigma)$ and

$$\frac{d^3}{d\sigma^3}\Delta^*(\sigma) > 0 \;\text{ if and only if }\; \varphi''' < 0. \tag{14}$$

Recall now that, for all function $V(\cdot)$ with $V''' > 0$, $E_H V(z) \geq E_{\hat{H}} V(z)$ whenever distribution $\hat{H}$ has more downside risk than distribution H (see, e.g., Menezes et al. 1980, theorem 2). The next statement is then a direct corollary of (14) and the above theorems.

Proposition 1. *Let m_A and m_B be two contingent monitoring policies having the same intensity, and suppose that m_A is relatively more upper-tailed than m_B. If $\varphi''' <$ $(>)$ 0, then the principal prefers policy m_A (m_B).*

This result corroborates the intuitive developments of Section 2. A *generalization of Baiman and Demski (1980)'s first characterization of optimal contingent monitoring* is also at hand: bearing in mind the inequalities in (13), the optimal contingent monitoring policy is bang-bang and upper-tailed (lower-tailed) if $\varphi''' < (>) \ 0$.[12]

The situation where $\varphi''' = 0$ is obviously one of neutrality towards downside risk. In this case, the principal cares only about the cost $K \cdot M$, and it can be checked that an optimal policy to implement action a could actually take any form (lower-tailed, upper-tailed, two-tailed, random, etc.). Note that $\varphi''' = 0$ is equivalent to having $P_u = 3R_u$, which means that the agent's preferences are captured by the utility function $u(z) = \sqrt{z}$.[13]

The methods underlying Theorems 1 and 2 can provide other extensions of key contingent monitoring results. Two important ones are pursued in the next section.

6. EXTENSIONS

Our analysis presupposed so far that the performance measures X and Y were conditionally independent, and that the precision of the additional signal Y was exogenous. This section will now lift these assumptions.[14]

6.1 *Correlated Performance Measures*

Suppose hereafter that the pair of random variables (X, Y) has a joint density function $h(x, y, a)$. In this context, $f(x, a) = \displaystyle\int_{\Gamma_Y} h(x, y, a)dy$ denotes the marginal density of X,

[12] This sufficient (non-parametric) condition for the optimality of lower-tailed contingent monitoring was already found by Dye (1986), in a discrete setup. Our more general derivation also brings an information-value perspective, refines the standard tradeoff between risk sharing and incentives, and conveys further intuition on the sufficient condition (i.e. $\varphi''' > 0$) for the optimality of lower-tailed contingent monitoring.

[13] This square-root utility function is frequently assumed in the literature, mainly for tractability reasons. The present remark suggests, however, that such a modelling choice might not be innocuous.

[14] Instead of the unit cost K, one could also assume that the cost of additional monitoring is strictly convex in the probabilities $m(x)$. It can be shown in this case that *two-tailed* contingent monitoring might be optimal (as in Lambert 1985 and Young 1986, but for different reasons), and that the optimal policy may not be bang-bang. A proof can be obtained from the authors upon request.

and $g(x,y,a) = \dfrac{h(x,y,a)}{f(x,a)}$ stands for the density of Y conditional on observing $X = x$.[15]

The likelihood ratio associated with a monitoring policy $m(X)$ is now

$$\begin{aligned} L_m &= \frac{f_a(x,a)}{f(x,a)} \quad \text{when no additional monitoring occurs, or} \qquad (15) \\ &= \frac{h_a(x,y,a)}{h(x,y,a)} = \frac{f_a(x,a)}{f(x,a)} + \frac{g_a(x,y,a)}{g(x,y,a)} \quad \text{when further monitoring happens.} \end{aligned}$$

The conditional mean and variance of the corresponding likelihood ratio distribution are then respectively

$$E(L_m \mid X = x) = L_X(x,a) \quad \text{and} \quad Var(L_m \mid X = x) = m(x) E_{Y/X=x}[(L_Y)^2] \ . \qquad (16)$$

Comparing (16) and (12) reveals that local variance depends here, not only on the probability $m(x)$, but also on the informational content of signal Y *when* $X = x$. If that content increases (decreases) in the value X takes, then the principal's preference for relatively upper-tailed (lower-tailed) contingent monitoring is in fact reinforced. This intuition is confirmed in the next proposition, which is proved in the Appendix.

Proposition 2: *Let m_A and m_B be two contingent monitoring policies having the same intensity, and suppose that m_A is relatively more upper-tailed than m_B. If $\varphi''' <$ $(>)$ 0 and the likelihood ratio distribution of Y incurs a mean-preserving spread as x increases (decreases), then the principal prefers policy m_A (m_B).*

The intuitive argument suggests there may exist a trade off between informativeness

[15] In a previous article, Rajan and Sarath (1997) considered optimal correlation levels in information systems where both selected signals X and Y are afterwards always observed (i.e., $m(X) \equiv 1$). Here, we examine the converse situation where correlation is already given but the principal has discretion over when to gather additional information. Note also that our framework allows signals to take more than two values.

and downside risk. Were these two factors to push instead in opposite directions (let $\varphi''' > 0$ while the informational content of Y grows with X, for instance), one can expect the principal to sometimes prefer two-tailed policies. Clearly, such policies will dominate if (as in Lambert 1985's example) downside risk is irrelevant (i.e., $\varphi''' = 0$) and the information conveyed by Y is more valuable at smaller and larger realizations of X.

6.2 Endogenous Precision

Consider a variant of contingent monitoring in which the principal may adjust the level of monitoring investment after observing X. In this case, $m(X) \equiv 1$ and signal Y is always observed, but its precision $b(x)$ is endogenous. The principal's expected cost of a policy $b(X)$ is now given by $K(b) = \int_{\Gamma_X} b(x)dF$, while the density function, distribution function and likelihood ratio of Y become respectively $g(y, b, a)$, $G(y, b, a)$ and $L_Y(Y, b, a)$. For two policies $b_A(X)$ and $b_B(X)$, we have that $L_Y(Y, b_A, a)$ is a mean-preserving spread of $L_Y(Y, b_B, a)$ when $b_A(x) > b_B(x)$.

With these modifications, the principal's cost function (6) and constraints (7) and (8) turn out to be respectively

$$EC = \int_{\Gamma_X}\int_{\Gamma_Y} \varphi(u_A(x) + \omega(x, y))dF(x, a)dG(y, b, a) + \int_{\Gamma_X} b(x)dF(x, a) \quad (17)$$

$$EU = \int_{\Gamma_X} u_A(x)dF(x, a) - a \geq \underline{U} \quad (18)$$

$$\int_{\Gamma_X} u_A(x)dF_a(x, a) + \int_{\Gamma_X}\int_{\Gamma_Y} \omega(x, y)dF(x, a)dG_a(y, b, a)] - 1 \geq 0. \quad (19)$$

Considering that a can take only two values a_H and a_L, Kim and Suh (1992, theorem 1) established that the optimal policy $b^*(X)$ in this case would be decreasing in the realizations of X, provided the agent's utility function exhibits nondecreasing absolute risk aversion (so $P_u - R_u \leq 0$) or a constant coefficient of relative risk aversion $-\frac{zu''}{u'}$ between $\frac{1}{2}$ and 1. This situation is now covered under condition $\varphi''' \geq 0$ of the next proposition. A formal proof can be found in the Appendix.

PROPOSITION 3: *Consider a set of feasible same-cost solutions to the principal-agent problem. If* $\varphi''' < (\geq)\ 0$, *then the principal prefers those solutions* $b(X)$ *which increase (decrease) with respect to* X.

As before, the principal trades off incentives and downside risk. This explains why the agent's prudence (via the function φ) plays again a key role.

7. CONCLUDING REMARKS

To obtain a ranking of contingent monitoring systems, this paper first brought together two important streams of literature in principal-agent theory: one that started with Holmstrom (1979) on ordering information systems, and the other that began with Baiman and Demski (1980) on imperfect audits. It now adds to the former by providing with Theorem 2 a generalization of Kim (1995)'s MPS criterion. It also offers new insights for the analysis of contingent monitoring, by stressing the allocation of downside risk and the key role of the agent's *prudence* in this case, and by reinterpreting and extending some "classical" results on conditional monitoring with endogenous precision (proposition 3) and on auditing policies with independent (proposition 1) or correlated signals (proposition 2).

Contingent monitoring is certainly not the only area where the allocation of downside risk matters. In their study of the shape of compensation contracts, for example, Hemmer et al. (2000) argue that convex rewards (like those linked to stock options) are appropriate when contracts are based on a signal (like some stock prices) that conveys significant downside risk. Applying the notions and tools we just introduced, such as mean and variance-preserving transformations and prudence, might now bring further insights to this literature.

APPENDIX

Proof of Theorem 1 : Take a contingent monitoring policy $m(X)$ of intensity M and associated likelihood ratio L_m. For a VNM decision maker with utility index $V(\cdot)$, we have

$$E[V(L)] = \int_{\underline{x}}^{\overline{x}} \{(1-m)V(L_X) + mE_Y[V(L_X + L_Y)]\}dF.$$

Assume that $V'''(\cdot) > 0$. According to Menezes et al. (1980, theorem 2), the distribution of L_{m_B} has more downside risk than the distribution of L_{m_A} if and only if

$$E[V(L_{m_A})] - E[V(L_{m_B})] = \int_{\underline{x}}^{\overline{x}} (m_A - m_B)\{E_Y[V(L_X + L_Y)] - V(L_X)\}dF \geqq 0 \ .$$

Now, consider the two functions Ω and Σ defined as

$$\Omega(x) = \int_{\underline{x}}^{x} (m_A - m_B)dF \text{ and } \Sigma(x) = E_Y[V(L_X + L_Y)] - V(L_X). \tag{20}$$

Clearly, $\Omega(\underline{x}) = \Omega(\overline{x}) = 0$. By assumption, $\Omega(x)$ is also non positive for all $x \in (\underline{x}, \overline{x})$, and $\Sigma(x)$ increases with x. Integrating by parts then gives

$$E[V(L_A)] - E[V(L_B)] = \int_{\underline{x}}^{\overline{x}} \Omega'(x)\Sigma(x)dx = -\int_{\underline{x}}^{\overline{x}} \Omega(x)\Sigma'(x)dx \geqq 0. \quad \blacksquare \tag{21}$$

Proof of the inequalities in (13): First notice that, for all $x \leq x^*$, $m(x) \leq m_{LT}(x) = 1$ so $\int_{\underline{x}}^{x} mdF \leq \int_{\underline{x}}^{x} m_{LT}dF$. For all $x > x^*$, moreover, $\int_{\underline{x}}^{x} mdF \leq M = \int_{\underline{x}}^{x} m_{LT}dF$. This proves the first inequality in (13).

Similarly, for all $x \geq x^{**}$, $m(x) \leq m_{UT}(x) = 1$ so $\int_{x}^{\overline{x}} mdF \leq \int_{x}^{\overline{x}} m_{UT}dF$. And for all $x < x^{**}$, $\int_{x}^{\overline{x}} mdF \leq M = \int_{x}^{\overline{x}} m_{UT}dF$. Since $\int_{\underline{x}}^{\overline{x}} mdF = \int_{\underline{x}}^{\overline{x}} m_{UT}dF = M$, the second inequality in (13) must hold. $\blacksquare$

Proof of Theorem 2: Let Γ_i , $H_i(t)$, $h_i(t)$, and L_i denote the support, distribution function, density function, and likelihood ratio associated with performance measure $i = Z, \check{Z}$ when an action a is implemented. The corresponding objective function, participation constraint, and incentive compatibility constraint of the principal-agent problem are now respectively written

$$\overline{EC}_i \equiv \int_{\Gamma_i} \varphi(u(t))h_i(t)dt \tag{22}$$

$$\int_{\Gamma_i} u(t)h_i(t)dt \geq a + \underline{U} \tag{23}$$

$$\int_{\Gamma_i} u(t)L_i(t)h_i(t)dt \geq 1. \tag{24}$$

Define as $\Lambda_i(u, \lambda, \mu)$ the Lagrangian function associated with this problem, that is:

$$\Lambda_i(u, \lambda, \mu) = -\int_{\Gamma_i} \varphi(u(t))h_i(t)dt + \lambda\{\int_{\Gamma_i} u(t)h_i(t)dt - a - \underline{U}\} + \mu\{\int_{\Gamma_i} u(t)L_i(t)h_i(t)dt - 1\}.$$

From the definition of the function Δ, we have

$$\Lambda_i(u, \lambda, \mu) = E_i[\Delta(u, \lambda + \mu L_i)] - \lambda(a + \underline{U}) - \mu. \tag{25}$$

From the necessary optimality conditions, we know that constraints (23) and (24) are binding, so there exist positive multipliers λ_i and μ_i such that the optimal schedule $u_i^*(\cdot)$ maximizes $\Delta(u, \lambda_i + \mu_i L_i)$. As a consequence, we have at the optimum that

$$-\overline{EC}_i^* = \Lambda_i(u_i^*, \lambda_i, \mu_i) = E_i[\Delta^*(\lambda_i + \mu_i L_i)] - \lambda_i(a + \underline{U}) - \mu_i. \tag{26}$$

Moreover, the following inequality holds whatever λ and μ:

$$\Lambda_i(u_i^*, \lambda_i, \mu_i) = \Lambda_i(u_i^*, \lambda, \mu) \leq Max_u \Lambda_i(u, \lambda, \mu) = E_i[\Delta^*(\lambda + \mu L_i)] - \lambda(a + \underline{U}) - \mu. \tag{27}$$

Using (26) and (27) now allows to compare $\overline{EC}_Z^*$ to $\overline{EC}_{\check{Z}}^*$, so

$$\begin{aligned}
\overline{EC}_{\check{Z}}^* - \overline{EC}_Z^* &= \Lambda_Z(u_Z^*, \lambda_Z, \mu_Z) - \Lambda_{\check{Z}}(u_{\check{Z}}^*, \lambda_{\check{Z}}, \mu_{\check{Z}}) \\
&= \Lambda_Z(u_Z^*, \lambda_Z, \mu_Z) - \Lambda_{\check{Z}}(u_{\check{Z}}^*, \lambda_Z, \mu_Z) \\
&\geqslant E_Z[\Delta^*(\lambda_Z + \mu_Z L_Z)] - E_{\check{Z}}[\Delta^*(\lambda_Z + \mu_Z L_{\check{Z}})].
\end{aligned}$$

To be sure, the principal prefers the information system generated by performance measure Z to the one generated by $\check{Z}$ if using the former is cheaper, i.e. if $\overline{EC}_{\check{Z}}^* - \overline{EC}_Z^* \geq 0$. Hence, she selects signal Z over signal $\check{Z}$ to implement an action a whenever $E_Z[\Delta^*(\lambda_Z + \mu_Z L_Z)] \geq E_{\check{Z}}[\Delta^*(\lambda_Z + \mu_Z L_{\check{Z}})]$, as claimed. ■

Proof of Proposition 2 : First, as in the proof of Theorem 1, we have to show that the function $\Sigma(x)$ defined in expression (20) increases with x under the present assumptions. Denote $\hat{G}(\cdot)$ the cumulative distribution of the likelihood ratio L_Y, and $\hat{g}(\cdot)$ the corresponding density. $\Sigma(\cdot)$ is now given by

$$\Sigma(x) = \int_{\underline{\ell}}^{\overline{\ell}} V(L_X + \ell)\hat{g}(\ell, x)d\ell - V(L_X)$$

Assume that $\hat{g}(\cdot)$ is differentiable with respect to x. The first derivative of $\Sigma(\cdot)$ is then

$$\Sigma'(x) = \int_{\underline{\ell}}^{\overline{\ell}} V(L_X + \ell)\frac{\partial \hat{g}(\ell, x)}{\partial x}d\ell + \frac{\partial L_X}{\partial x}\{\int_{\underline{\ell}}^{\overline{\ell}} V'(L_X + \ell)\hat{g}(\ell, x)d\ell - V'(L_X)\}. \quad (28)$$

The first term of the latter expression can be integrated by parts two times. This gives

$$\begin{aligned} \Sigma'(x) &= \int_{\underline{\ell}}^{\overline{\ell}} V''(L_X + \ell)\{\int_{\underline{\ell}}^{\ell} \frac{\partial \hat{G}(z, x)}{\partial x}dz\}d\ell \quad (29) \\ &+ \frac{\partial L_X}{\partial x}\{\int_{\underline{\ell}}^{\overline{\ell}} V'(L_X + \ell)\hat{g}(\ell, x)d\ell - V'(L_X)\}. \end{aligned}$$

Recall that the integral $\int_{\underline{\ell}}^{\ell} \frac{\partial \hat{G}(z, x)}{\partial x}dz$ is positive (negative) if L_Y spreads (contracts) as x increases. Hence, $\Sigma'(x) > (<)\ 0$ if $V(\cdot)$ is convex, $V''' \geqslant (\leq)\ 0$ and L_Y becomes more (less) variable as x gets larger.

The result then comes from verifying inequality (21) and applying Theorem 2. ■

Proof of proposition 3 : As before, let $\hat{G}(\cdot, b)$ represent the cumulative distribution of likelihood ratio L_Y, and $\hat{g}(\cdot, b)$ be the corresponding density. Note that the function $\int_{\underline{z}}^{z} \hat{G}(r, b)dr$ is strictly increasing and concave in b.

Take now a convex three-times differentiable VNM utility index $V(\cdot)$. According to

Theorem 2, we have to show that a policy $b^*(X)$ which maximizes the expectation

$$E[E[V(L) \mid b(X)]] = \int_{\underline{x}}^{\overline{x}} \int_{\underline{\ell}}^{\overline{\ell}} V(L_X + \ell)\hat{g}(\ell, b(x))f(x,a)d\ell dx$$

$$\text{subject to} \int_{\underline{x}}^{\overline{x}} b(x)f(x,a)dx = \overline{K}$$

has the stated property. With η standing for the Lagrange multiplier of the constraint, the necessary first-order condition for a solution to this problem is given by

$$\int_{\underline{\ell}}^{\overline{\ell}} V(L_X + l)\frac{\partial \hat{g}(l,b)}{\partial b}dl = \eta \ . \tag{30}$$

After integrating the right-hand side of (30) by parts twice, we then get

$$\int_{\underline{\ell}}^{\overline{\ell}} V''(L_X + \ell)\{\int_{\underline{\ell}}^{\ell} \frac{\partial \hat{G}(r,b)}{\partial b}dr\}d\ell = \eta \ . \tag{31}$$

The following equation now obtains from taking the derivative with respect to x on both sides of (31):

$$\frac{\partial L_X}{\partial x}\int_{\underline{\ell}}^{\overline{\ell}} V'''(L_X + \ell)\{\int_{\underline{\ell}}^{\ell} \frac{\partial \hat{G}(r,b)}{\partial b}dr\}d\ell +$$
$$b'(x)\int_{\underline{\ell}}^{\overline{\ell}} V''(L_X + \ell)\{\int_{\underline{\ell}}^{\ell} \frac{\partial^2 \hat{G}(r,b)}{\partial b^2}dr\}d\ell = 0.$$

Clearly, this condition can only be satisfied if $b'(\cdot)$ and $V'''(\cdot)$ have the same sign. ■

References

Baiman, S., and J. Demski (1980): "Economically Optimal Performance Evaluation and Control Systems," *Journal of Accounting Research*, 18, 184-220.

Conlon, J. R. (2006): "Two New Conditions Supporting the First-Order Approach to Multi-Signal Principal-Agent Problems," Working paper, Department of Economics, University of Mississippi.

Demougin, D., and C. Fluet (2000): "Ranking of Information Systems in Agency Models: An Integral Condition," *Economic Theory*, 17, 489-496.

Dewatripont, M., I. Jewitt, and J. Tirole (1999): "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*, 66, 183-198.

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

Eeckhoudt, L. and H. Schlesinger (2005): "Putting Risk in its Proper Place," CESifo working paper no. 1462.

Grossman, S„ and O. Hart (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7-45.

Hemmer, T., O. Kim, and R. E. Verecchia (2000): "Introducing Convexity into Optimal Compensation Contracts," *Journal of Accounting and Economics*, 28, 307-327.

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

Jewitt, I. (1997): "Information and Principal-Agent Problems," Mimeo, University of Bristol.

Keenan, D. C., and A. Snow (2002): "Greater Downside Risk Aversion," *Journal of Risk and Uncertainty*, 24, 267-277.

Kim, S. K. (1995): "Efficiency of an Information System in an Agency Model," *Econometrica*, 63, 89-102.

Kim, S. K., and Y. S. Suh (1992): "Conditional Monitoring under Moral Hazard," *Management Science*, 38, 1106-1120.

Kimball, M. S. (1990): "Precautionary Saving in the Small and in the Large," *Econometrica*, 58, 53-73.

Lambert, R. A. (1985): "Variance Investigation in Agency Settings," *Journal of Accounting Research*, 23, 633-647.

Menezes, C., C. Geiss, and J. Tressler (1980): "Increasing Downside Risk," *American Economic Review*, 70, 921-932.

Rajan, M. V., and B. Sarath (1997): "The Value of Correlated Signals in Agencies," *The Rand Journal of Economics*, 28, 150-167.

Sinclair-Desgagné B. (1994): "The First-Order Approach to Multi-Signal Principal-Agent Problems," *Econometrica*, 62, 459-465.

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.