Multi-task Incentive Contract and Performance Measurement with Multidimensional Types*

Dawen Meng †
Institute for Advanced Research
Shanghai University of Finance and Economics
Shanghai, China. 201600

Guoqiang Tian ‡
Department of Economics
Texas A&M University
College Station, Texas 77843

August 23, 2011

Abstract

This paper provides a new explanation for the dominance of the low-powered incentive contract over the high-powered incentive contract using a mixed model of moral hazard and adverse selection. We first show that the power of incentives in the second-best contract is lower than that in the first-best contract in the presence of either unobservable risk aversion or cost. We then consider the case that both risk aversion and cost of the agent are unobservable to the principal. We solve this multidimensional mechanism design problem under two assumptions with regard to the structures of performance measurement system and wage contract. It is shown that if the deterministic and stochastic components of different performance measures vary proportionally, the principal is inclined to provide a low-powered incentive contract. Moreover, it is shown that if the base wage depends only on a quadratic function rather than the direction of the performance wage vector, no incentive is provided for most of the performance measures in an orthogonal performance measurement system.

Keywords: multi-task incentive model, multidimensional mechanism design, low-powered incentive, missing incentive

---

*This research is supported by the 211 Project (Phase 3) for Shanghai University of Finance and Economics. The first author thanks the Humanities and Social Sciences Project at the Ministry of Education of China (09YJC790184), Innovation Program of Shanghai Municipal Education Commission and the second author thanks the National Natural Science Foundation of China (NSFC-70773073) and the Program to Enhance Scholarly and Creative Activities at Texas A&M University as well as the Cheung Kong Scholars Program at the Ministry of Education of China for financial support.

†Email: devinmeng@yahoo.com.cn

‡Email: tian@econmail.tamu.edu

1 Introduction

The central topic of moral hazard problem is to design incentives to elicit the agent's effort. Higher incentive pay will induce the agent to work harder and consequently bring higher surplus to the principal. However, the arrangements employers typically reach with their employees in reality look quite different from the incentive contracts derived by economic theorists. Low-powered incentives are very common in practice, especially within organizations. Many firms prefer to pay fixed compensation and offer continued employment to all but clearly unsatisfactory employees. Good examples are the government agencies and public firms, which are generally blamed for poor performance because their managers and workers lack high-powered incentives. Based on the standard transaction-cost and principal-agent economics, several theories have been provided to explain why low-powered incentives are employed even if objective performance measures are available and agents are highly responsive to incentive pay.

Williamson (1985) argues that weak incentive arises from opportunism and incompleteness of contracts. He shows that the use of high-powered incentives would raise undesirable side problems such as exploitation, inefficient asset utilization and accounting manipulations. For example, if supplying a single large customer would require a firm to make a large investment in an asset that cannot be used readily for other purposes, the supplier may reasonably fear exploitation by the customer: once the investment is made, the customer could force a lower price on the supplier. The problem is not simply that one party to the transaction may end up being treated unfairly. The bigger problem is that as people will anticipate this possibility, the transaction may not take place at all. Even if the manufacturer intends to keep his commitment, a transaction beneficial to both sides may be aborted because the supplier cannot trust him. One possible solution to this problem is to write a court enforceable contract specifying how each party must behave under a number of different contingencies. Unfortunately as Williamson points out, contracts are not always effective in preventing opportunism in that due to limits to human information-processing abilities, it is often impossible to anticipate all possible contingencies, let alone specify them in a contract. This leaves scope for opportunism, so the supplier and manufacturer have to replace the high-powered market transaction with low-powered incentive inside firms.

Holmstrom and Milgrom (1991) show that the power of incentives on some tasks relies on the principal's ability to monitor other aspects of the agent's performance. The agents may shift their effort from some activities where their individual contributions are poorly measured to the better-measured and well-compensated activities. For this reason, high-powered incentive may be dysfunctional in multi-task environment.

The conclusion of Holmstrom and Milgrom (1991) relies largely on the assumptions that agent is risk averse and the tasks are substitute. On the contrary, Baker (1992) shows that low-powered incentive might arise even with risk neutral agent when the performance measure used and the principal's true objective are weakly correlated. That means if the performance measure does not respond to the agent's actions in the same way that the principal's objective responds to these actions, the firm will reduce the intensity of the incentive contract.

Aside from piece rates or commissions, another way that firms use to compensate agent is by relative performance evaluation such as awarding promotions to the member of a work group who performs best. In fact, in some political organizations such as government agencies, the agents are rewarded mainly on relative performance measures rather than on their individual output. One function of relative performance evaluation is allowing the principal to use flatter incentives.[1] In this sense, the literature justifying relative performance evaluations also gives partial explanations on the arising of low-powered incentive.

Lazear and Rosen (1981) show in a standard single moral-hazard framework that the promotion-based incentive scheme can achieve the same results as other incentive schemes can. They argue that the dominance of the promotion-based incentive scheme over the piece-rate linear scheme and the standard bonus scheme arises from the fact that obtaining ordinal measures generally requires less resources than obtaining cardinal measures. Green and Stokey (1983) and Nalebuff and Stiglitz (1983) show also in the single moral-hazard framework that the relative performance evaluation incentive scheme may dominate the absolute performance evaluation scheme when the agents are risk-averse and there are shocks that are common to all the agents. Obviously, the promotion-based incentive scheme, by filtering out common randomness, can reduce the risk that would otherwise be imposed on the agents and requires compensation. Therefore, the relative performance evaluation improves the principal's efficiency.

Another class of literature closely related to the present paper is in the area of multidimensional mechanism design. The multidimensional mechanism design problem arises when the agent possesses multiple characteristics. Its implementability is much more complicated than that in the unidimensional mechanism design problem because of the lack of a natural order on types. The most notable publications about multidimensional mechanism design include Armstrong (1996), Rochet and Choné (1998) and Basov (2001) among many others. Armstrong (1996) was the first to formulate this problem in a multiproduct nonlinear pricing setting. In this seminal paper, he develops an integration along rays technique and characterizes the pricing contract for the case with cost-based tariff. Rochet and Choné(1998) analyze

[1]Rank order tournaments is a simple and widely used form of relative performance evaluation. This classical form of relative performance compensation has the particularity of using only an ordinal ranking of performance. By awarding high and low prizes based on relative performance, a principal can elicit higher effort level than with a scheme that involves the same wage bill but equal wages. An agent's performance is increasing in the spread between the winner and loser prize, ceteris paribus, rather than the absolute payment levels. It therefore allows the principal to elicit a higher effort level using lower performance pay.

a general multidimensional screening model. They show that, in general, the monopolist will use mechanisms in which there is bunching, i.e., different consumer-types will be treated equally. They develop a methodology – sweeping technique, for dealing with bunching in multiple dimensions. Basov (2001) takes advantage of control theoretic tools and develops a "generalized Hamiltonian approach" for solving the multidimensional mechanism design problem.

In this paper, we provide a new explanation for the dominance of low-powered over high-powered incentives from the perspective of multidimensional mechanism design. Our contributions are three-fold. First, we attribute the missing of incentive to multidimensionally asymmetric information. We show that in the presence of one-dimensional asymmetric information, the second-best incentive contract is flatter than the first-best contract. In the multidimensional screening model, however, if the base wage depends only on the $\Sigma-$norm of the performance wage, the efficient linear compensation rule contains no incentive component for most of the performance measures.

Secondly, in contrast to most of the existing literature dealing with the power-of-incentive issue in the framework of pure moral hazard, our analysis is made in a mixed model of both moral hazard and adverse selection. The standard moral hazard model concerns only the trade-off between insurance and incentives. In these environments, the compensation based on certain "risky" performance measure serves the dual functions of increasing both profits and risk. A tension between these two functions arises when the agent is risk averse. Higher pay induces the agent to exert a higher level of effort and thus increases the principal's profit. On the other hand, high wage also exposes the agent to unwanted risk, which requires an extra risk premium as compensation. Consequently, when choosing contract, the principal trades off the benefits of more effort against higher wage costs. Most of the existing studies assume that only the agent's actions are unobservable. In contrast, our paper assumes that the agent is privately informed about both his actions and types. The principal therefore is faced with an additional tradeoff – a tradeoff between efficiency and rent extraction. We show that the incentive contract is inclined to be flatter in mixed model than in pure moral hazard model.

Thirdly, we develop a "delegating" method for the complex multidimensional mechanism design problem. The intuition behind this method is a tradeoff between authority and complexity. As a centralized way of resource allocation, an incentive mechanism vests all the decision making authority in the principal, but it needs to process information transmitted by the agent. The multidimensional information increases the principal's information processing cost and complexity of writing a contract. The more authority the principal owns, the more information he has to process. Therefore, in order to save information processing cost and avoid complexity of writing a multidimensional contract, the principal may choose to delegate part of his authority to the agent. Two extreme cases of this tradeoff in practice are a decentralized market economy which distributes all the decision making authority to individual agents but their communications requirements are minimal(See Hurwicz (1972, 1979, 1986), Mount and Reiter (1974), Walker (1977), Osana (1978), Tian(1994, 2004, 2006) among others for detailed discussion); and socialist economy in which a central planner has almost all the authority but a great amount of information has to be processed. In this paper, part of the principal's authority is delegated to the agent under the assumption that the fixed component of compensation bases only on a quadratic form ($\Sigma - norm$) of the vector of incentive compensation coefficients. This assumption deprives some of the principal's degrees of freedom but decreases drastically the amount of information required. The multidimensional mechanism design problem is therefore easily solved.

The remainder of the paper is organized as follows. The basic multi-task principal-agent model is specified in Section 2, along with a characterization of the first-best contract. The results with unobservable risk aversion are examined in Section 3. The results with unobservable cost are discussed in Section 4. Section 5 considers the optimal incentive contract in a general environment where risk aversion and cost are both unobservable. Finally, in Section 6, some concluding remarks are given.

2 Basic Model

Consider a principal-agent relationship in which the agent controls n activities that influence the principal's payoff. The principal is risk neutral and her gross payoff is a linear function of the agent's effort vector e:

$$V(e) = \beta' e + \eta, \tag{1}$$

where the $n-$dimensional vector β characterizes the marginal effect of the agent's effort e on $V(e)$, and η is a noise term with zero mean. The agent chooses a vector of efforts $e = (e_1, \cdots, e_n)' \in \mathbb{R}^n_+$ at quadratic personal cost $\frac{e'Ce}{2}$, where C is a symmetric positive definite matrix. The agent's preferences are represented by the negative exponential utility function utility $u(x) = -e^{-rx}$, where r is the agent's absolute risk aversion and x is his compensation minus personal cost.

It is assumed that there is a linear relation between the agent's efforts and the expected levels of the performance measures:

$$P_i(e) = b_i' e + \varepsilon_i, i = 1, \cdots, m, \tag{2}$$

where $b_i \in \mathbb{R}^n$ captures the marginal effect of the agent's effort e on the performance measure $P_i(e)$; $B = (b_1, \cdots, b_m)'$ is an $m \times n$ matrix of performance parameters, and it is assumed that the matrix B has full row rank m so that every performance measure is indispensable; and $\epsilon = (\varepsilon_1, \cdots, \varepsilon_m)'$ is an $m \times 1$ vector of normally distributed variables with mean zero and variance-covariance matrix Σ.

Definition 1 *(Orthogonality) A performance system is said to be orthogonal if and only if* $b_i'b_j = 0$ *and* $Cov(\varepsilon_i, \varepsilon_j) = 0$, *for* $i \neq j$, *that is,* $B'B$ *and* Σ *are both diagonal matrices.*

Definition 2 *(Signal-noise ratio) The signal-noise ratio of a performance measure* $P_i = b_i'\beta + \varepsilon_i$ *is the ratio of the inner product of the expected marginal effect of activity on a measure divided by the variance of the noise of the measure:* $\gamma_i = \frac{b_i'b_i}{\sigma_i^2}$.

Definition 3 *(Congruence) The congruence of a performance measure* $P_i = b_i'\beta + \varepsilon_i$ *is measured by* $\Upsilon_i = \cos(\widehat{b_i, \beta})$, *where* $(\widehat{b_i, \beta})$ *is the angle between the vector of payoff sensitivities* β *and the vector of performance measure sensitivities* b_i.

According to this definition, performance measure $P_i = b_i'\beta + \varepsilon_i$ is incongruent if vector b_i and vector β are linearly independent, which in turn implies that $(\widehat{b_i, \beta}) \neq 0$. Moreover, a more congruent performance measure is characterized by a smaller angle $(\widehat{b_i, \beta})$, and hence, implies a higher measure of congruity Υ_i due to the definition of the cosine.

The principal compensates the agent's effort through a linear contract:

$$W(e) = w_0 + w'P(e), \tag{3}$$

where $P(e) = (P_1(e), \cdots, P_m(e))'$, w_0 denotes the base wage, and $w = (w_1, \cdots, w_m)'$ the performance wage. Under this linear compensation rule, the principal's expected profit is $\Pi_p = \beta'e - w_0 - w'Be$, and the agent's certainty equivalence is

$$CE_a = w_0 + w'Be - \frac{1}{2}e'Ce - \frac{r}{2}w'\Sigma w. \tag{4}$$

The principal's problem is to design a contract (w_0, w) that maximizes her expected profit Π_p while ensuring the agent's participation and eliciting his optimal effort.

The optimization problem of the principal is thus formulated as:

$$\begin{cases} \underset{\{w_0,w,e\}}{max} \; \beta'e - w_0 - w'Be \\ \quad \text{s.t:} IR : w_0 + w'Be - \frac{1}{2}e'Ce - \frac{r}{2}w'\Sigma w \geqslant 0 \\ \quad\quad IC : e \in argmax_{\tilde{e}} \left[w_0 + w'B\tilde{e} - \frac{1}{2}\tilde{e}'C\tilde{e} - \frac{r}{2}w'\Sigma w \right] \end{cases}.$$

The IR constraint ensures that the principal cannot force the agent into accepting the contract, and here the agent's reservation utility is normalized to zero; the IC constraint represents the rationality of the agent's effort choice.

We now consider the effort choosing problem of the agent for a given incentive scheme (w_0, w). Since the objective is concave by noting that the second-order derivative of CE_a with respect to e is a negative definite matrix $-C$, the maximizer can be determined by the first-order condition: $Ce = B'w$. After replacing e with $e^* = C^{-1}B'w$ and substituting the IR constraint written with equality into the principal's objective function, the principal's optimization problem simplifies to:

$$\underset{w \in \mathbb{R}^n}{max} \left[\beta'C^{-1}B'w - \frac{1}{2}w' \left(BC^{-1}B' + r\Sigma \right) w \right].$$

The optimal wage contract and effort to be elicited are therefore:

$$w^{fb} = \left[BC^{-1}B' + r\Sigma\right]^{-1} BC^{-1}\beta \tag{5}$$

$$w_0^{fb} = \frac{rw^{fb'}\Sigma w^{fb} - w^{fb'}BC^{-1}B'w^{fb}}{2} \tag{6}$$

$$e^{fb} = C^{-1}B'w^{fb}. \tag{7}$$

The resulting surplus of the principal is[2]

$$\pi^{fb} = \frac{1}{2}\beta' C^{-1}B' \left[BC^{-1}B' + r\Sigma\right]^{-1} BC^{-1}\beta. \tag{8}$$

A higher incentive pay could induce the agent to implement a higher effort, but it will also expose the agent to a higher risk. It therefore requires a premium to compensate the risk-averse agent for the risk he bears. The optimal power of incentive is therefore determined by the tradeoff between incentive and insurance. Moreover, the results above show that in multi-task agency relationships the degree of congruity of available performance measures and the agent's task-specific abilities also affects the power and distortion of incentive contract, which is in line with many previously known studies such as those of Feltham and Xie (1994), Baker (2002) and Thiele (2008).

3 The optimal contract with unobservable risk aversion

The first-best incentive contract stated above relies crucially on the agent's attitude towards risk. In the following, we assume that risk aversion r is private information of the agent, and its distribution function $F(r)$ and density function $f(r)$ supported on $[\underline{r}, \overline{r}]$ are common knowledge to all parties. This assumption is different to most of the previous studies in which risk aversion is regarded as a publicly observed variable. The principal then has to offer a contract menu $\{w_0(\hat{r}), w(\hat{r})\}$ contingent on the agent's reported "type" $\hat{r}$ to maximize her expected payoff. $\{w_0(\hat{r}), w(\hat{r})\}$ is said to be implementable if the following incentive compatibility condition is satisfied:

$$w_0(r) + \frac{1}{2}w(r)'\left[BC^{-1}B' - r\Sigma\right] w(r) \geqslant w_0(\hat{r}) + \frac{1}{2}w(\hat{r})'\left[BC^{-1}B' - r\Sigma\right] w(\hat{r}) \tag{9}$$

Let $U(r,\hat{r}) \equiv w_0(\hat{r}) + \frac{1}{2}w(\hat{r})'\left[BC^{-1}B' - r\Sigma\right] w(\hat{r})$, and $U(r) \equiv U(r,r)$. Then the implementability condition of $\{U(r), w(r)\}$ is stated equivalently as:

$$\exists w_0 : [\underline{r}, \overline{r}] \to \mathbb{R}_+, \forall (r,\hat{r}) \in [\underline{r}, \overline{r}]^2, U(r) = \max_{\hat{r}} \left\{ w_0(\hat{r}) + \frac{1}{2}w(\hat{r})'\left[BC^{-1}B' - r\Sigma\right] w(\hat{r}) \right\} \tag{10}$$

[2] Notice that the optimal incentive contract w^{fb} could be regarded as a "partial" generalized least squares regression of the payoff sensitivity β on performance measure sensitivities B'. If the agent is risk neutral ($r = 0$), and has no task-specific abilities across n independent tasks, i.e., $C = diag\{c, c, \cdots, c\}$, w^{fb} is actually the OLS regression parameter of β on B'.

The "Taxation Principle" [cf. Guesnerie (1981), Hammond (1979) and also Rochet (1985)] states that (**??**) is equivalent to the following very similar condition

$$\exists w_0 : \mathbb{R}^m \to \mathbb{R}_+, \forall r \in [\underline{r}, \overline{r}], U(r) = \underset{w}{max} \left\{ w_0(w) + \frac{1}{2} w'[BC^{-1}B' - r\Sigma]w \right\}. \quad (11)$$

It is possible to show that $U(\cdot)$ is continuous, convex [3](thus almost everywhere differentiable), and satisfies the envelop condition:

$$U'(r) = -\frac{1}{2} w'\Sigma w. \quad (12)$$

Conversely, if (**??**) holds and $U(r)$ is convex, then

$$U(r) \geqslant U(\hat{r}) + (r - \hat{r})U'(\hat{r}) = U(\hat{r}) - \frac{1}{2}(r - \hat{r})w'(\hat{r})\Sigma w(\hat{r}),$$

which implies the incentive compatibility condition $U(r) \geqslant U(r, \hat{r})$. Formerly, we have

Lemma 1 *The surplus function $U(r)$ and performance wage function $w(r)$ are implementable by the principal if and only if:*

*(1) envelop condition (**??**) is satisfied;*

(2) $U(r)$ is convex in r.

Substituting $U(r)$ into the principal's expected payoff, we get

$$\begin{aligned} \Pi &= \int_{\underline{r}}^{\overline{r}} [\beta' e^* - w_0(r) - w(r)'Be^*] f(r)\mathrm{d}r \\ &= \int_{\underline{r}}^{\overline{r}} \left\{ \beta' C^{-1} B' w(r) - \frac{1}{2} w(r)' \left[BC^{-1}B' + r\Sigma \right] w(r) - U(r) \right\} f(r)\mathrm{d}r. \end{aligned}$$

The principal's optimization problem is therefore:

$$\underset{U(r),w(r)}{max} \Pi, \text{ s.t.:} U(r) \geqslant 0, U'(r) = -\frac{1}{2} w(r)'\Sigma w(r), U(r) \text{ is convex.} \quad (13)$$

The following proposition summarizes the solution of the principal's problem.

Proposition 1 *If the hazard rate $\Phi(r)$ is nondecreasing, then the optimal wage contract is given by*

$$\begin{aligned} w^{sb}(r) &= \left[BC^{-1}B' + \Phi(r)\Sigma \right]^{-1} BC^{-1}\beta \quad (14) \\ w_0^{sb}(r) &= \frac{1}{2} \int_r^{\overline{r}} w^{sb}(\tilde{r})'\Sigma w^{sb}(\tilde{r}) d\tilde{r} - \frac{1}{2} w^{sb}(r)' \left[BC^{-1}B' - r\Sigma \right] w^{sb}(r), \quad (15) \end{aligned}$$

where $\Phi(r) \equiv r + \frac{F(r)}{f(r)}$.

[3]One way to define the convex functions is through representing them as maximum of the affine functions, that is, $s(x)$ is convex if and only if

$$s(x) = \underset{a,b \in \Omega}{max}(a \cdot x + b)$$

for some $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ and some $\Omega \subset \mathbb{R}^{n+1}$. In this example $a = -\frac{1}{2} w'\Sigma w, b = w_0(w) + \frac{1}{2} w' BC^{-1} B' w$, therefore $U(r) = \underset{(a,b)\in \mathbb{R}_- \times \mathbb{R}_+}{max} (ar + b)$ is a convex function in r.

Proof. See appendix. ■

To explore some of the properties of our model, let us now work with the special case in which there exists a one-to-one relationship between the performance measures and the activities: $B = diag\{b_{11}, b_{22}, \cdots b_{nn}\}$; the error terms are stochastically independent ($\Sigma = diag\{\sigma_1, \cdots, \sigma_n\}$) and the activities are technologically independent($C = diag\{c_1, \cdots, c_n\}$). Then

$$w_i^{fb}(r) = \frac{b_{ii}\beta_i}{b_{ii}^2 + rc_i\sigma_i^2}, \tag{16}$$

$$w_i^{sb}(r) = \frac{b_{ii}\beta_i}{b_{ii}^2 + \left(r + \frac{F(r)}{f(r)}\right) c_i\sigma_i^2}, i = 1, \cdots, n. \tag{17}$$

It is obvious that $w_i^{sb}(r) < w_i^{fb}(r)$ for all r except for $r = \underline{r}$. The following corollary is immediate.

Corollary 1 *Suppose that the tasks are technologically independent, the error terms are stochastically independent and the activities and performance measures are one-to-one corresponding to each other. Then the power of incentives on all tasks is lower than that in the first-best contract for all types except the least risk-averse one.*

If the risk aversion parameter is unobservable to the principal, the less risk-averse agent gains information rent by mimicking the more risk-averse one. The amount of information rent gained by an agent depends on the performance wage of agents with larger risk aversion, and therefore the basic tradeoff between efficiency and rent extraction leads to low-powered incentive for all but the least risk-averse types.

4 The optimal contract with unobservable cost

In this section we assume that the cost parameter is private information to the agent. To avoid the complicated multidimensional mechanism design issue momentarily, we assume that $C = cI$, that is, the tasks are technologically identical and independent. $\delta = \frac{1}{c}$ is assumed to be distributed on the support $[\underline{\delta}, \overline{\delta}]$, according to a cumulative distribution function $G(\delta)$ and density $g(\delta)$.

A contract menu $\{w_0(\delta), w(\delta)\}$ is said to be implementable if the following incentive compatibility condition is satisfied:

$$w_0(\delta) + \frac{1}{2}w(\delta)' \left[\delta BB' - r\Sigma\right] w(\delta) \geqslant w_0(\hat{\delta}) + \frac{1}{2}w(\hat{\delta})' \left[\delta BB' - r\Sigma\right] w(\hat{\delta}), \forall(\delta, \hat{\delta}) \in [\underline{\delta}, \overline{\delta}]^2. \tag{18}$$

Let $U(\delta, \hat{\delta}) \equiv w_0(\hat{\delta}) + \frac{1}{2}w(\hat{\delta})' \left[\delta BB' - r\Sigma\right] w(\hat{\delta})$, and $U(\delta) \equiv U(\delta, \delta)$. Then $\{U(\delta), w(\delta)\}$ is called implementable if

$$\exists w_0 : [\underline{\delta}, \overline{\delta}] \to \mathbb{R}_+, \forall(\delta, \hat{\delta}) \in [\underline{\delta}, \overline{\delta}]^2, U(\delta) = \max_{\hat{\delta}} \left\{ w_0(\hat{\delta}) + \frac{1}{2}w(\hat{\delta})' \left[\delta BB' - r\Sigma\right] w(\hat{\delta}) \right\} \tag{19}$$

or equivalently,

$$\exists w_0 : \mathbb{R} \to \mathbb{R}_+, \forall \delta \in [\underline{\delta}, \overline{\delta}], U(\delta) = \underset{w \in \mathbb{R}^m}{max} \left\{ w_0(w) + \frac{1}{2} w' \left[\delta BB' - r\Sigma \right] w \right\}. \tag{20}$$

$U(\delta)$ is necessarily continuous, increasing and convex in δ [4] and satisfies the envelop condition:

$$U'(\delta) = \frac{1}{2} w' BB' w. \tag{21}$$

Conversely, similar to the case with unobservable risk aversion, the convexity of $U(\delta)$ and envelop condition (**??**) implies

$$U(\delta) \geqslant U(\hat{\delta}) + (\delta - \hat{\delta}) U'(\hat{\delta}) = U(\hat{\delta}) + \frac{1}{2} (\delta - \hat{\delta}) w' BB' w = U(\delta, \hat{\delta})$$

, which in turn implies the implementability of contract. We summarize the above discussion in the following lemma.

Lemma 2 *The surplus function $U(\delta)$ and wage function $w(\delta)$ are implementable by the principal if and only if*

(1) $U'(\delta) = \frac{1}{2} w' BB' w$;

(2) $U(\delta)$ is convex in δ.

The second-best $\delta-$contingent contract solves the following optimization problem:

$$\begin{cases} \underset{w(\delta), U(\delta)}{max} \int_{\underline{\delta}}^{\overline{\delta}} \left\{ \delta \beta' B' w(\delta) - \frac{1}{2} w(\delta)' \left[\delta BB' + r\Sigma \right] w(\delta) - U(\delta) \right\} g(\delta) \mathrm{d}\delta \\ \text{s.t: } U(\delta) \geqslant 0, U'(\delta) = \dfrac{w' BB' w}{2}, U(\delta) \text{ is convex} \end{cases}.$$

Proposition 2 *With unobservable cost, if $\delta H(\delta)$ is decreasing, then the optimal wage is given by*

$$w^{sb}(\delta) = \left(H(\delta) BB' + \frac{r\Sigma}{\delta} \right)^{-1} B\beta \tag{22}$$

$$w_0^{sb}(\delta) = \frac{1}{2} \int_{\underline{\delta}}^{\delta} w^{sb}(\tilde{\delta})' BB' w^{sb}(\tilde{\delta}) d\tilde{\delta} - \frac{1}{2} w^{sb}(\delta)' \left[\delta BB' - r\Sigma \right] w^{sb}(\delta), \tag{23}$$

where $H(\delta) \equiv 1 + \frac{1 - G(\delta)}{\delta g(\delta)}$.

Proof. See appendix. ■

We now consider a special case of orthogonal performance measurement system, that is, $BB' = diag\{b_1' b_1, \cdots, b_m' b_m\}$, and $\Sigma = diag\{\sigma_1^2, \cdots, \sigma_m^2\}$. The first-best and second-best wage contracts are:

$$w_i^{fb}(\delta) = \frac{b_i' \beta}{b_i' b_i + \frac{r\sigma_i^2}{\delta}}, \tag{24}$$

$$w_i^{sb}(\delta) = \frac{b_i' \beta}{\left(1 + \frac{1 - G(\delta)}{\delta g(\delta)} \right) b_i' b_i + \frac{r\sigma_i^2}{\delta}}, i = 1, \cdots, n. \tag{25}$$

It is apparent that $w_i^{sb}(\delta) < w_i^{fb}(\delta)$ for all i and all $\delta \in [\underline{\delta}, \overline{\delta})$.

[4] In this case, let $a = \frac{1}{2} w' BB' w, b = w_0(w) - \frac{1}{2} w' \Sigma w$, then $U(\delta) = \underset{a,b}{max}(a\delta + b)$ is convex in δ.

Corollary 2 *For an orthogonal performance measurement system, the power of incentives on all tasks is strictly lower than that in the first-best contract for all but the most efficient types.*

When the agent possesses private information on his own cost, an agent with higher δ could accrue information rent by mimicking the agents with smaller δ. To minimize agency costs, optimality requires a downward distortion of the power of inefficient types' incentive wage.

5 The optimal contract with both unobservable cost and risk aversion

In this section we assume that both efficiency parameter δ and risk aversion r are unobservable to the principal. They are jointly distributed according to density function $f(\delta, r)$ on region $[\underline{\delta}, \overline{\delta}] \times [\underline{r}, \overline{r}]$. It is known that solutions to the multidimensional mechanism design models differ markedly from and are significantly more complex than their one-dimensional counterparts, essentially because different types of agents cannot be unambiguously ordered. Lacking methodology in the most general sense, different authors use different assumptions and methods to solve the multidimensional mechanism design models in the existing literature. Armstrong (1996) adopts an integration along rays procedure solving the relaxed problem of the principal, but the envelop condition could be satisfied by the pointwise maximizer only by accident, let alone the convexity condition. In order for the contract to be implementable, he makes two "separable" assumptions on the indirect and density functions. Rochet and Choné (1998) develop a general technique for dealing with the multi-dimensional screening problem, but it is workable only in the case where the dimensionality of type space is as same as the number of the principal's available instruments. The generalized Hamiltonian approach developed by Basov (2005) circumvents this difficulty but it obtains the optimal contract from a system of partial differential equations, which usually has no analytic solution. Therefore, one often has to rely upon the numerical techniques except for some very special function form.

In the following, we will treat the choosing of performance wage as a multidimensional mechanism design problem. In order to get an explicit analytic solution, we impose restrictions on the set of implementable allocations by assuming that the performance evaluation system is such that $BB' = k\Sigma$ or the base wage is based on the $\Sigma - norm$ of performance wage vector.

5.1 The performance measurement system with proportionally-varying deterministic and stochastic components: $BB' = k\Sigma$

If there exists a constant $k \in \mathbb{R}_+$ such that $BB' = k\Sigma$, then the deterministic and stochastic parts of a performance measurement system vary in similar ways. With this assumption, the agent's surplus could be represented as a function of a scalar $\theta_1 \equiv k\delta - r$

$$U(\delta, r) = \underset{w}{max} \left[w_0(w) + \frac{1}{2}\theta_1 w'\Sigma w \right] \equiv u(\theta_1). \tag{26}$$

Then, as in the previous sections, we get the convexity and envelop conditions ($u(\theta_1)$ is convex in θ_1 and $u'(\theta_1) = \frac{1}{2}w'\Sigma w$), which in turn implies the implementability of contract. We also define $\theta_2 = k\delta + r$. Then type vector (δ, r) is transformed linearly to (θ_1, θ_2). Notice that θ_1 is the only variable affecting the agent's choice, θ_2 is irrelevant and has no informative value to both parties.

For the convenience of discussion, we introduce some new notations. Let $D \equiv \big\{(\delta, r) \in \mathbb{R}_+^2 | \underline{\delta} \leqslant \delta \leqslant \overline{\delta}, \underline{r} \leqslant r \leqslant \overline{r}\big\}$ and $\Theta \equiv \big\{ (\theta_1, \theta_2) \in \mathbb{R}^2 \big| (\delta, r) \in D \big\}$ denote the domain of the original and transformed types. $\underline{\theta}_1 = k\underline{\delta} - \overline{r}, \overline{\theta}_1 = k\overline{\delta} - \underline{r}$ are minimal and maximal values of θ_1. Let $\varphi(\theta_1, \theta_2) = f\left(\frac{\theta_1+\theta_2}{2}, \frac{\theta_2-\theta_1}{2k}\right)\mathbf{J} = f\left(\frac{\theta_1+\theta_2}{2}, \frac{\theta_2-\theta_1}{2k}\right)\frac{1}{2k}$ denote the joint density of (θ_1, θ_2), where $\mathbf{J} \equiv \left|\det\left(\frac{\partial(\delta, r)}{\partial(\theta_1, \theta_2)}\right)\right| = \frac{1}{2k}$ is the jacobian of the transformation. $\varphi_1(\theta_1) \equiv \int_{\Theta_2(\theta_1)} \varphi(\theta_1, \theta_2)\mathrm{d}\theta_2$ and $\Phi_1(\theta_1) \equiv \int_{\underline{\theta}_1}^{\theta_1} \varphi_1(\theta_1)\mathrm{d}\theta_1$ represent the marginal density and marginal cumulative functions of θ_1, where $\Theta_2(\theta_1) \equiv \{\theta_2 \in \mathbb{R}_+ | (\theta_1, \theta_2) \in \Theta\}$. Denote by $\mathbb{H}(\theta_1) \equiv \frac{1-\Phi_1(\theta_1)}{\varphi_1(\theta_1)}$ the inverse hazard rate of θ_1.

Assumption 1 *The inverse hazard rate* $\mathbb{H}(\theta_1) = \frac{1-\Phi_1(\theta_1)}{\varphi_1(\theta_1)}$ *is nonincreasing in* θ_1.

Assumption 2 $k \leqslant \frac{\sigma_r}{\sigma_\delta}$, *$\sigma_\delta$ and σ_r are respectively standard deviations of r and δ.*

We further define the following regimes in accordance with three different information structures. The case where both θ_1 and θ_2 (or equivalently both δ and r) are observable is labeled as the first-best regime; the case where only θ_1 is observable is called the second-best regime; the case where neither θ_1 nor θ_2 is observable is called the third-best regime. We hereafter index the optimal contract and the resulting surplus with a superscript $i \in \{fb, sb, tb\}$. Equipped with the above notations and definitions, the principal's objective is rewritten as:

$$\begin{aligned} \Pi &= \iint\limits_D \left[\delta w'B\beta - \frac{1}{2}w'(\delta BB' + r\Sigma)w - U(\delta, r)\right] f(\delta, r)\mathrm{d}\delta\mathrm{d}r \\ &= \int_{\underline{\theta}_1}^{\overline{\theta}_1} \left[\frac{w'B\beta}{2k}g(\theta_1) - \frac{1}{2}w'\Sigma w h(\theta_1) - u(\theta_1)\varphi_1(\theta)\right]\mathrm{d}\theta_1, \end{aligned}$$

where $g(\theta_1) \equiv \int_{\Theta_2(\theta_1)}(\theta_1 + \theta_2)\varphi(\theta_1, \theta_2)\mathrm{d}\theta_2$, $h(\theta_1) \equiv \int_{\Theta_2(\theta_1)} \theta_2\varphi(\theta_1, \theta_2)\mathrm{d}\theta_2$.

As a consequence, the principal's optimal contract design problem simplifies to a unidimensional mechanism design problem:

$$\underset{w(\cdot), u(\cdot)}{max} \Pi, s.t. : u'(\theta_1) = \frac{1}{2}w'\Sigma w, u(\theta_1) \text{ is a convex function }, u(\theta_1) \geqslant 0. \tag{27}$$

Using the integration by parts technique, we obtain

$$\int_{\underline{\theta}_1}^{\overline{\theta}_1} u(\theta_1)\varphi_1(\theta_1)\mathrm{d}\theta_1 = \int_{\underline{\theta}_1}^{\overline{\theta}_1} \frac{1}{2}w'\Sigma w\left[1-\Phi_1(\theta_1)\right]\mathrm{d}\theta_1.$$

Then the principal's objective can be expressed as:

$$\Pi = \int_{\underline{\theta}_1}^{\overline{\theta}_1} \left\{\frac{1}{2k}g(\theta_1)w'B\beta - \frac{w'\Sigma w}{2}[h(\theta_1)+1-\Phi_1(\theta_1)]\right\}. \tag{28}$$

We ignore momentarily the convexity condition and simply maximize this expression point-wise with respect to w to get:

$$w^{tb}(\theta_1) = \frac{1}{2k}\frac{g(\theta_1)}{h(\theta_1)+1-\Phi_1(\theta_1)}\Sigma^{-1}B\beta = \rho(\theta_1)\Sigma^{-1}B\beta, \tag{29}$$

where

$$\rho(\theta_1) \equiv \frac{1}{2k}\frac{g(\theta_1)}{h(\theta_1)+1-\Phi_1(\theta_1)} = \frac{1}{2k}\frac{\theta_1+E_{\theta_2}(\theta_2|\theta_1)}{\mathbb{H}(\theta_1)+E_{\theta_2}(\theta_2|\theta_1)}.$$

The only task left is to check the convexity of function $u(\theta_1)$. Because

$$u''(\theta_1) = \left(\frac{\partial w}{\partial \theta_1}\right)'\Sigma w = \rho'(\theta_1)\rho(\theta_1)\beta'B'\Sigma^{-1}B\beta,$$

$u(\cdot)$ is convex if and only if $\rho(\cdot)$ is nondecreasing. It holds provided that: (i) $\mathbb{H}(\theta_1)$ is nonincreasing and (ii) $\eta(\theta_1) \equiv E_{\theta_2}(\theta_2|\theta_1)$ is nonincreasing. Condition (i) is the familiar monotone hazard rate property, while condition (ii) is equivalent to the requirement that $Cov(\theta_1,\theta_2) < 0$(See Lemma **??** in appendix). It holds if and only if Assumption **??** is satisfied because $Cov(\theta_1,\theta_2) = k^2\sigma_\delta^2 - \sigma_r^2$.

Substituting (**??**) into (**??**), we get the principal's expected profit[5]

$$\Pi^{tb} = \frac{1}{8k^2}E_{\theta_1}\left[\frac{(\theta_1+E_{\theta_2}(\theta_2|\theta_1))^2}{\mathbb{H}(\theta_1)+E_{\theta_2}(\theta_2|\theta_1)}\right]\beta'B'\Sigma^{-1}B\beta. \tag{30}$$

If θ_1 is observable, we only need to consider the participation constraint $u(\theta_1) \geqslant 0$ in (**??**). Then the second-best wage contract and surplus are:

$$w^{sb}(\theta_1) = \frac{1}{2k}\frac{\theta_1+E_{\theta_2}(\theta_2|\theta_1)}{E_{\theta_2}(\theta_2|\theta_1)}\Sigma^{-1}B\beta \tag{31}$$

$$\Pi^{sb} = \frac{1}{8k^2}E_{\theta_1}\left[\frac{(\theta_1+E_{\theta_2}(\theta_2|\theta_1))^2}{E_{\theta_2}(\theta_2|\theta_1)}\right]\beta'B'\Sigma^{-1}B\beta. \tag{32}$$

If both θ_1 and θ_2 are observable, we get the first-best contract and surplus as follows:

$$w^{fb}(\theta_1,\theta_2) = \frac{1}{2k}\frac{\theta_1+\theta_2}{\theta_2}\Sigma^{-1}B\beta, \tag{33}$$

$$\Pi^{fb} = \frac{1}{8k^2}E_\theta\left[\frac{(\theta_1+\theta_2)^2}{\theta_2}\right]\beta'B'\Sigma^{-1}B\beta, \tag{34}$$

where $\theta = (\theta_1,\theta_2)$. It is obvious that $w^{tb}(\theta_1) \leqslant w^{sb}(\theta_1)$, which is resulted from the traditional rent extraction-efficiency trade-off, but $w^{tb}(\theta_1)$ and $w^{fb}(\theta_1,\theta_2)$ are ambiguously ordered. The

[5] $E_{\theta_i}(\cdot)$ is the expectation operator with respect to θ_i.

intuition behind this is that owning more information about θ_2 is not necessarily helpful to the principal, since θ_2 is irrelevant to the agent's decision making. The principal's profits in these cases are well ordered as:

$$\begin{aligned}
\Pi^{fb} &= \frac{1}{8k^2}E_\theta\left[\frac{(\theta_1+\theta_2)^2}{\theta_2}\right]\beta' B'\Sigma B\beta \\
&= \frac{1}{8k^2}E_{\theta_1}E_{\theta_2}\left[\frac{(\theta_1+\theta_2)^2}{\theta_2}\middle|\theta_1\right]\beta' B'\Sigma B\beta \\
&\geqslant \frac{1}{8k^2}E_{\theta_1}\left[\frac{(\theta_1+E_{\theta_2}(\theta_2|\theta_1))^2}{E_{\theta_2}(\theta_2|\theta_1)}\right]\beta' B'\Sigma B\beta = \Pi^{sb} \\
&\geqslant \frac{1}{8k^2}E_{\theta_1}\left[\frac{(\theta_1+E_{\theta_2}(\theta_2|\theta_1))^2}{\mathbb{H}(\theta_1)+E_{\theta_2}(\theta_2|\theta_1)}\right]\beta' B'\Sigma B\beta = \Pi^{tb}.
\end{aligned}$$

The second line follows from the law of iterated expectation, the third line follows from Jensen's inequality since $\frac{(\theta_1+\theta_2)^2}{\theta_2}$ is convex in θ_2.

We summarize the above discussion in the following proposition:

Proposition 3 *Suppose that a performance system $\mathcal{P}=(B,\Sigma)$ is such that $BB'=k\Sigma$, and that Assumptions* **??** *and* **??** *are satisfied. Then we have*

1. *the power of incentive in the third-best wage contract is lower than that in the second-best wage contract: $w_i^{fb}(\theta_1) < w_i^{sb}(\theta_1)$ for all $i=1,\cdots,m$ and $\theta_1 \in [\underline{\theta}_1, \overline{\theta}_1)$, but it is unambiguously ordered compared with the first-best wage $w^{fb}(\theta_1,\theta_2)$.*
2. *the principal's expected surpluses in these regimes are ordered as:*

$$\Pi^{tb} \leqslant \Pi^{sb} \leqslant \Pi^{fb}.$$

Corollary 3 *For an orthogonal performance measurement system, under Assumptions* **??** *and* **??***, the third-best performance wage and resulting expected surplus of the principal are given by:*

$$w^{tb}(\theta_1) = \frac{1}{2k}\frac{\theta_1+E_{\theta_2}(\theta_2|\theta_1)}{\mathbb{H}(\theta_1)+E_{\theta_2}(\theta_2|\theta_1)}\begin{pmatrix}\frac{b_1'\beta}{\sigma_1^2}\\ \frac{b_2'\beta}{\sigma_2^2}\\ \vdots \\ \frac{b_m'\beta}{\sigma_m^2}\end{pmatrix} \tag{35}$$

$$\Pi^{tb} = \frac{1}{8k^2}E_{\theta_1}\left[\frac{(\theta_1+E_{\theta_2}(\theta_2|\theta_1))^2}{\mathbb{H}(\theta_1)+E_{\theta_2}(\theta_2|\theta_1)}\right]\sum_{i=1}^m\left(\frac{b_i'\beta}{\sigma_i}\right)^2. \tag{36}$$

The efficiency parameter δ and the risk aversion parameter r affect the agent's payoff in different ways. δ affects his effort provision ($e^*=\delta B'w$) and thus the expectation of his net surplus ($w_0+w'Be^*-\frac{1}{2\delta}e^{*'}e^*$), while r affects his risk premium ($\frac{r}{2}w'\Sigma w$). Misreporting these two parameters helps the agent get information rents with two degrees of freedom. However, if the variations of the deterministic and stochastic components of performance measures are propositional such that $BB' = k\Sigma$, all the information relevant to the agent's decision-making is contained in a scalar θ_1, and the agent is in fact deprived of one of his degrees of freedom. Therefore, the multidimensional mechanism design problem simplifies to the traditional one-dimensional problem. It is worth noting that if there is a single performance measure ($m = 1$), the condition $BB' = k\Sigma$ is necessarily satisfied. Thus the multidimensional mechanism design problem arises only in the joint presence of multidimensional types and multiple performance measures.

5.2 The $\Sigma - norm$ based base wage: $w_0 = w_0(w'\Sigma w)$

To reduce the information required by a mechanism and thus simplify the model, we need to impose some restrictions on the principal's authority and then delegate part of it to the agent. We assume that the base wage is based only on the $\Sigma - norm$ of performance wage vector w, that is, $w_0 = w_0(w'\Sigma w)$ where $w_0(\cdot)$ is a function of a scalar variable. That is to say, the employer determines the base wage solely on the $\Sigma-$ adjusted length of wage vector w rather than on the allocations of intensity among different performance measures. In this case, the contract menu (U, w) is called $\Sigma-$implementable if it belongs to

$$\mathcal{M}^{\Sigma} = \left\{ \begin{array}{l} (U, w) \in \mathbb{R}_+ \times \mathbb{R}^m | \, \exists w_0 : \mathbb{R}_+ \to \mathbb{R}_+, \text{ such that} \\ U(\delta, r) = \underset{\tilde{w} \in \mathbb{R}^m}{max} \left[w_0(\tilde{w}'\Sigma\tilde{w}) + \frac{1}{2}\tilde{w}'(\delta BB' - r\Sigma)\tilde{w} \right] \text{ and} \\ w(\delta, r) = \underset{\tilde{w} \in \mathbb{R}^m}{argmax} \left[w_0(\tilde{w}'\Sigma\tilde{w}) + \frac{1}{2}\tilde{w}'(\delta BB' - r\Sigma)\tilde{w} \right] \end{array} \right\}. \quad (37)$$

Let

$$\mathcal{M} = \left\{ \begin{array}{l} (U, w) \in \mathbb{R}_+ \times \mathbb{R}^m | \, \exists w_0 : \mathbb{R}^m \to \mathbb{R}_+, \text{ such that} \\ U(\delta, r) = \underset{\tilde{w} \in \mathbb{R}^m}{max} \left[w_0(\tilde{w}) + \frac{1}{2}\tilde{w}'(\delta BB' - r\Sigma)\tilde{w} \right] \text{ and} \\ w(\delta, r) = \underset{\tilde{w} \in \mathbb{R}^m}{argmax} \left[w_0(\tilde{w}) + \frac{1}{2}\tilde{w}'(\delta BB' - r\Sigma)\tilde{w} \right] \end{array} \right\} \quad (38)$$

be the set of implementable allocations. It is obvious that a $\Sigma-$implementable mechanism is implementable but it is not true vice versa: $\mathcal{M}^{\Sigma} \subset \mathcal{M}$. When $(U, w) \in \mathcal{M}^{\Sigma}$, the agent's information rent is

$$\begin{aligned} U(\delta, r) &= \underset{w \in \mathbb{R}^m}{max} \left\{ w_0(w'\Sigma w) + \frac{1}{2}w'[\delta BB' - r\Sigma]w \right\} \\ &= \underset{x}{max} \underset{w: w'\Sigma w = x^2}{max} \left\{ w_0(w'\Sigma w) + \frac{1}{2}w'[\delta BB' - r\Sigma]w \right\} \\ &= \underset{x}{max} \left[w_0(x^2) + \frac{1}{2}\left(\delta \underset{w'\Sigma w = x^2}{max} \frac{w'BB'w}{w'\Sigma w} - r \right) x^2 \right] \\ &= \underset{x}{max} \left[w_0(x^2) + \frac{1}{2}\vartheta_1 x^2 \right] \\ &\equiv u(\vartheta_1), \end{aligned} \quad (39)$$

where $\vartheta_1 = \delta\lambda_1 - r$,

$$\lambda_1 = \underset{w'\Sigma w = x^2}{max} \frac{w'BB'w}{w'\Sigma w} = \lambda_1(\Sigma^{-1/2}BB'\Sigma^{-1/2}) = \lambda_1(BB'\Sigma^{-1}) \quad (40)$$

is the first (largest) eigenvalue of matrix $BB'\Sigma^{-1}$. The corresponding set of optimal wages for the agent is

$$\mathcal{W}(x) = \left\{ w \in \mathbb{R}^m | \, \Sigma^{1/2} w \in \mathcal{N}(\Sigma^{-1/2} BB' \Sigma^{-1/2} - \lambda_1 I), w'\Sigma w = x^2 \right\}, \qquad (41)$$

where $\mathcal{N}(\Sigma^{-1/2}BB'\Sigma^{-1/2} - \lambda_1 I)$ denotes the eigenspace of matrix $\Sigma^{-1/2}BB'\Sigma^{-1/2}$ corresponding to λ_1. (See Lemma **??** in appendix for detailed discussion.) As discussed in previous sections, (**??**) implies the envelop condition $u'(\vartheta_1) = \frac{1}{2}x^2$ and the convexity of $u(\vartheta_1)$ in ϑ_1, which is conversely sufficient for the implementability of contract. We further assume that $\vartheta_2 = \delta\lambda_1 + r$. Then

$$\begin{aligned} \delta &= \frac{\vartheta_1 + \vartheta_2}{2\lambda_1} \\ r &= \frac{\vartheta_2 - \vartheta_1}{2}. \end{aligned}$$

The principal's optimization problem is thus formulated as:

$$\begin{aligned} & \max_x \iint\limits_{\Theta} \left[\frac{\vartheta_1 + \vartheta_2}{2\lambda_1} \max_{w \in \mathcal{W}(x)} w'B\beta - \frac{1}{2}\vartheta_2 x^2 - u(\vartheta_1) \right] \psi(\vartheta_1, \vartheta_2) \mathrm{d}\vartheta_1 \mathrm{d}\vartheta_2 \\ & \qquad s.t. : u'(\vartheta_1) = \frac{1}{2}x^2, u(\cdot) \text{ is a convex function, } u(\vartheta_1) \geqslant 0, \end{aligned} \qquad (42)$$

where $\psi(\vartheta_1, \vartheta_2) \equiv f\left(\frac{\vartheta_1+\vartheta_2}{2\lambda_1}, \frac{\vartheta_2-\vartheta_1}{2}\right) \frac{1}{2\lambda_1}$ represents the joint density of $(\vartheta_1, \vartheta_2)$.

$$\Theta = \left\{ (\vartheta_1, \vartheta_2) | \, \frac{(\vartheta_1 + \vartheta_2)}{2\lambda_1} \in [\underline{\delta}, \overline{\delta}], \frac{(\vartheta_2 - \vartheta_1)}{2} \in [\underline{r}, \overline{r}] \right\}$$

denotes the region of transformed variables. Letting $y = \Sigma^{1/2}w$, the embedded program $\max_{w \in \mathcal{W}(x)} w'B\beta$ can be expressed as

$$\max_y y'\Sigma^{-1/2}B\beta, s.t. : y'y = x^2, y \in \mathcal{N}\left(\Sigma^{-1/2}BB'\Sigma^{-1/2} - \lambda_1 I\right).$$

Applying Lemma **??** in the Appendix, we get the maxima and maximized value of this program

$$\begin{aligned} y^* &= x \frac{Q_k Q_k' \Sigma^{-1/2} B\beta}{\sqrt{\beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B\beta}} \\ \Pi^* &= x\sqrt{\beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B\beta}. \end{aligned}$$

Q_kQ_k' is the spectral projector of matrix $\Sigma^{-1/2}BB'\Sigma^{-1/2}$ corresponding to the first eigenvalue λ_1. Following from the spectral representation theorem in linear algebra, Q_kQ_k' is unique although Q_k is usually not. (See Lemma **??** in the appendix for detailed discussion.)

The maxima to the original program $\max_{w \in \mathcal{W}(x)} w'B\beta$ is therefore

$$w^* = x \frac{\Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B\beta}{\sqrt{\beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B\beta}}. \qquad (43)$$

Substituting this expression into (**??**), we can rewrite the optimization problem of the principal as

$$\begin{aligned}\max_{u,x} \iint_{\Theta} \left[\frac{\vartheta_1+\vartheta_2}{2\lambda_1}\sqrt{\beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B\beta}\, x - \frac{1}{2}\vartheta_2 x^2 - u(\vartheta_1)\right]\psi(\vartheta_1,\vartheta_2)\mathrm{d}\vartheta_1\mathrm{d}\vartheta_2 \\ s.t.: u'(\vartheta_1) = \frac{1}{2}x^2, u(\vartheta_1) \text{ is a convex function }, u(\vartheta_1) \geqslant 0.\end{aligned} \tag{44}$$

For expositional convenience we define the following notations

$$\begin{aligned}\mu(\vartheta_1) &\equiv \int_{\Theta_2(\vartheta_1)} (\vartheta_1+\vartheta_2)\psi(\vartheta_1,\vartheta_2)\mathrm{d}\vartheta_2 \\ \varrho(\vartheta_1) &\equiv \int_{\Theta_2(\vartheta_1)} \vartheta_2\psi(\vartheta_1,\vartheta_2)\mathrm{d}\vartheta_2 \\ \psi_1(\vartheta_1) &\equiv \int_{\Theta_2(\vartheta_1)} \psi(\vartheta_1,\vartheta_2)\mathrm{d}\vartheta_2 \\ \Psi_1(\vartheta_1) &\equiv \int_{\underline{\vartheta}_1}^{\vartheta_1} \psi_1(s)ds \\ \mathcal{H}(\vartheta_1) &\equiv \frac{1-\Psi_1(\vartheta_1)}{\psi_1(\vartheta_1)},\end{aligned}$$

where $\Theta_2(\vartheta_1) = \{\vartheta_2 \in \mathbb{R}_+ | (\vartheta_1,\vartheta_2) \in \Theta\}$, $\underline{\vartheta}_1 = \lambda_1\underline{\delta} - \overline{r}$, and make the following two assumptions:

Assumption 3 *$\mathcal{H}(\vartheta_1)$ is decreasing in ϑ_1.*

Assumption 4 *$\lambda_1 \leqslant \frac{\sigma_r}{\sigma_\delta}$, σ_r and σ_δ are respectively standard deviations of δ and r.*

Integrating with respect to ϑ_2, the above optimization can be simplified to a standard one-dimensional screening problem:

$$\begin{aligned}\max_{u,x} \int_{\Theta_1} \left[x\frac{\mu(\vartheta_1)}{2\lambda_1}\sqrt{\beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B\beta} - \frac{x^2}{2}\varrho(\vartheta_1) - u(\vartheta_1)\psi_1(\vartheta_1)\right]\mathrm{d}\vartheta_1 \\ s.t.: u'(\vartheta_1) = \frac{x^2}{2}, u(\vartheta_1) \text{ is a convex function }, u(\vartheta_1) \geqslant 0.\end{aligned} \tag{45}$$

Ignoring for a while the convexity condition and applying the standard technique, we obtain the solution to the relaxed problem:

$$\begin{aligned}x^*(\vartheta_1) &= \frac{\sqrt{\beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B\beta}}{2\lambda_1}\frac{\mu(\vartheta_1)}{\varrho(\vartheta_1)+[1-\Psi_1(\vartheta_1)]} \\ &= \frac{\sqrt{\beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B\beta}}{2\lambda_1}\frac{\vartheta_1 + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}{\mathcal{H}(\vartheta_1)+E_{\vartheta_2}(\vartheta_2|\vartheta_1)}.\end{aligned} \tag{46}$$

We now need to verify the convexity of $u(\cdot)$, which is equivalent to say that $x(\cdot)$ is an increasing function. It holds provided that (i) Assumption **??** is satisfied, (ii) $\xi(\vartheta_1) \equiv$

$E_{\vartheta_2}(\vartheta_2|\vartheta_1)$ is decreasing in ϑ_1. Condition (ii) is satisfied if and only if Assumption **??** holds because $Cov(\vartheta_1, \vartheta_2) = \lambda_1^2\sigma_\delta^2 - \sigma_r^2 \leqslant 0$. Substituting (**??**) into (**??**) we get the optimal wage

$$w^*(\vartheta_1) = \frac{1}{2\lambda_1}\frac{\vartheta_1 + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}{\mathcal{H}(\vartheta_1) + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}\Sigma^{-1/2}Q_kQ_k'\Sigma^{-1/2}B\beta. \tag{47}$$

The information rent accrued to the agent and surplus of the principal are also easily obtained:

$$u^*(\vartheta_1) = \frac{\beta'B'\Sigma^{-1/2}Q_kQ_k'\Sigma^{-1/2}B\beta}{8\lambda_1^2}\int_{\underline{\vartheta}_1}^{\vartheta_1}\left(\frac{\vartheta_1 + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}{\mathcal{H}(\vartheta_1) + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right)^2 \mathrm{d}\vartheta_1 \tag{48}$$

$$\Pi^* = \frac{1}{8\lambda_1^2}E_{\vartheta_1}\left[\frac{(\vartheta_1 + E_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right]\beta'B'\Sigma^{-1/2}Q_kQ_k'\Sigma^{-1/2}B\beta. \tag{49}$$

The above analysis can be summarized in the following proposition.

Proposition 4 *Suppose that Assumptions* **??** *and* **??** *are satisfied. Then the* $\Sigma-$*implementable allocations are given by (***??***) and (***??***), and the resulting surplus is given by (***??***).*

In the original model, the contract $\{w_0(\delta, r), w(\delta, r)\}$ is implementable if for all $(\delta, \hat{\delta}, r, \hat{r}) \in [\underline{\delta}, \overline{\delta}]^2 \times [\underline{r}, \overline{r}]^2$, the following incentive compatibility condition is satisfied:

$$w_0(\delta, r) + \frac{1}{2}w(\delta, r)'(\delta BB' - r\Sigma)w(\delta, r) \geqslant w_0(\hat{\delta}, \hat{r}) + \frac{1}{2}w(\hat{\delta}, \hat{r})'(\delta BB' - r\Sigma)w(\hat{\delta}, \hat{r}). \tag{50}$$

Let

$$U(\delta, r) \equiv w_0(\delta, r) + \frac{1}{2}w(\delta, r)'(\delta BB' - r\Sigma)w(\delta, r)$$

and

$$U(\hat{\delta}, \hat{r}; \delta, r) \equiv w_0(\hat{\delta}, \hat{r}) + \frac{1}{2}w(\hat{\delta}, \hat{r})'(\delta BB' - r\Sigma)w(\hat{\delta}, \hat{r}).$$

Then $\{U(\delta, r), w(\delta, r)\}$ is implementable if

$$U(\delta, r) = \underset{(\hat{\delta}, \hat{r}) \in [\underline{\delta}, \overline{\delta}] \times [\underline{r}, \overline{r}]}{max}\left\{w_0(\hat{\delta}, \hat{r}) + \frac{1}{2}w(\hat{\delta}, \hat{r})'[\delta BB' - r\Sigma]w(\hat{\delta}, \hat{r})\right\} \tag{51}$$

Applying "taxation principle", it could be equivalently represented as:

$$U(\delta, r) = \underset{w \in \mathbb{R}^m}{max}\left\{w_0(w) + \frac{1}{2}w'[\delta BB' - r\Sigma]w\right\}.$$

It implies that (i) the envelop conditions $\frac{\partial U}{\partial \delta} = \frac{1}{2}w'BB'w$, $\frac{\partial U}{\partial r} = -\frac{1}{2}w'\Sigma'w$ hold; (ii) $U(\delta, r)$ is convex in (δ, r). Conversely, given the envelop and convexity conditions, we have the following incentive compatibility condition:

$$\begin{aligned} U(\delta, r) &\geqslant U(\hat{\delta}, \hat{r}) + (\delta - \hat{\delta})\frac{\partial U}{\partial \delta} + (r - \hat{r})\frac{\partial U}{\partial r} \\ &= U(\hat{\delta}, \hat{r}) + \frac{1}{2}(\delta - \hat{\delta})w(\hat{\delta}, \hat{r})'BB'w(\hat{\delta}, \hat{r}) - \frac{1}{2}(r - r')w(\hat{\delta}, \hat{r})'\Sigma w(\hat{\delta}, \hat{r}) \\ &= U(\hat{\delta}, \hat{r}, \delta, r). \end{aligned}$$

$U(\delta, r)$ and $w(\delta, r)$ are therefore implementable. Thus the principal's optimization problem is

$$\begin{aligned}
\max_{U,w} \iint_D & \left[\delta w' B\beta - \frac{1}{2} w'(\delta BB' + r\Sigma)w - U(\delta, r)\right] \mathrm{d}\delta \mathrm{d}r \\
s.t. : & \frac{\partial U}{\partial \delta} = \frac{1}{2} w' BB' w, \\
& \frac{\partial U}{\partial r} = -\frac{1}{2} w' \Sigma' w, \\
& U(\delta, r) \geqslant 0, \\
& U(\delta, r) \text{ is convex.}
\end{aligned}$$

Ignoring momentarily the convexity condition, the principal's relaxed problem could be regarded as an optimal control problem with multiple controls and double-fold integrals. The generalized Hamiltonian approach offered by Basov (2005) is applicable to this problem. His method however ensures the existence of solution to the relaxed problem rather than offers a feasible way for getting it. One often has to rely upon the numerical techniques to get solution from a system of partial differential equations. A more serious drawback of his approach is that the solution to the relaxed problem usually cannot solve the complete problem because the convexity condition could only be satisfied by accident. In fact the envelop and convexity conditions require that the vector field $(\frac{1}{2}w'BB'w, -\frac{1}{2}w'\Sigma'w)$ has a convex potential function. This puts severe restrictions on the set of implementable wages and makes the multidimensional problem much more complex than its unidimensional counterpart because the latter requires only that $\frac{1}{2}w'BB'w$ or $-\frac{1}{2}w'\Sigma w$ has a convex primitive function.

In order to get an explicit solution to the complete problem, we therefore sacrifice some of the principal's degrees of freedom by restricting our attention in the set of $\Sigma-$implementable allocations $\mathcal{M}^{\Sigma}$. We decompose the information contained in vector w into two aspects: its $\Sigma-$norm ($\sqrt{w'\Sigma w} = x$) and direction. Meanwhile, the type vector (δ, r) is transformed linearly to $(\vartheta_1, \vartheta_2)$. Notice that, the $\Sigma-$norm of wage vector depends only on ϑ_1, while its direction is at free disposal of the agent and depends on neither ϑ_1 nor ϑ_2. Our $\Sigma-$norm-based assumption on the base wage w_0 limits greatly the authority of the principal since he now has only the discretion to choose x contingent on the agent's report $\hat{\vartheta}_1$. The authority of the agent, on the contrary, is augmented since he is vested the authority of choosing the direction of w. Thus, this procedure is virtually a process of delegating part of the principal's authority to the agent. Under the assumptions we made, the multidimensional mechanism problem is solved with the same amount of computational work as in the one-dimensional screening problem after performing integration with respect to the irrelevant variable ϑ_2.

In a special case of orthogonal performance measurement system, we have the following corollary.

Corollary 4 *For an orthogonal performance measurement system, there is no incentive in the performance measures with non-largest signal-noise ratio.*

Proof. See appendix. ■

As mentioned above, the wage vector is determined by two aspects: its overall intensity ($\Sigma - norm$) and relative allocation among performance measures (direction). In our dimensionality-reducing procedure, the authority of choosing relative allocation is delegated to the agent. Then for an orthogonal system in which performance measures are totally independent to each other, the agent inclines to allocate the overall intensity to the measures with larger sensitivity (measured by $\|b_i\|^2$) and smaller randomness (measured by σ_i^2). Therefore he will put the overall intensity of incentives on the measures with largest signal-noise ratio $\|b_i\|^2/\sigma_i^2$, and the measures with non-largest signal-noise ratios will be assigned zero incentive.

Holmstrom and Milgrom (1990) show that missing incentive clauses are commonly observed in practice, even when good, objective output measures are available and agents are highly responsive to incentive pay. In their model, there exist multiple performance measures with varying degrees of accuracy (the tasks and performance measures are one-to-one corresponding to each other, that is, $B = I$), and the tasks are substitute to each other. In this setup, employees will concentrate their attention (effort) on improving the performance measure tied to high compensation, to the exclusion of hard-to-measure or even non-observable but important tasks. Therefore an optimal incentive contract can be to pay a fixed wage independent of measured performance. Our Corollary **??** offers a different explanation to the missing incentive phenomenon. Notice that in this corollary, we assume the performance measures are orthogonal to each other, which is quite different to the substitute condition required by Holmstrom and Milgrom's paper.

The following corollary provides a comparison of surpluses obtained using two performance measurement systems with the same largest signal-noise ratios.

Corollary 5 *If two orthogonal performance measurement systems* $\mathcal{P}_1 \equiv (B_1, \Sigma_{11})$ *and* $\mathcal{P}_2 \equiv (B_2, \Sigma_{22})$ *are such that matrices* $\Sigma_{11}^{-1/2} B_1 B_1' \Sigma_{11}^{-1/2}$ *and* $\Sigma_{22}^{-1/2} B_2 B_2' \Sigma_{22}^{-1/2}$ *have the same first eigenvalues* λ_1 *and the multiplicities of* λ_1 *in these two matrices are, respectively,* k_1 *and* k_2*, then* $\Pi^*(\mathcal{P}_1) \geqslant \Pi^*(\mathcal{P}_2)$ *if and only if the sum of squares of congruences of the first* k_1 *performance measures in* $\mathcal{P}_1$ *is larger than that of the first* k_2 *performance measures in* $\mathcal{P}_2$*,i.e.,* $\sum_{i=1}^{k_1} \Upsilon_{i1}^2 \geqslant \sum_{i=1}^{k_2} \Upsilon_{i2}^2$.

Proof. See appendix. ■

We next discuss the value of additional performance measures to an existing set. Let $\mathcal{P}_1 = (B_1, \Sigma_{11})$ represent a performance measurement system that reports m_1 measures and let

$$\mathcal{P} \equiv (B, \Sigma) = \left[\begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right]$$

represent a system that reports an additional m_2 measures $\mathcal{P}_2 = (B_2, \Sigma_{22})$. $\mathcal{P}_1$ and $\mathcal{P}_2$ are supposed to be orthogonal to each other. That is to say, $\Sigma_{12} = \mathbf{0}, \Sigma_{21} = \mathbf{0}, B_1'B_2 = \mathbf{0}, B_2'B_1 = \mathbf{0}$. Denote the set of eigenvalues of $\Sigma_{11}^{-1/2}B_1B_1'\Sigma_{11}^{-1/2}$ and $\Sigma_{22}^{-1/2}B_2B_2'\Sigma_{22}^{-1/2}$, respectively, by $\lambda_i, i = 1, \cdots m_1$ and $\mu_j, j = 1, \cdots, m_2$. The first eigenvalues $\lambda_1 = max_{1 \leqslant i \leqslant m_1}\lambda_i$ and $\mu_1 = max_{1 \leqslant j \leqslant m_2}\mu_j$ have multiplicities k_1 and k_2 respectively. The following corollary provides a specification of the incremental expected value of the additional performance measures provided by $\mathcal{P}$.

Corollary 6 *If* $\lambda_1 > \mu_1$, *then* $\Pi^*(\mathcal{P}) = \Pi^*(\mathcal{P}_1)$; *if* $\lambda_1 = \mu_1$, *then* $\Pi^*(\mathcal{P}) = \Pi^*(\mathcal{P}_1) + \Pi^*(\mathcal{P}_2) > \Pi^*(\mathcal{P}_1)$; *if* $\lambda_1 < \mu_1$, *then* $\Pi^*(\mathcal{P}) = \Pi^*(\mathcal{P}_2)$.

Proof. See appendix. ■

In the environment of complete information (with observable costs and risk aversion), Feltham and Xie (1994) show that the incremental value of additional performance measures is always non-negative because the principal can always assign zero incentive to the additional measures. In this case, the principal's surplus obtained using the original performance system $\mathcal{P}_1 = (B_1, \Sigma_{11})$ is

$$\pi^{fb}(\mathcal{P}_1) = \frac{\delta}{2}\beta' B_1' \left(B_1B_1' + \frac{r}{\delta}\Sigma_{11} \right)^{-1} B_1\beta;$$

the surplus obtained using the augmented performance measurement system

$$\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2) = \left[\begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right]$$

is

$$\pi^{fb}(\mathcal{P}_1) = \frac{\delta}{2}\beta' B' \left(BB' + \frac{r}{\delta}\Sigma \right)^{-1} B\beta.$$

The incremental value of additional performance measures is thus:

$$\Delta\pi = \pi^{fb}(\mathcal{P}) - \pi^{fb}(\mathcal{P}_1) = \frac{\delta}{2}\beta'(D - D_1)\beta,$$

where

$$
\begin{aligned}
D_1 &= B_1' \left(B_1 B_1' + \frac{r}{\delta}\Sigma\right)^{-1} B_1 \\
D &= B' \left(BB' + \frac{r}{\delta}\Sigma\right)^{-1} B \\
&= (B_1', B_2') \begin{pmatrix} B_1 B_1' + \frac{r}{\delta}\Sigma_{11} & B_1 B_2' + \frac{r}{\delta}\Sigma_{12} \\ B_2 B_1' + \frac{r}{\delta}\Sigma_{21} & B_2 B_2' + \frac{r}{\delta}\Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \\
&= (B_1', B_2') \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}^{-1} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \\
&= (B_1', B_2') \begin{pmatrix} H_{11}^{-1} + H_{11}^{-1} H_{12} H_{22\cdot 1}^{-1} H_{21} H_{11}^{-1} & -H_{11}^{-1} H_{12} H_{22\cdot 1}^{-1} \\ -H_{22\cdot 1}^{-1} H_{21} H_{11}^{-1} & H_{22\cdot 1}^{-1} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \\
&= B_1' H_{11}^{-1} B_1 + B_1' H_{11}^{-1} H_{12} H_{22\cdot 1}^{-1} H_{21} H_{11} B_1 - B_1' H_{11} H_{12} H_{22\cdot 1}^{-1} B_2 \\
&\quad - B_2' H_{22\cdot 1}^{-1} H_{21} H_{11}^{-1} B_1 + B_2' H_{22\cdot 1}^{-1} B_2 \\
H_{ij} &= B_i B_j' + \frac{r}{\delta}\Sigma_{ij}, \\
H_{22\cdot 1} &= H_{22} - H_{21} H_{11}^{-1} H_{12}.
\end{aligned}
$$

It follows that $D - D_1 = B_1' H_{11}^{-1} H_{12} H_{22\cdot 1}^{-1} H_{21} H_{11} B_1 - B_1' H_{11} H_{12} H_{22\cdot 1}^{-1} B_2 - B_2' H_{22\cdot 1}^{-1} H_{21} H_{11}^{-1} B_1 + B_2' H_{22\cdot 1}^{-1} B_2 = [B_1' H_{11}^{-1} H_{12} - B_2'] H_{22\cdot 1}^{-1} [H_{21} H_{11}^{-1} B_1 - B_2]$ is a semi-positive definite matrix. It in turn implies that $\Delta\pi = \frac{\delta}{2}\beta'(D - D_1)\beta \geqslant 0$. As a special case, if $\mathcal{P}_1$ is orthogonal to $\mathcal{P}_2$, $D - D_1 = B_2' H_{22}^{-1} B_2$; therefore, $\Delta\pi = \frac{\delta}{2}\beta' B_2 \left(B_2 B_2' + \frac{r}{\delta}\Sigma_{22}\right)^{-1} B_2 \beta = \pi^{fb}(\mathcal{P}_2) \geqslant 0$.

The incremental value then is zero if and only if the measures provided by the original performance measurement system are a sufficient statistic for the measures provided by the augmented system, with respect to the agent's effort. According to this result, adding a performance measurement system which is orthogonal to the original one will increase the surplus for sure. Our result, on the contrary, states that the incremental value is zero if $\lambda_1 > \mu_1$; is positive if $\lambda_1 = \mu_1$; is ambiguous if $\lambda_1 < \mu_1$. These new results come from the assumption that w_0 is based on the $\Sigma - norm$ of w. Under this assumption, the performance measures associated with non-largest eigenvalues are in fact redundant, and the incremental value is therefore determined by the first eigenvalues of the original and new performance measurement systems.

6 Conclusion

In this paper, we explain the phenomenon of low-powered incentives from a new perspective. We consider a case where the agent possesses private information about his own risk aversion and the cost of efforts. Besides the rents eliciting the agent's efforts, the principal has to give up some additional information rents to the agent in order to elicit his truthtelling. She has to consider two tradeoffs when choosing the optimal incentive contract. One is the tradeoff between insurance and incentives; the other is the tradeoff between efficiency and rent extraction. The former is the fundamental issue in moral hazard problem; while the latter lies in the core of adverse selection problem. These two "tradeoffs" together lead to lower-powered incentives. We further show that in the presence of mere unobservable risk aversion or cost, the second-best incentive contract is flatter than the first-best one. In the case with multidimensionally asymmetric information, we first assume that the deterministic and stochastic components of a performance measurement system vary in a similar way. Under this assumption, the agent's private information relevant to his decision making is captured in a single scalar variable. The power of incentive is lower than that in the case where this scalar variable is observable. Furthermore, we reduce the complexity of computation by delegating part of the principal's decision-making authority to the agent. In this setup, we find that most performance measures are redundant and are compensated by fixed wage. This provides a new explanation to the frequently-observed phenomenon of missing incentive.

References

[1] Armstrong, M. (1996) "Multiproduct Nonlinear Pricing", *Econometrica*, 64: 51-75

[2] Baker, G. 1992. "Incentive Contracts and Performance Measurement", *Journal of Political Economy* 100, 598-614

[3] Baker, G.P. (2002). "Distortion and Risk in Optimal Incentive Contracts", *Journal of Human Recources*, 37, 728-751.

[4] Basov, S., (2001) "Hamiltonian Approach to Multi-dimensional Screening", *Journal of Mathematical Economics*, 36: 77-94.

[5] Feltham G. A., and J. Xie. "Performance Measure Congruity and Diversity in Multi-Task Principal-Agent Relations", *The Accounting Review*, 1994, 69: 429-453

[6] Frant, H. (1996) "High-Powered and Low-Powered Incentives in the Public Sector", *Journal of Public Administration Research and Theory* 6:365-82.

[7] Guesnerie, R. (1981): "On Taxation and Incentives: Further Reflections on the Limits of Redistribution." Discussion Paper No.89, Sonderforschungsbereich 21, University of Bonn.

[8] Green, J., and Stokey, N. (1983), "A Comparison of Tournaments and Contracts", *Journal of Political Economy* 91: 349-364.

[9] Hurwicz, L., 1972, On Informationally Decentralized Systems, in: Radner, R. and C. B. McGuire, eds., *Decision and Organization, Volume in Honor of J. Marschak (North-Holland)* 297-336.

[10] Hurwicz L (1979), "On the Dimension Requirements of Informationally Decentralized Pareto-satisfactory Processes", In: *Arrow K, Hurwicz L (eds) Studies in Resource Allocation Processes*, Cambridge University Press

[11] Hurwicz L (1986), "On Informational Decentralization and Efficiency in Resource Allocation Mechanism", In: Reiter, S. (ed) , *Studies in Mathematical Economics*, Mathematical Association of America

[12] Hammond, P. (1979), "Straightforward Individual Incentive Compatibility in Large Economies", *Review of Economic Studies*, 46:263-282

[13] Holmstrom, B. and P. Milgrom (1991), "Multitask Principal-Agent Analyses", *Journal of Law, Economics, and Organization* 7: 24-52

[14] Lazear, E. and Rosen, S. 1981, "Rank-Order Tournaments as Optimum Labor Contracts", *Journal of Political Economy*, 89: 841-864.

[15] Mount, K. and Reiter,S. (1974) "Informational Size of Message Spaces", *Journal of Economic Theory* 8:161-191

[16] Nalebuff, B., and Stiglitz, J. (1983), "Prices and Incentives: Towards a General Theory of Compensation and Competition", *Bell Journal of Economics* 14: 21-43.

[17] Osana H (1978), "On the Informational Size of Message Spaces for Resource Allocation Process", *Journal of Economic Theory*, 17:66-78

[18] Rochet, J.-C. (1985), "The Taxation Principle and Multi-time Hamilton-Jacobi Equations", *Journal of Mathematical Economics*, 14, 113-128

[19] Rochet, J.-C., and P. Choné (1998), "Ironing, Sweeping, and Multidimensional Screening", *Econometrica* 66: 783-826.

[20] Rochet, J.-C. (1987), "A Necessary and Sufficient Condition for Rationalizability in a Quasi-linear Context", *Journal of Mathematical Economics* 16: 190-200.

[21] Tian, G (1994), "On Informational Efficiency and Incentive Aspects of Generalized Ratio Equilibria", *Journal of Mathematical Economics*, 23:323-337

[22] Tian, G (2004), "A Unique Informationally Efficient Allocation Mechanism in Economies with Consumption Externalities," *International Economic Review*, 45:79-111

[23] Tian, G (2006), "The Unique Informational Efficiency of the Competitive Mechanism in Economies with Production," *Social Choice and Welfare*, 26, 155-182.

[24] Thiele, V (2010) "Task-Specific Abilities in Multi-Task Principal-Agent Relationships", *Labour Economics*, 17(4), 690-698.

[25] Walker, M (1977), "On the Informational Size of Message Spaces", *Journal of Economic Theory*, 15:366-375

[26] Williamson, O. (1985) "The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting", New York: Free Press.

Appendix A.

Proof of Proposition ??. Using the envelop condition $U'(r) = -\frac{1}{2}w'\Sigma w$, the participation constraint $U(r) \geqslant 0$ simplifies to $U(\bar{r}) \geqslant 0$. Incentive compatibility implies that only the participation constraint of the most risk averse type can be binding, i.e., $U(\bar{r}) = 0$. We therefore get

$$U(r) = \int_r^{\bar{r}} \frac{1}{2} w(\tilde{r})' \Sigma w(\tilde{r}) dr. \tag{A.1}$$

The principal's objective function becomes

$$\Pi = \int_{\underline{r}}^{\bar{r}} \left\{ \beta' C^{-1} B' w(r) - \frac{1}{2} w(r)' \left[BC^{-1}B' + r\Sigma \right] w(r) - \int_r^{\bar{r}} \frac{1}{2} w(\tilde{r})' \Sigma w(\tilde{r}) dr \right\} f(r) \mathrm{d}r$$

which, by an integration of parts, gives

$$\int_{\underline{r}}^{\bar{r}} \left\{ \beta' C^{-1} B' w(r) - \frac{1}{2} w(r)' \left[BC^{-1}B' + \left(r + \frac{F(r)}{f(r)} \right) \Sigma \right] w(r) \right\} f(r) \mathrm{d}r.$$

Maximizing pointwise the above expression, we get

$$w^{sb}(r) = \left[BC^{-1}B' + \Phi(r)\Sigma \right]^{-1} BC^{-1}\beta$$

and

$$w_0^{sb}(r) = \frac{1}{2} \int_r^{\bar{r}} w^{sb}(\tilde{r})' \Sigma w^{sb}(\tilde{r}) d\tilde{r} - \frac{1}{2} w^{sb}(r)' \left[BC^{-1}B' - r\Sigma \right] w^{sb}(r).$$

The only work left is to verify the convexity of $U(r)$. Notice that

$$U''(r) = -(D_r w^{sb})' \Sigma w^{sb} = \Phi'(r) w^{sb}(r)' \Sigma \left[BC^{-1}B' + \Phi(r)\Sigma \right]^{-1} \Sigma w^{sb}(r).$$

The second equality comes from the fact that the derivative of w^{sb} with respect to r is[6]

$$\begin{aligned} D_r w^{sb} &= -\left[BC^{-1}B' + \Phi(r)\Sigma \right]^{-1} \Phi'(r)\Sigma \left[BC^{-1}B' + \Phi(r)\Sigma \right]^{-1} BC^{-1}\beta \\ &= -\Phi'(r) \left[BC^{-1}B' + \Phi(r)\Sigma \right]^{-1} \Sigma w^{sb}. \end{aligned}$$

It is clear that $U''(r) \geqslant 0$ because $\Phi'(r) \geqslant 0$ and the matrix $\Sigma \left[BC^{-1}B' + \Phi(r)\Sigma \right]^{-1} \Sigma$ is positive definite. The proof is completed.

Proof of Proposition ??. Using integration by parts, we get

$$\int_{\underline{\delta}}^{\bar{\delta}} U(\delta) g(\delta) = \int_{\underline{\delta}}^{\bar{\delta}} \left[\frac{1 - G(\delta)}{g(\delta)} \right] \frac{w'BB'w}{2} dG(\delta).$$

Substituting it into the expression of the principal's expected surplus and optimizing it with respect to w, we get the second-best performance wage $w^{sb}(\delta)$, and $w_0^{sb}(\delta)$ is also easily

[6]Let A be a nonsingular, $m \times m$ matrix whose elements are functions of the scalar parameter α, then

$$\frac{\partial A^{-1}}{\partial \alpha} = -A^{-1} \frac{\partial A}{\partial \alpha} A^{-1}.$$

obtained. We now check the convexity of $U(\delta)$. The first order derivative of $w^{sb}(\delta)$ is

$$
\begin{aligned}
D_\delta w^{sb}(\delta) &= -\left[H(\delta)BB' + \frac{r}{\delta}\Sigma\right]^{-1}\left[H'(\delta)BB' - \frac{r\Sigma}{\delta^2}\right]\left[H(\delta)BB' + \frac{r\Sigma}{\delta}\right]^{-1} B\beta \\
&= -\left[H(\delta)BB' + \frac{r\Sigma}{\delta}\right]^{-1}\left[H'(\delta)BB' - \frac{r\Sigma}{\delta^2}\right] w^{sb}(\delta) \\
&= -\left[H(\delta)BB' + \frac{r\Sigma}{\delta}\right]^{-1}\left\{-\frac{H(\delta)}{\delta}BB' - \frac{r\Sigma}{\delta^2} + \left[\frac{H(\delta)}{\delta} + H'(\delta)\right]BB'\right\} w^{sb}(\delta) \\
&= \frac{1}{\delta}\left\{(BB')^{-1} - \left[H(\delta)BB' + \frac{r\Sigma}{\delta}\right]^{-1}[H(\delta) + \delta H'(\delta)]\right\} BB'w^{sb}(\delta).
\end{aligned}
$$

It can be verified that the matrix $\frac{1}{\delta}\left\{(BB')^{-1} - \left[H(\delta)BB' + \frac{r\Sigma}{\delta}\right]^{-1}[H(\delta)+\delta H'(\delta)]\right\}$ is positive definite since $\delta + \frac{1-G(\delta)}{g(\delta)} = \delta H(\delta)$ is decreasing. Therefore

$$
\begin{aligned}
U''(\delta) &= D_\delta w^{sb}(\delta)BB'w^{sb}(\delta) \\
&= \frac{1}{\delta} w^{sb}(\delta)'BB'\left\{(BB')^{-1} - \left[H(\delta)BB' + \frac{r\Sigma}{\delta}\right]^{-1}[H(\delta) + \delta H'(\delta)]\right\} BB'w^{sb}(\delta) \geqslant 0,
\end{aligned}
$$

which implies the convexity of $U(\delta)$.

Lemma A.1 *$E(Y|X)$ is nonincreasing in X if and only if $Cov(X,Y) \leqslant 0$.*

Proof. Because $Cov(X,Y) = E(X)E(Y) - E(XY) = E(X)E[E(Y|X)] - E[E(XY|X)] = E(X)E[E(Y|X)] - E[XE(Y|X)] = Cov[X, E(Y|X)]$, $Cov(X,Y) \leqslant 0$ if and only if $E(Y|X)$ is a nonincreasing function of X.[7] ■

Lemma A.2 *Let A, B be $m \times m$ symmetric matrices and $B > 0$, then*

$$\max_{x \neq 0} \frac{x'Ax}{x'Bx} = \lambda_1(B^{-1/2}AB^{-1/2})$$

and the optimal x satisfies: $B^{-1/2}x \in \mathcal{N}(B^{-1/2}AB^{-1/2} - \lambda_1 I)$.

Proof. Let $\frac{B^{1/2}x}{\sqrt{x'Bx}} = y$. Then

$$\max_{x \neq 0} \frac{x'Ax}{x'Bx} = \max_{\|y\|=1} y'B^{-1/2}AB^{-1/2}y.$$

Since $B^{-1/2}AB^{-1/2}$ is a symmetric matrix, there exists an orthogonal matrix P such that $P'B^{-1/2}AB^{-1/2}P = diag\{\lambda_1, \cdots, \lambda_m\}$. $\lambda_1, \cdots, \lambda_m$ are eigenvalues of $B^{-1/2}AB^{-1/2}$ in descending order; λ_1 has multiplicity k. Let $P'y = z$, then

$$\max_{\|y\|=1} y'B^{-1/2}AB^{-1/2}y = \max_{\|z\|=1} z' diag\{\lambda_1, \cdots, \lambda_m\} z = \max_{1 \leqslant j \leqslant m} \lambda_j.$$

The optimal solution to this problem is $z = (z_1, \cdots, z_k, 0, \cdots, 0)'$ with $\sum_{j=1}^k z_j^2 = 1$. We get $y = Pz = \sum_{i=1}^k z_i p_i$. $p_i, i = 1, \cdots, k$ are eigenvectors associated with $\lambda_1 = \cdots = \lambda_k$, therefore $y \in \mathcal{N}(B^{-1/2}AB^{-1/2} - \lambda_1 I)$, which in turn implies that $B^{-1/2}x \in \mathcal{N}(B^{-1/2}AB^{-1/2} - \lambda_1 I)$. ■

[7]Here we use a result in probability theory: $Cov(\varphi_1(X), \varphi_2(X)) \leqslant 0$ iff $\varphi_1'(X)\varphi_2'(X) \leqslant 0$.

Lemma A.3 *The maxima x^* and maximized value Π^* to program*

$$\max_x : \alpha' x, s.t. : \|x\| = a, x \in \mathcal{N}(A - \lambda I) \tag{A.2}$$

are

(i) *If $\alpha' Q_k Q_k' \alpha \neq 0$*

$$\begin{aligned} x^* &= a \frac{Q_k Q_k' \alpha}{\sqrt{\alpha' Q_k Q_k' \alpha}}, \\ \Pi^* &= a \sqrt{\alpha' Q_k Q_k' \alpha} \end{aligned}$$

(ii) *If $\alpha' Q_k Q_k' \alpha = 0$*

x^* is an arbitrary non-null element in $\mathcal{N}(A - \lambda I)$ with norm a

$\Pi^* = 0$

where λ is an eigenvalue of symmetric matrix A with multiplicity k, $\mathcal{N}(A-\lambda I)$ represents the eigenspace of A associated with λ, $Q_k = (q_1, q_2, \cdots, q_k)$ are a set of orthonormal eigenvectors of A corresponding to λ.

Proof. Since A is a real symmetric matrix, there exists an orthogonal matrix $Q = (q_1, \cdots, q_n) = (Q_k, Q_{-k})$ such that

$$Q' A Q = diag\{\lambda, \cdots \lambda, \lambda_{k+1}, \cdots, \lambda_n\}.$$

$Q_k = (q_1, q_2, \cdots, q_k)$ are a set of orthonormal eigenvectors associated with λ, $Q_{-k} = (q_{k+1}, \cdots, q_n)$ is the set of remaining orthonormal eigenvectors. Applying the spectral decomposition theorem in matrix algebra, the spectral projector matrix $Q_k Q_k'$ is unique although Q_k is in general not unique.

$$(A - \lambda I)x = 0 \Longleftrightarrow Q diag\{0, \cdots, 0, \lambda - \lambda_{k+1}, \cdots, \lambda - \lambda_n\} Q' x = 0$$

Letting $Q' x = y$, we get

$$diag\{0, \cdots, 0, \lambda - \lambda_{k+1}, \cdots, \lambda - \lambda_n\} y = 0,$$

which implies $y_i = 0, i = k+1, \cdots n$. Then the program (**??**) can be rewritten as

$$\max_y : \alpha' Q y, s.t. : \|y\| = a, y = (y_1, \cdots, y_k, 0, \cdots, 0).$$

- If $\alpha' Q_k Q_k' \alpha = 0$, then $Q_k' \alpha = 0$. Therefore,

$$\alpha' Q y = (0, \cdots, 0, \alpha' q_{k+1}, \alpha' q_n) \begin{pmatrix} y_1 \\ \vdots \\ y_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0$$

The maxima to the program (**??**) is therefore an arbitrary non-null vector in $\mathcal{N}(A-\lambda I)$ with norm a.

- If $\alpha' Q_k Q_k' \alpha \neq 0$, it is optimal to choose

$$y^* = \frac{a}{\sqrt{\alpha' Q_k Q_k' \alpha}} (Q_k, \mathbf{0})' \alpha,$$

  the corresponding optimal value is

$$\Pi^* = a\sqrt{\alpha' Q_k Q_k' \alpha}.$$

  The maxima for the original program is therefore

$$x^* = Qy^* = a\frac{Q_k Q_k' \alpha}{\sqrt{\alpha' Q_k Q_k' \alpha}}.$$

■

Lemma A.4 *(Uniqueness of Spectral Representation) A represents an $n \times n$ symmetric matrix, Q represents an $n \times n$ orthogonal matrix, $D = diag\{d_1, \cdots, d_n\}$ is an $n \times n$ diagonal matrix such that $Q'AQ = D$.(Note that every real symmetric matrix is orthogonally diagonalizable.) The ith columns of Q are $\mathbf{q}_i, i = 1, \cdots, n$,respectively. $\lambda_1, \cdots, \lambda_k$ represent the distinct eigenvalues of A, $\nu_1, \cdots \nu_k$ represent the (algebraic or geometric) multiplicities of $\lambda_1, \cdots, \lambda_k$, respectively. For $j = 1, \cdots, k$, $S_j = \{i : d_i = \lambda_j\}$ represent the set comprising the ν_j values of i such that $d_i = \lambda_j$. Then A can be expressed uniquely (aside from the ordering of the terms) as*

$$A = \sum_{j=1}^{k} \lambda_j E_j \tag{A.3}$$

where (for $j = 1, \cdots, k$) $E_j = \sum_{i \in S_j} \mathbf{q}_i \mathbf{q}_i'$, $\mathbf{q}_i, i \in S_j$ are eigenvectors associated with λ_j.

Proof. Suppose that P is an $n \times n$ orthogonal matrix and $D^* = \{d_i\}$ an $n \times n$ diagonal matrix such that $P'AP = D^*$ (where P and D^* are possibly different from Q and D). Further, denote the first,$\cdots nth$ columns of P by $p_1, \cdots, p_n$, respectively, and (for $j = 1, \cdots, k$) let $S_j = \{i : d_i^* = \lambda_j\}$. Then, analogous to the decomposition $A = \sum_{j=1}^{k} \lambda_j E_j$, we have the decomposition

$$A = \sum_{j=1}^{k} \lambda_j F_j$$

where (for $j = 1, \cdots, k$) $F_j = \sum_{j \in S_j^*} p_j p_j'$. Now, for $j = 1, \cdots, k$, let $Q_j = (\mathbf{q}_{i_1}, ..., \mathbf{q}_{i_{\nu_j}}$) and $P_j = (\mathbf{p}_{i_1^*}, \cdots, \mathbf{p}_{i_{\nu_j}^*})$ where $i_1, \cdots, i_{\nu_j}$ and $i_1^*, \cdots, i_{\nu_j}^*$ are the elements of S_j and S_j^* , respectively. Then, $\mathcal{C}(P_j) = \mathcal{N}(A - \lambda_j I) = \mathcal{C}(Q_j)$ (the symbol $\mathcal{C}(A)$ denotes the column space of a matrix A), so that $P_j = Q_j L_j$ for some $\nu_j \times \nu_j$ matrix L_j. Moreover, since clearly $Q_j' Q_j = I_{\nu_j}$ and $P_j' P_j = I_{\nu_j}$,

$$L_j' L_j = L_j' Q_j' Q_j L_j = P_j' P_j = I,$$

implying that L_j is an orthogonal matrix. Thus,

$$F_j = P_j P_j' = Q_j L_j L_j' Q_j' = Q_j I Q_j' = Q_j Q_j' = E_j.$$

We conclude that the decomposition $A = \sum_{j=1}^{k} \lambda_j F_j$ is identical to the decomposition $A = \sum_{j=1}^{k} \lambda_j E_j$, and hence that the decomposition $A = \sum_{j=1}^{k} \lambda_j E_j$ is unique (aside from the ordering of terms). ■

Proof of Corollary ??. Let $\mathcal{P} = (B, \Sigma)$ be an orthogonal performance measurement system with $\Sigma = diag\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_m^2\}$, $BB' = diag\{\|b_1\|^2, \cdots, \|b_m\|^2\}$, then $\Sigma^{-\frac{1}{2}} BB' \Sigma^{-\frac{1}{2}} = diag\{\lambda_1, \cdots, \lambda_m\}$, $\lambda_i = \|b_i\|^2/\sigma_i^2$, $i = 1, \cdots, m$ are eigenvalues in descending order. $\lambda_1 = \lambda_2 = \cdots = \lambda_k > \lambda_{k+1} \geqslant \cdots \geqslant \lambda_m$. Let $\mathbf{p} = (p_1, \cdots, p_m)' \in \mathbb{R}^m$ be the normalized eigenvector associated with λ_1. Then

$$\mathbf{p}' \Sigma^{-1/2} BB' \Sigma^{-1/2} \mathbf{p} = \lambda_1$$

It follows that

$$\lambda_1 \sum_{j=1}^{k} p_j^2 + \sum_{j=k+1}^{m} \lambda_j p_j^2 = \lambda_1 \sum_{j=1}^{m} p_j^2$$

Then we obtain

$$p_j = 0, \forall j = k+1, \cdots m.$$

Therefore we write

$$Q_k = \begin{pmatrix} \tilde{Q}_k \\ \mathbf{0} \end{pmatrix},$$

where $\tilde{Q}_k$ is a $k \times k$ orthogonal matrix. Substituting it into (**??**), we get

$$w^*(\vartheta_1) = \frac{1}{2\lambda_1} \frac{\vartheta_1 + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}{\mathcal{H}(\vartheta_1) + E_{\vartheta_2}(\vartheta_2|\vartheta_1)} \begin{pmatrix} b_1'\beta/\sigma_1^2 \\ \vdots \\ b_k'\beta/\sigma_k^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{A.4}$$

The optimal wages paid for the performance measures associated with the non-largest eigenvalues are zero: $w_i^*(\vartheta_1) = 0$, for all $i = k+1, \cdots, m$.

Proof of Corollary ??. For an orthogonal measurement system $\mathcal{P} = (B, \Sigma)$, eigenvalues of diagonal matrix $\Sigma^{-1/2} BB' \Sigma^{-1/2}$ are in fact the signal-noise ratios of measures in $\mathcal{P}$: $\lambda_i = \frac{b_i' b_i}{\sigma_i^2}$. Suppose that $\mathcal{P}_1$ and $\mathcal{P}_2$ are orthogonal systems with the same first eigenvalues $\lambda_1 = \lambda_1(\Sigma_{11}^{-1/2} B_1 B_1' \Sigma_{11}^{-1/2}) = \lambda_1(\Sigma_{22}^{-1/2} B_2 B_2' \Sigma_{22}^{-1/2})$, the multiplicities of λ_1 are respectively k_1 and k_2 in $\Sigma_{11}^{-1/2} B_1 B_1' \Sigma_{11}^{-1/2}$ and $\Sigma_{22}^{-1/2} B_2 B_2' \Sigma_{22}^{-1/2}$. Then the surplus obtained using system $\mathcal{P}_1$ is

$$\begin{aligned}
\Pi^*(\mathcal{P}_1) &= \kappa(\lambda_1)\beta' B_1' \Sigma_{11}^{-1/2} Q_1 Q_1' \Sigma_{11}^{-1/2} B_1 \beta \\
&= \kappa(\lambda_1)\beta' B_1' \begin{pmatrix} \Lambda_{k_1} & \mathbf{0} \\ \mathbf{0} & \Lambda_{m_1-k_1} \end{pmatrix} \begin{pmatrix} I_{k_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \Lambda_{k_1} & \mathbf{0} \\ \mathbf{0} & \Lambda_{m_1-k_1} \end{pmatrix} B_1 \beta \\
&= \kappa(\lambda_1)\beta' \sum_{i=1}^{k_1} \frac{b_i^1 b_i^{1\prime}}{\sigma_i^2} \beta \\
&= \kappa(\lambda_1) \sum_{i=1}^{k_1} \frac{\|b_i^1\|^2 \|\beta\|^2 \cos^2(\widehat{b_i^1, \beta})}{\sigma_i^2} \\
&= \kappa(\lambda_1)\lambda_1 \|\beta\|^2 \sum_{i=1}^{k_1} \cos^2(\widehat{b_i^1, \beta}) \\
&= \kappa(\lambda_1)\lambda_1 \|\beta\|^2 \sum_{i=1}^{k_1} \Upsilon_{i1}^2
\end{aligned}$$

Where

$$\begin{aligned}
\kappa(\lambda_1) &\equiv \frac{1}{8\lambda_1^2} E_{\vartheta_1}\left[\frac{(\vartheta_1 + E_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right] \\
&= \frac{1}{2} E_{\vartheta_1}\left[\frac{(E(\delta|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right] \\
&= \frac{1}{2}\int \left[\frac{\left(\frac{\int \delta f(\delta, \lambda_1\delta - \vartheta_1)\mathrm{d}\delta}{\psi_1(\vartheta_1)}\right)^2}{\frac{1-\Psi_1(\vartheta_1)}{\psi_1(\vartheta_1)} + \frac{\int \vartheta_2 \psi(\vartheta_1, \vartheta_2)\mathrm{d}\vartheta_2}{\psi_1(\vartheta_1)}} \psi_1(\theta_1)\right] \mathrm{d}\vartheta_1 \qquad \text{(A.5)} \\
&= \frac{1}{2}\int \left[\frac{\left(\int \delta f(\delta, \lambda_1\delta - \vartheta_1)\mathrm{d}\delta\right)^2}{\Pr(\lambda_1\delta - r \geqslant \vartheta_1) + \int \vartheta_2 f\left(\frac{\vartheta_1+\vartheta_2}{2\lambda_1}, \frac{\vartheta_2-\vartheta_1}{2}\right)\frac{1}{2\lambda_1}\mathrm{d}\vartheta_2}\right] \mathrm{d}\vartheta_1 \\
&= \frac{1}{2}\int \left[\frac{\left(\int \delta f(\delta, \lambda_1\delta - \vartheta_1)\mathrm{d}\delta\right)^2}{\Pr(\lambda_1\delta - r \geqslant \vartheta_1) + \left(\int (2\lambda_1\delta - \vartheta_1) f(\delta, \lambda_1\delta - \vartheta_1)\mathrm{d}\delta\right)}\right] \mathrm{d}\vartheta_1
\end{aligned}$$

[8] $\Lambda_{k_1} = diag\left\{\frac{1}{\sigma_1}, \cdots, \frac{1}{\sigma_{k_1}}\right\}, \Lambda_{m_1-k_1} = diag\left\{\frac{1}{\sigma_{k_1+1}}, \cdots, \frac{1}{\sigma_{m_1}}\right\}$, $b_i^1, i = 1, \cdots, m_1$ are the columns of B_1', $\Upsilon_{i1} = \cos(\widehat{b_i^1, \beta})$. Similarly, the surplus obtained using system $\mathcal{P}_2$ is:

$$\Pi^*(\mathcal{P}_2) = \kappa(\lambda_1)\lambda_1 \|\beta\|^2 \sum_{i=1}^{k_2} \cos^2(\widehat{b_i^2, \beta}) = \kappa(\lambda_1)\lambda_1 \|\beta\|^2 \sum_{i=1}^{k_2} \Upsilon_{i2}^2, \qquad \text{(A.6)}$$

where $b_i^2, i = 1, \cdots, m_2$ are columns of matrix B_2', $\Upsilon_{i2} = \cos(\widehat{b_i^2, \beta})$. It follows that $\Pi^*(\mathcal{P}_1) \geqslant \Pi^*(\mathcal{P}_2)$ if and only if $\sum_{i=1}^{k_1} \Upsilon_{i1}^2 \geqslant \sum_{i=1}^{k_1} \Upsilon_{i2}^2$.

Proof of Corollary ??. The matrix

$$\Sigma^{-1/2} B B' \Sigma^{-1/2} = \begin{pmatrix} \Sigma_{11}^{-1/2} B_1 B_1' \Sigma_{11}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22}^{-1/2} B_2 B_2' \Sigma_{22}^{-1/2} \end{pmatrix}$$

has $m_1 + m_2$ eigenvalues $\lambda_1, \cdots, \lambda_{m_1}, \mu_1, \cdots, \mu_{m_2}$.

1. If $\lambda_1 > \mu_1$, then the first eigenvalue of $\Sigma^{-1/2} B B' \Sigma^{-1/2}$ is λ_1, its multiplicity is still k_1. If $\mathbf{q} \in \mathbb{R}^{m_1}$ is an eigenvector of $\Sigma_{11}^{-1/2} B_1 B_1' \Sigma_{11}^{-1/2}$ associated with λ_1, then $\hat{\mathbf{q}} = (\mathbf{q}, \mathbf{0})' \in$

[8]Here we drop the limits of integrations.

$\mathbb{R}^{m_1+m_2}$ is clearly the eigenvector of $\Sigma^{-1/2}BB'\Sigma^{-1/2}$ associated with λ_1. Subsequently, if Q_1Q_1' is the spectral projector of $\Sigma_{11}^{-1/2}B_1B_1'\Sigma_{11}^{-1/2}$ associated with λ_1, then

$$\hat{Q}_1\hat{Q}_1' = \begin{pmatrix} Q_1 \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} Q_1' & \mathbf{0} \end{pmatrix} = \begin{pmatrix} Q_1Q_1' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

is the spectral projector of $\Sigma^{-1/2}BB'\Sigma^{-1/2}$ associated with λ_1. Then the expected revenue of the principal with augmented performance system $\mathcal{P}$ is

$$\begin{aligned} \Pi^*(\mathcal{P}) =& \frac{1}{8\lambda_1^2} E_{\vartheta_1}\left[\frac{(\vartheta_1 + E_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right] \beta' B'\Sigma^{-1/2}\hat{Q}_1\hat{Q}_1'\Sigma^{-1/2}B\beta \\ =& \frac{1}{8\lambda_1^2} E_{\vartheta_1}\left[\frac{(\vartheta_1 + E_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right] \beta' B_1'\Sigma_{11}^{-1/2}Q_1Q_1'\Sigma_{11}^{-1/2}B_1\beta \\ =& \Pi^*(\mathcal{P}_1) \end{aligned} \tag{A.7}$$

2. If $\lambda_1 = \mu_1$, then the first eigenvalue of $\Sigma^{-1/2}BB'\Sigma^{-1/2}$ is λ_1, but its multiplicity is now $k_1 + k_2$. Let $Q_1(Q_2)$ represent an $m_1 \times k_1$ ($m_1 \times k_1$) matrix whose columns are orthonormal eigenvectors of $\Sigma_{11}^{-1/2}B_1B_1'\Sigma_{11}^{-1/2}$ ($\Sigma_{22}^{-1/2}B_2B_2'\Sigma_{22}^{-1/2}$) associated with λ_1 (μ_1). Then the columns of matrix

$$\hat{Q} = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} \in \mathbb{R}^{m_1+m_2}_{k_1+k_2}$$

form an orthonormal basis for eigenspace $\mathcal{N}(\Sigma^{-1/2}BB'\Sigma^{-1/2} - \lambda_1 I)$

$$\begin{aligned} \Pi^*(\mathcal{P}) &= \frac{1}{8\lambda_1^2} E_{\vartheta_1}\left[\frac{(\vartheta_1 + E_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right] \beta' B'\Sigma^{-1/2}\hat{Q}\hat{Q}'\Sigma^{-1/2}B\beta \\ &= \frac{1}{8\lambda_1^2} E_{\vartheta_1}\left[\frac{(\vartheta_1 + E_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + E_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right] \times \\ &\quad \beta'\left(B_1'\Sigma_{11}^{-1/2}Q_1Q_1'\Sigma_{11}^{-1/2}B_1 + B_2'\Sigma_{22}^{-1/2}Q_2Q_2'\Sigma_{22}^{-1/2}B_2\right)\beta \\ &= \Pi^*(\mathcal{P}_1) + \Pi^*(\mathcal{P}_2) \\ &> \Pi^*(\mathcal{P}_1) \end{aligned}$$

3. The case $\lambda_1 < \mu_1$ is similar to $\lambda_1 > \mu_1$, the proof is thus omitted.