Optimal Incentive Contract with Costly and Flexible Monitoring

Anqi Li\textsuperscript{1} \quad Ming Yang\textsuperscript{2}

\textsuperscript{1}Department of Economics, Washington University in St. Louis

\textsuperscript{2}Fuqua School of Business, Duke University

May 2016
Motivation

Firm’s choice of monitoring technology has a significant impact on employee productivity.

With the recent advent of IT, human resource managers have more flexibility in deciding

- What kinds of information to consider in performance appraisals;
- What types of feedback to give to employees;
- Whether employees should be evaluated on an individual or group basis.
Standard agency models limit the principal’s choice over monitoring technologies.

Hard to reconcile model predictions with stylized facts:

1. Aggregation of performance information into simple grades;
2. Coexistence of individual and group-based incentive schemes among firms with similar production technologies.
A principal-agent model with costly and flexible monitoring:

- **Flexibility**: the principal can implement any partitional monitoring technology;
- **Cost**: increase in the fineness of the partition.

Characterize the optimal monitoring technology through the trade-off between compensation cost and monitoring cost.

Explain stylized facts:

1. Aggregation of various performance information into simple grades;
2. Coexistence of individual and group-based incentive schemes among firms with similar production technologies.
Agenda

1. Baseline model
2. Extensions
3. Conclusion
Agenda

1. Baseline model
2. Extensions
3. Conclusion
A risk-neutral principal and a risk-averse agent.

The agent’s payoff $u(w) - c(a)$:

- Consumption $w \geq 0$: $u(0) = 0$, $u' > 0$, $u'' < 0$;
- Effort $a \in \{0, 1\}$: $c(1) = c > c(0) = 0$.

Each effort $a$ generates a probability space $(\Omega, \Sigma, P_a)$, where $P_a$ has a well-defined probability density function $p_a$.

The principal’s goal: induce high effort.
Incentive Contract

A pair of monitoring technology $\mathcal{P}$ and wage scheme $w : \mathcal{P} \rightarrow \mathbb{R}_+$:

- $\mathcal{P}$: a finite partition of $\Omega$ whose elements belong to $\Sigma$;
- $w(\cdot)$: maps each $A \in \mathcal{P}$ to a $w(A) \geq 0$.

$\mathcal{P} = (A_1, \cdots, A_N)$ and $a \in \{0, 1\}$ induce $\vec{p}_a = (p_{a,1}, \cdots, p_{a,N})$, where $p_{a,n} = P_a(A_n)$.
Timeline:

1. Principle commits to $\langle \mathcal{P}, w(\cdot) \rangle$;
2. The agent privately exerts $a \in \{0, 1\}$;
3. Nature draws $\omega \in \Omega$ according to $P_a$;
4. $A(\omega) \in \mathcal{P}$ is publicly realized;
5. The principal pays $w(A(\omega))$. 
Implementation Cost

Implementation cost (conditional on $a$):

$$\sum_{A \in \mathcal{P}} w(A)P_a(A) + H(\vec{p}_a).$$

1. $\sum_{A \in \mathcal{P}} w(A)P_a(A)$: compensation cost;
2. $H(\vec{p}_a)$: monitoring cost, including the cost of
   - Processing information;
   - Compiling performance grade;
   - Communicating the result to the agent.
Assumption 1.

For any \( N \in \mathbb{N} \) and \( \vec{p} \in \Delta^N \),

(i) \( H(p_1, \cdots, p_N) = H(p_{\pi(1)}, \cdots, p_{\pi(N)}) \) for all permutation \( \pi \) of \( \{1, \cdots, N\} \);

(ii) \( H(p_1, \cdots, p_N) < H(p'_1, p''_1, \cdots, p_N) \) for all \( p'_1, p''_1 > 0 \) such that \( p'_1 + p''_1 = p_1 \).

The monitoring cost is

- Invariant to the labeling of monitoring outcomes;
- Increasing in the fineness of the partition.

E.g.: \( H(\vec{p}) = -\sum_{n=1}^{N} p_n \log(p_n) \) or \( f(N) \).
Main Tradeoff

As $\mathcal{P}$ becomes finer:

- The compensation cost decreases;
- The monitoring cost increases;

The optimal monitoring technology balances the trade-off between the compensation cost and the monitoring cost.
Define a random variable $Z : \Omega \to \mathbb{R}$ by

$$Z = 1 - \frac{p_0}{p_1}.$$ 

For each $A \in \Sigma$, define

$$z(A) = 1 - \frac{P_0(A)}{P_1(A)} = \mathbb{E}[Z|A, a = 1].$$

A contract is incentive compatible for the agent if

$$\sum_{A \in \mathcal{P}} u(w(A))z(A)P_1(A) \geq c.$$  \hspace{1cm} (IC)

$z$-value contains all the useful information for deterring shirking.
Optimal Incentive Contract

The optimal contract \( \langle P^*, w^*(\cdot) \rangle \) solves

\[
\min_{\langle P, w(\cdot) \rangle} \sum_{A \in P} w(A) P_1(A) + H(\tilde{p}_1),
\]

s.t. (IC) and (LL).
Benchmark: Exogenous Monitoring

Standard agency models take $\mathcal{P}$ as exogenously given and solve for

$$\min_{w:\mathcal{P}\rightarrow \mathbb{R}} \sum_{A \in \mathcal{P}} w(A)P_1(A), \text{ s.t. (IC) and (LL)}.$$ 

Lemma 1.

*For any given $\mathcal{P}$, there exists $\lambda > 0$ such that for all $A \in \mathcal{P}$, $u'(w^*(A)) = \frac{1}{\lambda z(A)}$ if and only if $w^*(A) > 0$.***
**Z-convexity**

For each $A \in \Sigma$, let $Z(A)$ denote the image of $A$ under $Z$.

$Z(\Omega)$ is the set of all feasible $z$-values.

**Definition 1.**

* A set $A \in \Sigma$ is Z-convex if for all $\omega', \omega'' \in A$,

$$\{ \omega \in \Omega : \exists s \in (0, 1) \text{ s.t. } Z(\omega) = (1 - s) \cdot Z(\omega') + s \cdot Z(\omega'') \} \subset A.$$

When $Z(\Omega) \subset \mathbb{R}$ is connected, the Z-convexity of $A$ reduces to the convexity of $Z(A)$. 
Theorem 1.

Under Assumption 1, for any $\mathcal{P}^* = \{A_1, \ldots, A_N\}$,

(i) Any $A \in \mathcal{P}^*$ is $Z$-convex;

(ii) In case $Z(\Omega)$ is connected, there exists $\{\hat{z}_n\}_{n=0}^N$ such that each $A_n$ contains $\{\omega : Z(\omega) \in (\hat{z}_{n-1}, \hat{z}_n)\}$. 
Aggregation of (potentially) high-dimension performance information into a coarse, single-dimensional grade;

The distributions of the optimal performance measure satisfy the strong MLRP with respect to $\mathbf{z} \prec (A \mathbf{z} \prec A' \text{ if } z(A) \leq z(A'))$.

Alternative theories on endogenous monitoring technology:
- Costly verification: Banker and Datar (1980), Dye (1986);
- Linear contract with normally distributed signals.
Application

Design and implementation of
- Multi-source feedback systems;
- Grading schemes.
Agenda

1. Baseline model
2. Extensions
   - Multiple actions
   - Multiple agents
3. Conclusion
Multiple Actions

The agent’s action space $\mathcal{A}$ is finite, and the principal wants to induce $a^* \in \mathcal{A}$.

Define $\mathcal{D} = \mathcal{A} - \{a^*\}$. For each $a \in \mathcal{D}$, define

$$Z_a = 1 - \frac{p_a}{p_{a^*}}.$$  

For each $A \in \mathcal{P}$, define

$$z_a(A) = 1 - \frac{P_a(A)}{P_{a^*}(A)}.$$  

A contract is incentive compatible if for each $a \in \mathcal{D}$,

$$\sum_{A \in \mathcal{P}} u(w(A)) P_{a^*}(A) z_a(A) \geq c(a^*) - c(a). \quad (IC_a)$$


Take any profile \( \{\lambda_a\}_{a \in \mathcal{D}} \) of non-negative reals, and define 

\[
\bar{Z} = \sum_{a \in \mathcal{D}} \lambda_a \cdot Z_a.
\]

**Definition 2.**

A set \( A \in \Sigma \) is \( \bar{Z} \)-convex if for all \( \omega', \omega'' \in A \),

\[
\{ \omega \in \Omega : \exists s \in (0, 1) \text{ s.t. } \bar{Z}(\omega) = (1 - s) \cdot \bar{Z}(\omega') + s \cdot \bar{Z}(\omega'') \} \subset A.
\]
Corollary 1.

Under Assumption 1, for any \( P^* = \{ A_1, \cdots, A_N \} \), there exist non-negative reals \( \{ \lambda_a \}_{a \in D} \) with \( \max_{a \in D} \lambda_a > 0 \) such that

(i) Any \( A \in P^* \) is \( \overline{Z} \)-convex;

(ii) In case \( \overline{Z}(\Omega) \) is connected, there exists \( \{ \hat{z}_n \}_{n=0}^N \) such that each \( A_n \) contains \( \{ \omega : \overline{Z}(\omega) \in (\hat{z}_{n-1}, \hat{z}_n) \} \).
Implications

Balanced scorecard (Kaplan and Norton (1992)):

- $Z_a$: performance in aspect $a$;
- $\lambda_a$: weight of aspect $a$.

$\{\lambda_a\}_{a \in D}$ determines how resources are allocated across the monitoring of various deviations.

A big $\lambda_a$ means:

- The boundary of $A \in \mathcal{P}^*$ varies sensitively with $Z_a$;
- Many resources are spent on the monitoring of deviation $a$. 
Application: Multiple Tasks

The agent can exert either the high effort \((a_i = 1)\) or the low effort \((a_i = 0)\) in each of two tasks \(i = 1, 2\).

Each \(\vec{a} = a_1a_2\) generates \((\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, P_{a_1} \times P_{a_2})\).

The principal wants to induce \(\vec{a} = 11\).
Multiple Tasks

A special case of the multi-action problem:

- \( A = \{11, 10, 01, 00\} \), \( a^* = 11 \), \( D = A - \{11\} \);
- Independence \( \implies Z_{00} = Z_{10} + Z_{01} - Z_{10} \cdot Z_{01} \).

The optimal monitoring technology conducts overall evaluations that aggregate performance information across tasks.

\( (\lambda_{00} + \lambda_{01})/(\lambda_{00} + \lambda_{10}) \) captures how resources are allocated across the monitorings of task 1 and task 2.
The Optimal Monitoring Technology (Cont’d)

Figure: Optimal bi-partition when only $\lambda_{00}$ is binding
The Optimal Monitoring Technology

**Figure:** Optimal bi-partition when $\lambda_{00}$ is slack
How does \((\lambda_{00} + \lambda_{01})/(\lambda_{00} + \lambda_{10})\) depend on:

- Whether tasks are substitutes or complements?
- Underlying signal distribution?
Agenda

1. Baseline model
2. Extensions
   - Multiple actions
   - Multiple Agents
3. Conclusion
Background

Holmstrom (1979)’s sufficient principle suggests that in the absence of technological linkage or statistical correlation, agents should be evaluated on an individual basis.

Standard theory uses technological linkage to justify group reward, and common productivity shock to explain tournament.

Hard to reconcile with recent empirical findings, that individual and group-based compensation schemes coexist among firms with similar production technologies.
Multiple Agents

Each of two agents $i = 1, 2$ can exert either the high effort ($a_i = 1$) or the low effort ($a_i = 0$).

Each $a_i$ independently generates $(\Omega, \Sigma, P_{a_i})$.

Each $\bar{a} = a_1a_2$ generates $(\Omega \times \Omega, \Sigma \otimes \Sigma, P_{a_1} \times P_{a_2})$.

The principal wants both agents to exert the high effort.
An incentive contract consists of

1. \( \mathcal{P} \): a finite partition of \( \Omega \times \Omega \);
2. \( (w_1, w_2) : \mathcal{P} \to \mathbb{R}_+^2 \).

Implementation cost:

\[
\sum_{A \in \mathcal{P}} \left[ w_1(A) + w_2(A) \right] \cdot P_\bar{a}(A) + \mu \cdot H(\bar{p}_\bar{a})
\]

where \( \mu \) is the marginal cost of monitoring.
z-Value

For each \( i = 1, 2 \), define

\[
Z_i = 1 - \frac{p_{a_i=0, a_{-i}=1}}{p_{11}}.
\]

For each \( A \in \mathcal{P} \) and \( i \), define

\[
z_i(A) = 1 - \frac{P_{a_i=0, a_{-i}=1}(A)}{P_{11}(A)}.
\]

A contract is incentive compatible if for each \( i \),

\[
\sum_{A \in \mathcal{P}} u(w_i(A)) \cdot P_{11}(A) \cdot z_i(A) \geq c_i
\]
Z-Convexity

Define $\tilde{Z} = (Z_1, Z_2)$.

**Definition 1.**
A set $A \in \Sigma \otimes \Sigma$ is $\tilde{Z}$-convex if for all $\tilde{\omega}', \tilde{\omega}'' \in A$, 
\[
\left\{ \tilde{\omega} : \exists s \in (0, 1) \text{ s.t. } \tilde{Z}(\tilde{\omega}) = s \cdot \tilde{Z}(\tilde{\omega}') + (1 - s) \cdot \tilde{Z}(\tilde{\omega}'') \right\} \subset A.
\]
Theorem 2.

Under Assumption 1,

(i) any $A \in \mathcal{P}^*$ is $\vec{Z}$-convex;

(ii) When $\vec{Z}(\Omega \times \Omega)$ is connected and the distribution of $\vec{Z}$ is atomless, the boundaries of each $A \in \mathcal{P}^*$ consist of straight line segments in the space of $(Z_1, Z_2)$.
Quad-Partition: Individual Evaluation

\[ W_1 > 0 \quad W_2 > 0 \]
\[ W_1 > 0 \quad W_2 = 0 \]
\[ W_1 = 0 \quad W_2 > 0 \]
\[ W_1 = 0 \quad W_2 = 0 \]

\( Z_2 \)

\( Z_1 \)
Bi-Partition: Group Evaluation

\[ W_1 = 0 \]
\[ W_2 = 0 \]
\[ W_1 > 0 \]
\[ W_2 > 0 \]
\[ Z_1 \]
\[ Z_2 \]
Bi-Partition: Tournament

\[ Z_1 \]
\[ Z_2 \]

\[ W_1 = 0 \]
\[ W_2 > 0 \]

\[ W_1 > 0 \]
\[ W_2 = 0 \]
Special Case: \( H(\bar{p}) = f(N) \)

For each \( N \in \mathbb{N} \), define

\[
C(N) = \min_{\mathcal{P} : |\mathcal{P}| = N} \sum_{A \in \mathcal{P}} (w_1(A) + w_2(A))P_{11}(A),
\]

s.t. (IC) and (LL).
Optimal Multi-Agent Contract

Implementation cost

\[ C(4) + \mu \cdot f(4) \]
\[ C(3) + \mu \cdot f(3) \]
\[ C(2) + \mu \cdot f(2) \]

\( \mu \) values:
- \( \mu(2) \)
- \( \mu(3) \)
Implications

Explain the use of different multi-agent contracts by factors affecting $\mu$ (e.g., Bloom and Van Reenen (2010)):

- IT, labor market regulation, access to advanced managerial practices;
- Human capital share of production, product market competition.

Individual evaluation is sub-optimal when $\mu$ is large.
Literature Review

Moral hazard:
- Single-agent: Holmstrom (1979);
- Multi-agent: Holmstrom (1982), Mookherjee (1984);

Heterogeneity in managerial practices:
- Theory: Chassang (2010), Halac and Prat (working paper).

Conclusion

A principal-agent model with costly and flexible monitoring.

Endogenize the choice of monitoring technology through the trade-off between compensation cost and monitoring cost.

Explain stylized facts:

1. Information aggregation; overall performance evaluation.
2. Use of different incentive schemes among firms with similar production technologies.