Vendor-induced Performance Monitoring, Signaling, and Incentive Contracts in IT Outsourcing

Vijay Gurbaxani  
vgurbaxa@uci.edu

Shivendu Shivendu  
sshind@uci.edu

Qiang Zeng  
qzeng@uci.edu

The Paul Merage School of Business, University of California Irvine, Irvine, CA 92697

Abstract
Performance monitoring is often used by the principal to lower the rent paid to the agent due to moral hazard. However, in global sourcing of IT services, we observe a phenomenon wherein vendors voluntarily provide costly monitoring systems which allow the client to obtain accurate and timely information on their own performance. It is puzzling as why do vendors themselves offer mechanisms that lower their own rent extraction opportunities? We develop a model of IT outsourcing wherein IT services quality is non-verifiable, vendors have private information about their capabilities and quality-improving effort is unobserved by the client. Given the strategic and long-term nature of IT services contract, identifying the high capability vendor and inducing him to exert optimal level of quality-improving effort is critically important for the client. We show that performance monitoring is used as a credible truth-revealing signaling device by vendors to obtain favorable contracts. By committing to costly performance monitoring and convening verifiable signals on his private information to the client, a high capability vendor is able to separate his expected payoff from that of the free-riding low capability vendor. The dynamics of truth-revealing signaling with performance monitoring involves that the client credibly commits to a long-term relationship with the appropriate vendor with incentive contracts contingent on the verifiable signals from performance monitoring. When the monitoring-effort-output space can be partitioned, client achieves efficient screening. Otherwise, in order to achieve separation and optimize on information rent to the high type vendors, the client sacrifices allocative efficiency and efficient effort is not induced.

Key words: IT services, outsourcing, information asymmetry, performance monitoring, adverse selection, signaling, moral hazard, non-verifiable quality

WORKING PAPER (DRAFT)
04/17/2011
“Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted.”

Albert Einstein

1. **Introduction**

Principal employs monitoring mechanisms to mitigate contractual inefficiencies in the presence of moral hazard, i.e. when agent’s unobservable quality-improving effort has stochastic impact on quality (Ross 1973, Holmstrom 1979, Grossman and Hart 1983, Baron & Besanko 1987, Brazen and Suen 1988). Therefore, one would expect to observe principal-mandated monitoring in IT services contracts (Kern & Willcocks 2000). However, surprisingly we observe that, in IT outsourcing (see Dibbern et al. 2004 and Lacity et al. 2009 for comprehensive literature review), some vendors voluntarily offer monitoring systems to their clients. Aron et al. (2008) provide some examples:

> “Wipro InfoTech, a large India-based outsourcing firm offered many of its buyers additional layers of control using a technology driven mechanism called ‘The Cocoon’. This platform enables buyers to monitor the status of various projects and get an estimate of the eventual output quality level. Other firms such as HCL Ltd. (based in India) permitted their buyers to use real-time monitoring mechanisms to track how specific functions were being executed. Tata Consulting Services Ltd. (TCS) developed in-house management information systems that would enable the buyer of its BPO services, a large UK financial services firm, to monitor the working of project teams as work progressed.”

Why do the vendors voluntarily offer such costly monitoring technology that seems to limit their own ability to extract information rent from the client due to information asymmetry? It seems that any such vendor-induced monitoring will only benefit the client and may potentially lower vendor profit. Also, it’s not clear how do vendors determine level of investment in such monitoring technologies and whether these investments are uniform across different types of vendors. Therefore, in view of the strategic nature of IT outsourcing contracts, economic as well as strategic benefits of such vendor-induced monitoring mechanisms to the vendors and to the client need to be studied.

IT outsourcing is motivated by strategic, economic, and technological benefits (Lee 2001) and profitability of outsourcing relationships depend crucially on the efficacy of contractual arrangements (Lacity & Willcocks 1998; Gopal et al. 2003). These contracts are unique compared to other contracts because of following four key characteristics (Gurbaxani et al. 2011): (i) quality of IT services is impacted by intrinsic capability of vendor and quality-improving effort exerted by the vendor, (ii) vendor’s intrinsic capability is private knowledge and is not observed by the client, (iii) quality-improving effort of vendor is unobservable and its impact on quality of services is not deterministic, and (iv) the evaluation of the quality of the outsourcing services is difficult and often subjective (Ramachandran & Gopal 2010) and thus service quality is not verifiable though is perfectly observed when services are
delivered (Holmstrom 1985; Snir & Hitt 1994). These characteristics are key to any analytical investigation of IT outsourcing. Thus, outsourcing contracts are confounded by adverse selection, moral hazard and non-verifiability issues. Contractual inefficiencies created by informational asymmetry impact client as well as vendors. Information asymmetry impacts high ability more because they may not be able to attract favorable contracts as clients may not be able to identify them. On the other hand the client may bear substantial risks by being locked in with a low ability vendor (Aubert et al. 1998; Gurbaxani et al. 2011) or not being able to incentivize vendor to exert quality-improving effort (Susrala et al. 2009).

In this paper, we develop a model of outsourcing contracts which abstracts information asymmetry and non-verifiable quality of IT services on the lines of Gurbaxani et al. 2011 and monitoring mechanism on the lines of Lewis and Sappington 1991. We show that the vendor-induced performance monitoring is an effective signaling mechanism employed by heterogeneous vendor to signal their private information. By announcing an appropriate pricing scheme based only on output of the monitoring mechanism, client can achieve Separating Perfect Bayesian Equilibrium which results in higher payoff to the client as well as to the contracting high capability vendor. Counter to institution, our analysis demonstrates that vendor-induced monitoring is offered by the vendors, especially high capability vendors in self-interest and results in higher payoff to them. Monitoring not only mitigates adverse selection issues, it also provides an additional instrument to the client to improve contracting and thus induces efficient effort by the vendor. When vendor’s quality improving effort has large impact on IT services, client offers contracts that induce optimal effort and efficient screening is achieved. Not surprisingly, vendor-induced monitoring is also social welfare enhancing under certain conditions, and quality of delivered IT services is higher.

1.1 Performance monitoring and risk mitigation in IT outsourcing

While there are fruitful insights provided by the principal-agent models in the economics literature that focus on issues arising due to non-verifiability of quality in presence of moral hazard and/or adverse selection (Baron & Bensko 1984b; Sappington & Sibly 1986; Laffont & Tirole 1988, Sappington 2005), little work has been done in the IT outsourcing literature to study the contractual mechanisms available to the clients and vendors. Gurbaxani et al. 2011 show that client firms may use IT asset transfer as an effective screening mechanism as well as to induce efficient effort when quality of IT services is non-verifiable, the vendors’ intrinsic capabilities are private information, and their effort or relationship-specific investment is not observed. In this paper, our model captures the essential characteristics of the IT outsourcing contracts and posits that vendor-induced costly performance monitoring is used by the heterogeneous vendor to credibly signal their capability and the verifiable,
though stochastic, output of the monitoring system mitigates agency issues and improves contractual efficiency.

Since the seminal work by Spence (1973), there have been number of studies in economics (Milgrom & Roberts 1982; Kihlstrom & Riordan 1984; Bagwell & Ramsey 1988) that have studied signaling mechanisms that are used by more informed party to convey credible information to the less informed party to get better contractual terms. In IS literature, some research has done analytical modeling work in the framework of signaling, monopolistic screening, non-linear pricing and incentive contracts under information asymmetry (Chellappa and Shivendu 2010; Levi and Nault 2004; Meldelson and Whang 1990; Sundrarajan 2004). Our work adds to this growing stream of literature in the IS and is first analytical paper to our knowledge that studies monitoring in IT services contracts with non-verifiable quality.

Performance monitoring has been studied in IT outsourcing design (Elitzur and Wensley 1997, Ren and Zhou 2008, Chen and Bharadwaj 2009, Aron and Liu 2009), given that the outsourcing firms often have different and multiple objectives (Fitoussi and Gurbaxani 2011). Recently, with technological advances in remote monitoring tools, effort monitoring by the clients has become more popular in outsourcing relationships. For example, two large British banks have set up monitoring arrangements with their service providers that enable the client to monitor the working of the employees of the vendor (Aron et al. 2008). Since the client cannot directly observe the effort of the vendor and the intrinsic characteristics of the vendor are not fully known to the client, adverse selection followed by moral hazard present substantial risks to the outsourcing client (Aubert et al. 1998; Gurbaxani et al. 2011). Besides mitigating the risks of moral hazard, costly monitoring may enable the client to collect information on the vendor’s intrinsic characteristics and thus to identify the appropriate vendor (Bar-Ilan 1991). Information collected by monitoring may strengthen the incentive contract for vendor selection (Snir and Hitt 1994). Recently, researchers have found that performance monitoring mitigate moral hazard risks in IT outsourcing relationships (Aron et al. 2008, Aron and Liu. 2009). When the vendor’s effort affects the quality of outsourced services and is not directly observable, costly monitoring by the client, along with incentive contracts, reduces the opportunistic behavior by the vendor and thus provides assurance on the service quality.

In this paper, we show that vendor-induced costly performance monitoring can be used as a credible signaling device by vendors with high capability to obtain favorable contracts. By committing to costly performance monitoring and by generating verifiable signals that are directly correlated to the IT services quality, the high capability vendor is able to separate his expected payoff and that of the free-riding (one who does not exert effort) or mimicking (one who pretends to be of higher capability than his
actual type) vendors. The separation can be constructed in such a way that it is the self-interest of the vendor to reveal their type by self-selecting monitoring level that no free-riding or mimicking type vendor would find profitable to accept. The dynamics of truth-revealing -signaling with performance monitoring involves that the client credibly commits to a long-term relationship with the appropriate vendor with incentive contracts contingent on the verifiable output from monitoring.

Rest of the paper is organized as follows: In §2, we describe the model. In §3 we develop the benchmark contracts when service quality is verifiable. In §4, we develop contracts when quality is non-verifiable and show how vendor’s employ self-induced monitoring technology as a signaling mechanism. In §5 we analyze, signaling mechanisms and contracts under linear and concave pricing schemes. In §6 we discuss our results and identify suitable theoretical and managerial implications.

2 Model

We consider a principal-agent setting. The client firm is the principal who wishes to outsource implementation of an IT project that provides one unit of IT services. There are a large number of vendors or agents who are capable and willing to implement the project. The agents are heterogeneous in their intrinsic capability $\theta \in [\underline{\theta}, \overline{\theta}]$ with probability distribution $f(\theta)$ and cumulative probability distribution $F(\theta)$ which is differentiable everywhere. Further, we assume $f(\theta)$ is single-peaked (uni-modal) and is everywhere positive on its support such that hazard function $h(\theta) = \frac{f(\theta)}{F(\theta)}$, satisfies the monotone hazard rate property i.e., $h'(\theta) \leq 0$. Most common distributions satisfy these assumptions. Intrinsic capability $\theta$ is private knowledge of the vendor, though the distribution of the same is common knowledge.

In our conceptualization, following prior literature (Gurbaxani et.al. 2011), quality of the IT services delivered has two distinct components: one that is verifiable and thus contractible and the other that is observed by the client but is non-verifiable and thus is not contractible. Without loss of generality, we normalize the verifiable component of delivered IT services which is often governed through SLAs as zero. In the rest of the paper we focus on the non-verifiable component of quality of IT services which depends on vendor’s intrinsic capability $\theta$ and quality-improving effort $e \in [0,1]$. Note that vendor’s effort $e$ is unobserved by the client and has stochastic impact on quality $q$. If the stochastic impact of effort $e$ on quality is denoted by $y$, then specifically:

$$q(\theta, e) = \theta + y$$
Note that our setting is akin to Laffont & Martimort (2002) model of adverse selection followed by moral hazard. For tractability, without loss of generality, we model stochastic effort-related quality $y$ as $y \in \{0, Y\}$ with $Pr(y = Y | e) = e$. Thus if a vendor of $\theta$ type exerts $e(\theta)$ quality-improving effort, then quality of delivered IT services is:

$$q = \begin{cases} 
\theta + Y & \text{with probability } e(\theta) \\
\theta & \text{with probability } 1 - e(\theta)
\end{cases}$$

Further, the cost of quality improving effort is $\psi(e)$. The fixed cost of IT service provisioning is $c$. Note that quality delivered by a vendor of type $\theta$ is stochastic for any all effort levels $e \in (0,1)$. $Y$ denotes the quality premium that is associated with effort and for tractability, without loss of generality we model it to be binary.

Vendor firms offer monitoring technology $M$ which consists of probability $m \in (0,1]$ to client. The monitoring technology $M$ is imperfect in the sense that with probability $m$ it provides a verifiable output $\mu$ which is directly related to the non-verifiable IT services quality $q$ and with probability $1 - m$ it provides no information about the quality of services. For tractability we assume $\mu = q$. More formally: monitoring technology $M$ provides an output $\mu \in \{\mu, \phi\}$ such that $prob(\mu = q | m) = m$ and $prob(\mu = \phi | m) = 1 - m$. Cost of monitoring technology $M$ with probability $m$ is $C(m)$. Note that $m = 1$ denotes perfect monitoring and quality is perfectly verifiable, $m = 0$ denotes no monitoring and any $m \in (0,1)$ denotes imperfect monitoring wherein sometimes IT services quality is fully verifiable (when $\mu = q$) and sometimes it remains non-verifiable (when $\mu = \phi$).

When vendor firm with intrinsic capability $\theta$ adopts $m(\theta)$ as monitoring mechanism and exerts $e(\theta)$ quality improving effort then with probability $m(\theta)$ the client firm observes the quality of delivered IT services $\mu = q = \begin{cases} 
(\theta + Y) & \text{with probability } e(\theta) \\
\theta & \text{with probability } 1 - e(\theta)
\end{cases}$ and with probability $1 - m(\theta)$ observes nothing, i.e. $\mu = \phi$. Therefore, ex-post the client firm can receive three monitoring mechanism output $\mu : \theta + Y$ with probability $m(\theta) e(\theta)$ or $\theta$ with probability $m(\theta)(1 - e(\theta))$ or $\phi$ with probability $(1 - m(\theta))$. More specifically:
\[
\begin{align*}
\mu & \in \left\{ \begin{array}{l}
\theta + Y \text{ with probability } m(\theta)e(\theta) \\
\theta \text{ with probability } m(\theta)(1 - e(\theta)) \\
\phi \text{ with probability } (1 - m(\theta))
\end{array} \right. 
\end{align*}
\]

Client firm offers a contract contingent upon \( \mu \) after observing \( m(\theta) \). Note that monitoring technology as well as the output of the monitoring is observed and verifiable. If the client offers a contract \( \{ p(\mu) \} \), after observing \( m(\theta) \) then the vendor’s ex-ante expected payoff is given as:

\[
\pi_V(\theta) = m(\theta)p(\mu) - C(m(\theta)) - \psi(e(\theta))
\]

The client’s ex-ante expected payoff when she\(^1\) contracts with a vendor of \( \theta \) intrinsic capability who offers \( m(\theta) \) as monitoring technology and exerts \( e(\theta) \) quality-improving effort, is given as:

\[
\pi_C(\mu) = \theta + e(\theta)Y - m(\theta)p(\mu)
\]

Summary of notations is provided in Appendix A. We make following assumptions:

1. Cost of quality-improving effort \( \psi(e) \) is increasing and convex in \( e \): \( \psi'(e) > 0, \psi''(e) > 0 \) and \( \psi(0) = 0 \). Moreover, to ensure interior solutions, the Inada conditions \( \psi'(0) = 0 \) and \( \psi'(1) = \infty \) are met.

2. Cost of monitoring mechanism \( C(m) \) is increasing and convex in \( m \): \( C'(m) > 0, C''(m) > 0 \) and \( C(0) = 0 \). To ensure interior solutions, the Inada conditions \( C'(0) = 0 \) and \( C'(1) = \infty \) are met.

3. Quality premium associated with effort \( Y \) is strictly greater than the maximum vendor capability premium, i.e., \( Y > (\bar{\theta} - \theta) \). This assumption allows us to partition the monitoring-effort-output space to simplify signaling information rent-optimal effort inducing information rent tradeoff. In §4.4 we relax this assumption and briefly discuss the solution.

4. Client firm and vendor firms are risk neutral.

5. Fixed cost of IT service provisioning is constant and same across all vendors.

6. Stochastic impact of quality-improving effort and monitoring technology are independent, i.e., probability \( e(\theta) \) and \( m(\theta) \) are independent.

\(^1\)Throughout this paper, our client is she and vendor is he.
Before we analyze optimal contracts and viability of monitoring technology as signaling device by the vendors we have following definitions:

**Definition 1:** A contracting mechanism is said to induce efficient quality-improving effort, if its optimal for the vendor to exert 
\[e^* = \arg \max_e eY - \psi(e).\] The solution is given by \[\psi'(e^*) = Y.\]

**Definition 2:** A contracting mechanism is said to be efficient screening if (i) vendors’ are perfectly screened by their intrinsic capability by using monitoring technology as a signal and (ii) vendors’ exert efficient quality-improving effort.

**Definition 3:** A contracting mechanism is said to be the first-best if it implements efficient screening and no information rent is paid to the vendors.

Figure 1 describes the timing of the events.

![Figure 1: Timing of Events](image)

Figure 1: Timing of Events

Note that when the monitoring technology \( m = 1 \), i.e. monitoring is perfect, with Probability \[ \mu = q \] = 1, the client observes the quality with certainty and the same is verifiable. Further, any pricing scheme has to be such that \[ p \leq 0 \] when \[ \mu = \phi \]. The logic behind this is simple. If a vendor who has low capability, exerts zero effort, and invest in a zero level of monitoring technology is paid a positive price (above the total costs of providing the basic services), the client will be better off by lowering the price and the vendor still has the incentive to participate. The logic is true until the point at which the price is lowered at least to zero.

First, in §3 we develop the benchmark contracts when delivered IT services quality is verifiable and then in §4 we develop contracts with non-verifiable quality.

3. **Benchmark Contracts**

It is straight to see that if the client firm provisions the IT services in-house, then she always exerts efficient quality-improving effort \( e^* \). Further, the client always prefers to contract with the highest
capability vendor of $\bar{\theta}$ type and her *ex-ante* optimal expected quality of delivered services is $\bar{\theta} + e^*Y$. Also, for any given effort level $\bar{e}$, the client always prefers to contract with $\theta''$ than with any $\theta'$, $\forall \theta'' > \theta'$.

First, let us consider the simple case when there is no adverse selection and moral hazard, i.e., intrinsic capability of vendors is fully observed and is common knowledge, and quality-improving effort is observed by the client. Since services quality is verifiable, it is straight to see that the client contracts with $\bar{\theta}$ and offers the following contract to achieve the *first best*:

$$\{e, p\} = \{e^*, c + \psi(e^*)\}$$

In another setting with verifiable service quality, when there is no adverse selection but quality-improving effort is unobservable, i.e., there is moral hazard; the client contracts with $\bar{\theta}$ and the following contract achieves the *first best*:

$$p = \begin{cases} 
  c + \frac{\psi(e^*)}{e^*} & \text{when } q = \bar{\theta} + Y \\
  c & \text{when } q = \bar{\theta} + Y 
\end{cases}$$

Therefore, when quality is verifiable, moral hazard does not distort allocative efficiency in terms of optimal effort choice of the agent as long as there is no adverse selection. This is due to our assumption that the client firm and vendor firms are risk neutral and is consistent with the literature (Laffont and Mortimort 2002). It is easy to see that inefficiencies in quality-improving effort due to moral hazard will arise when the vendor is no longer risk-neutral. There are two alternative ways to study the rent-efficiency tradeoff in these transactions. One is to maintain risk neutrality for payoffs but impose a limited-liability constraint, which requires non-negative prices. Second is to let the vendor be strictly risk-averse. We will take the alternative one to study such impact in §3.1 under verifiable quality and will assume limited liability for our analysis in §4 and §5 under non-verifiable quality.

Before we analyze contracts under adverse selection and moral hazard under verifiable quality, let us look at a special case of moral hazard, which we term as *false moral hazard* following literature (Laffont and Mortimort 2002). When the impact of agent’s effort on the output is deterministic, then the adverse selection parameter and moral hazard unknown variable are blended together, in a deterministic way, into a single observed output available for contracting. This leads to no randomness at all in the benefit obtained by the principal when dealing with the agent as the link between effort, types and the observed quality is completely deterministic. Under false moral hazard, when a vendor of $\theta$ type exerts quality-improving effort $e$, the quality of delivered IT services is:

$$q = \theta + eY$$
Note that here $e$ is not a probability and quality of services is a continuous variable. It is easy to see that the following simple incentive contract achieves the *first best* in the presence of both adverse selection and *false moral hazard*:

$$p = \begin{cases} 
  c + \psi(e^*) & \text{when } q = \bar{\theta} + e^*Y \\
  c & \text{when } q < \bar{\theta} + e^*Y
\end{cases}$$

Only $\bar{\theta}$ type vendor who puts in efficient quality-improving effort $e^*$ gets zero utility and all others get strictly negative utility, and thus this contract implements *first best*.

### 3.1 Contracts under verifiable quality with adverse selection and moral hazard

First, we develop the contracts with unlimited liability which means that the pricing scheme announced by the client firm is such that under certain states of nature, vendor pays a penalty to the client. Or in other words, prices can be negative. The client announces a contract consisting of two price-quality pairs $\{p_1^*, q^*\}, \{p_2^*, q^*\}$, such that she pays price $p_1^*$ if observed quality is $q^*$, and pays price $p_2^*$ if observed quality is not $q^*$.

**LEMMA 1:** When quality is verifiable, with unlimited liability the unique Nash equilibrium in the price-effort setting game in the presence of adverse selection and moral hazard is $\{p_1^*, p_2^*, \hat{e}\}$ such that

$$p = \begin{cases} 
  p_1^* = Y + c + \psi(e^*) - Ye^* & \text{when } q = q^* = \bar{\theta} + Y \\
  p_2^* = c + \psi(e^*) - Ye^* & \text{when } q = \bar{\theta} + Y
\end{cases}$$

and $\hat{e} = e^*$.

For all proofs, see Appendix B.

It is easy to see that the pricing schedule given in Lemma 1 perfectly screens vendors and induces efficient quality-improving effort on the part of the vendor whose expected payoff is:

$$\pi_v(e^*) = e^* \left( Y + c + \psi(e^*) - Ye^* \right) + (1 - e^*)(c + \psi(e^*) - Ye^*) - \psi(e) - c = e^*Y + \psi(e^*) - Ye^* - \psi(e) = 0$$

The vendor’s optimal choice of effort $\hat{e} = e^*$ and his optimal expected payoff is zero as any other level of quality-improving effort leads to negative expected payoff for him. The client’s expected payoff is:

$$\pi_c = \bar{\theta} + Ye^* - e^* \left( Y + c + \psi(e^*) - Ye^* \right) - (1 - e^*)(c + \psi(e^*) - Ye^*) = \bar{\theta} + Ye^* - c - \psi(e^*)$$

**PROPOSITION 1:** With unlimited liability, when quality of IT services is verifiable the client firm can achieve first-best contracting even when there is asymmetric information in form of adverse selection and moral hazard.

This expected payoff of the client is optimal since it matches her expected payoff if she operates internally. As discussed above, these results are consistent with the literature (Laffont & Martimort 2002),
that with unlimited liability and risk-neutral agents, *first best* can be achieved without loss of efficiency due to moral hazard. The intuition behind this somewhat surprising result is that when quality is verifiable, the principal can induce efficient effort on the part of the agent by penalizing the agent if the favorable state of nature is not realized.

### 3.1.1 Contracts with limited liability under verifiable quality

Now we turn our attention to the more interesting case when vendors cannot be penalized by the client by charging a penalty if unfavorable state of nature is realized (i.e., \( Y = 0 \) when \( e > 0 \)). In other words, the vendors are protected by limited liability, i.e., penalty is infeasible. Now, it is possible that a free-riding (a vendor with \( \theta < \bar{\theta} \) who exerts zero effort) picks up the contact. It is easy to see that following contract with any \( x > 0 \), will deter free-riding (who exert zero effort) low type vendors:

\[
p = \begin{cases} 
  c + x & \text{when } q \geq \bar{\theta} \\
  0 & \text{when } q < \bar{\theta}
\end{cases}
\]

Only the \( \bar{\theta} \) type vendor or a mimicking type vendors (a vendor pretending to be of higher intrinsic capability than he actually is and exerting quality-improving effort) may accept the contract since free-riding low type vendors never provide quality \( q \geq \bar{\theta} \) and thus get a negative payoff \(-c\).

Now let us consider mimicking low type vendors (who are willing to exert high effort to achieve a higher level of quality). The free-riding proof contract as given above, will also prevent any vendors whose intrinsic value \( \theta < \bar{\theta} - Y \) from mimicking since the highest level of quality they can achieve will always be \( q < \bar{\theta} \). Since by assumption no. 3, \( Y > (\bar{\theta} - \bar{\theta}) \), there is no possibility that any mimicking type vendor can get a positive payoff by accepting the contract.

Now, we analyze the situation when the assumption no. 3 on quality premium due to effort \( Y \) is relaxed, that is there may be some type of vendors for whom mimicking by exerting higher effort is viable. For the vendors with intrinsic capability \( \theta \) which is such that \( \bar{\theta} > \theta > \bar{\theta} - Y \), the expected payoff of the vendor who exerts effort \( e \) is:

\[
e \left[ c + x \right] - c - \psi(e)
\]

In order to make mimicking a loss making exercise for the vendors, we must have:

\[
e \left[ c + x \right] - c - \psi(e) < 0
\]

Thus any \( x < \frac{e + \psi(e)}{e} - c \) always ensures that there is no mimicking. The right-hand side is convex and increasing in \( e \). When a mimicking low type vendor exerts quality-improving effort \( e = 1 \), the price should be set such that \( x < \psi(1) \). Thus the following incentive contract achieves perfect screening:
\[ p = \begin{cases} c + \psi(1) \text{ when } q \geq \theta \\ 0 \text{ when } q < \theta \end{cases} \]

The economic intuition here is that, for any mimicking low type vendor who exerts his effort \( e \leq 1 \), his expected payoff, \( e[c + \psi(1)] - c - \psi(e) \leq 0 \). However, the above contract would incentivize the \( \theta \) type vendor to exert zero effort to:

\[
\begin{cases}
    p = \frac{c + \psi(e^*)}{e^*} \text{ when } q = \theta + Y \\
    0 \text{ when } q \neq \theta + Y
\end{cases}
\]

In this game, the client sets the price and then the vendor optimizes on his choice of effort. Anticipating the vendor’s reaction over her price choice, the client will set the price that maximizes her own payoffs given the vendor’s reaction function. We therefore use backward induction to solve the Nash equilibrium in this price-effort setting game under the scenario that quality of services is verifiable and there is a random effect of unobserved effort on quality.

The vendor’s optimization problem is:

\[
\max_e p e - \psi(e) - c
\]

And the solution is given as \( \psi'(\hat{e}) = p \). Incorporating this reaction function of the vendor in choosing quality-improving effort level given an announced price \( p \), client’s optimization problem is:

\[
\max_p \bar{\theta} + \hat{e}(Y - p), \quad \text{s.t. } \psi'(\hat{e}) = p
\]

**LEMMA 2:** With limited liability, when quality premium due to effort is not large \( (Y < (\bar{\theta} - \theta)) \), the unique Nash equilibrium in the price-effort setting game under the scenario that quality of services is verifiable and that there is a random effect of unobserved effort on quality has \( \{p^*, \hat{e}\} \) is given by \( (Y - p^*)h'(p^*) = h(p^*) \) where \( \hat{e} = h(p^*) \).

**PROPOSITION 2:** With limited liability, when quality premium due to effort \( Y \) is not large \( (Y < (\bar{\theta} - \theta)) \), the client firm does not achieve first-best. Allocative efficiency of quality-improving effort is distorted to reduce rent. The optimal contact offered by the client firm achieves perfect screening; vendor exerts less than efficient effort and gets a positive rent.

**Example:** Let \( \psi(e) = \alpha e^2 \) and \( c = 0 \) (since the cost of providing the basic services is constant and common knowledge, this assumption is equivalent to this cost being paid by the client upon signing the
contract). The optimal price contract is
\[ p^* = \begin{cases} \frac{Y}{2} & \text{when } q = \theta + Y \\ 0 & \text{when } q \neq \theta + Y \end{cases} \]
and optimal effort by vendor is
\[ \hat{e} = \frac{Y}{4\alpha}. \]
Note that efficient level of effort is
\[ e^* = \frac{Y}{2\alpha} \] and \( \hat{e} < e^* \). Vendor’s \textit{ex-ante} expected payoff is
\[ \frac{Y^2}{16\alpha}. \]
Thus vendor has a positive expected rent \textit{ex-ante}.

Interestingly, with limited liability under certain conditions \( Y < (\theta - \theta) \), the client is not able to achieve the \textit{first-best}. The intuition is that the client’s marginal benefit of internal effort (equal to the quality differential) is higher than the vendor’s marginal benefit of effort given the client’s strategic choice on price (equal to half of the quality differential). Another interesting result is that the \( \theta \) type vendor is able to get a positive rent. Here, the tradeoff involves inducing effort and giving up an \textit{ex-ante} limited liability rent. In this case, the client chooses to economize on limited liability rent instead of inducing efficient effort.

In the next section, we analyze the general case in which the quality of services is non-verifiable and the effect of unobserved effort on quality is stochastic. The vendors are also protected by limited liability.

4. \textbf{Contracts with non-verifiable quality}

When the quality of delivered IT services is non-verifiable, then price is the only contracting variable available to the client firm in our setting. It is easy to see that if the client observes vendor’s intrinsic capability \( \theta \) and quality-improving effort \( e \), that there is no adverse selection and moral hazard, then by contracting with \( \theta \) type and requiring him to exert efficient effort \( e^* \), \textit{first-best} is achieved. In the presence of only moral hazard, client contracts with \( \theta \) type and offers \( p = c \) contract; vendor exerts zero effort and client gets quality \( q = \theta \). Similarly, in the presence of adverse selection only, client requires the vendor to exert \( e^* \) and contracts randomly with any vendor and offers
\[ p = c + \psi(e^*) \] and client \textit{ex-ante} gets expected quality
\[ q = \int_{\theta} \theta f(\theta) d\theta + e^* Y. \]
In the presence of adverse selection and moral hazard, client offers \( p = c \), contract randomly with any vendor who exerts zero quality-improving effort and client gets expected quality
\[ q = \int_{\theta} \theta f(\theta) d\theta. \] Thus in the presence of either adverse selection or moral hazard, or both, client does not achieve \textit{first-best} or even \textit{efficient screening}. 

13
From this analysis it’s clear that in the presence of adverse selection and moral hazard, when delivered IT services quality is non-verifiable, outsourcing contracts are inefficient because of (i) client cannot distinguish between vendors of heterogeneous capabilities, and (ii) client cannot induce efficient level of effort. This allocative inefficiency lowers client’s payoff, high intrinsic capability vendors fail to have any advantage in getting contracts, and vendors strictly get their outside option (zero payoff). Thus this inefficiency in contracting strictly lowers social welfare. Further, it’s clear that client cannot do much to mitigate this inefficiency as she has only one contracting variable, i.e., price. Given the nature of outsourcing contracts and non-verifiability of quality of services, client’s ability to employ some other contracting instrument is very limited.

Inefficient contracting hurts vendors as much as it hurts client firm. It is in the interest of the vendors, especially high capability vendors, to use some mechanism to convey their private information to the client to increase the probability of winning the contract. At the same time this “mechanism” to convey private information or “signaling” has to be such that it’s credible, i.e., it’s not in the interest of the vendor to convey favorable but untruthful information. Naturally, low capability vendors have the incentive to convey messages that indicate higher capabilities – mimic the high capability vendors – in hope of getting more favorable contacts. Consequently, a rational client views a rosy self-report on vendor’s capability skeptically unless it is backed by some credible evidence that may be too costly for low capability vendor to produce. Therefore, it is in the client’s own best interests that she offers contract terms that eliminate any incentive for a vendor of low capability to mimic a vendor of high capability.

And anticipating the client’s behavior, a vendor of high capability chooses a signaling mechanism that a vendor of low capability would find too costly to choose and would therefore, never choose. In this context, under the premise that the client is willing to pay a higher price when output of the monitoring mechanism is higher, vendor-induced performance monitoring may serve as a credible signaling mechanism. Vendors choose monitoring level to convey their private information on intrinsic capability. The output of the monitoring technology provides an additional strategic instrument to the client to mitigate contracting inefficiencies arising due to adverse selection and moral hazard. This in turn helps high capability vendors realize higher payoff through higher information rent.

4.1 Monitoring-Signaling Model

We proceed to develop a monitoring-signaling model following the literature (Baron & Myerson 1982; Mailath 1987; Ramsey 1996; Laffont & Martimort 2002) with a continuum of agent types. When

---

2 For limitations of service quality audit, SLAs etc. in improving contracting efficiency, see McFarlan & Nolan 1995; Kern & Willcocks 2000; Kern and Willcocks 2002.
vendor-induced monitoring technology is employed, the client receives a stochastic but verifiable output on quality $\mu$. We posit (and later show) that vendors choose the level of costly monitoring technology to convey information about their intrinsic capability and the output of monitoring $\mu$ provides an additional contracting variable to the client. Observing that the vendor chooses some level of monitoring technology, the client has two options of offering price schedule. One option is to contingent the price only on the verifiable message $\mu$ from the monitoring system and the other option is to contingent it on the output $\mu$ and the level of monitoring technology choice $m(\theta)$. In the model setting vendor of type $\theta$ has two choice variables: self-induced monitoring level $m(\theta)$ and after award of the contract, quality-improving effort level $e(\theta)$.

In our setting, the client firm derives value only from the quality of delivered IT services, which in turn is impacted only by type $\theta$ and effort $e$. So the monitoring does not directly impact quality, though it may induce vendor to exert effort as the client can compensate him through verifiable output of monitoring. Thus monitoring is useful to the extent that (i) it allows vendors to convey their private information on type and (ii) provides an additional instrument to the client to mitigate moral hazard related inefficiencies. Since client has to make sure that vendors’ payoffs are non-negative, though monitoring is paid for by the vendor, client has to keep the cost of monitoring in view in her pricing scheme. Thus, client needs to ensure that there is no overinvestment by the vendor in monitoring technology. Therefore, it is never optimal for the client to offer a $p(m,\mu)$ schedule. Client always offers $p(\mu)$ schedule such that $\frac{\partial p(\mu)}{\partial \mu} \geq 0$. We can see the logic of this property of the contract by contradiction. Let us assume that offering a $p(m,\mu)$ schedule is optimal for the client. Then we must have $\pi_c(p(m,\mu)) > \pi_c(p(\mu))$ for the client when she picks a vendor of type $\theta$. And then $\theta + e(\theta)Y - m(\theta)p(m,\mu) > \theta + e(\theta)Y - m(\theta)p(\mu)$, and $p(m,\mu) < p(\mu)$. By construction, we have $p(m,\mu) > p(\mu)$ for all $m \in M$. Since the above two statements contradict, the assumption that offering a $p(m,\mu)$ schedule is optimal for the client must be false.

Now we proceed to develop the signaling model and study some properties of the contract. Given a vendor of type $\theta$’s choice of $m(\theta)$, if he chooses $e(\theta)$ after the contract is awarded, his expected payoff is:

$$\pi_V(\theta) = m(\theta)p(\theta) + Ye(\theta) - C(m(\theta)) - \psi(e(\theta))$$  \hspace{1cm} (1)
Based on Revelation Principle (Green & Laffont 1977; Dasgupta et al. 1979; Myerson 1979), we are interested in a direct mechanism that incentivizes the vendors to truthfully reveal their private information through their choices on the level of monitoring technology. When different types of vendors choose the same level of monitoring technology, the client cannot infer their true types; we refer to this as a pooling equilibrium (or bunching solution). The client does infer their true types if she expects that certain monitoring level is chosen only by a certain type of vendors, we refer to this as a separating equilibrium. A separating equilibrium partitions the space of feasible choices of monitoring $M$, such that there is a one-to-one mapping from $m(\theta)$ to $\theta$, namely, $m(\theta)$ is a strictly monotonic in $\theta$. We use the standard equilibrium concept of signaling games – Perfect Bayesian Equilibrium – for our solutions. We add following additional definitions:

Definition 4: A Perfect Bayesian Equilibrium (PBE) of the monitoring-signaling game is a quartet
\[
\{m(\theta), \sigma(\theta), e(\theta)\}
\]
that satisfies the following requirements:

1. The client has a belief about which type(s) has (have) sent signal $m$ by choosing monitoring technology $m$. These beliefs can be described as a probability distribution $\sigma(\theta_j \mid m_j)$, the probability that the vendor has type $\theta_j$ if he chooses monitoring level as a signal $m_j$. The sum over all types $\theta_j$ of these probabilities is 1, conditional on any message $m_j$, i.e., $\forall m_j \in M, \int_{\theta} \sigma(\theta \mid m_j) d\theta = 1$.

2. The vendor’s choices $\{m(\theta), e(\theta)\}$ maximize his expected payoffs given the client’s strategy.

3. The client’s choice $\sigma(\theta)$ maximizes her expected payoffs given the vendors’ strategies and her own beliefs on the vendors’ types.

4. The client’s beliefs on the vendors’ types $\sigma$ are updated according to Bayes Rule.

Definition 5: A separating PBE is a PBE in which $m(\theta)$ is strictly monotonic in $\theta$.

Definition 6: A pooling PBE or bunching PBE is a PBE in which $m(\theta) = m(\hat{\theta})$ for at least one pair $\{\theta, \hat{\theta}\}$ in $[\theta, \hat{\theta}]$, where $\theta \neq \hat{\theta}$.

Based on assumption 3, it is clear that the client can induce efficient level of effort by providing an incentive when she observes $q = \theta + Y$ from the monitoring systems. In fact, it is straightforward to show that the following pricing schedule achieves the objective of inducing efficient effort since it aligns the vendor’s expected marginal benefit of effort with her own if the services are provided in-house:
\[ p = \begin{cases} p(\theta) + \frac{Y}{m(\theta)} & \text{when } q = \theta + Y \text{ and } m(\theta) \neq 0 \\ p(\theta) & \text{when } q = \theta \end{cases} \]

Now we proceed with this pricing schedule.

If a vendor of type \( \theta \) were to misrepresent his type as \( \hat{\theta} \) by choosing a monitoring level \( m(\hat{\theta}) \), his expected payoff is:

\[ \pi_v(\hat{\theta}, \theta) = m(\hat{\theta})p(\theta) + Ye(\theta) - C(m(\hat{\theta})) - \psi(\epsilon(\theta)) \]  

Therefore, the truth-telling incentive compatibility constraint is:

\[ \pi_v(\theta, \theta) \geq \max_{\hat{\theta}} \pi(\hat{\theta}, \theta) \quad \forall \theta \text{ in } \Theta \]  

and the vendor’s individual rationality constraint is:

\[ \pi_v(\theta, \theta) \geq 0 \quad \forall \theta \text{ in } \Theta \]

Given the vendors are truthfully revealing their types by choosing monitoring technology, the client’s payoff writes:

\[ \int_{\theta}^{\hat{\theta}} (p(\theta) - \Phi(\max(m(\theta))) - m(\theta)p(\theta)) f(\theta)d\theta \]

where \( \Phi(m) \) is the inverse function of \( m(\theta) \).

**LEMMA 3:** A Perfect Bayesian Equilibrium of the monitoring-signaling game satisfies the following conditions for all \( \theta \) in \([\theta, \hat{\theta}]\):

(i) \( m(\hat{\theta})p(\theta) - m(\hat{\theta})p(\hat{\theta}) \leq \pi(\theta) - \pi(\hat{\theta}) \leq m(\theta)p(\theta) - m(\theta)p(\hat{\theta}) \)

(ii) \( p(\mu) \geq p(\hat{\mu}) \quad \forall \mu \geq \hat{\mu} \)

(iii) \( m(\theta) \geq m(\hat{\theta}) \quad \forall \theta \geq \hat{\theta} \)

The condition (i) ensures that PBE satisfies incentive compatibility constraints and it’s always optimal for a vendor of type \( \theta \) to pick up \( m(\theta) \) monitoring technology. Condition (ii) ensures that client commits to have a non-decreasing payment schedule contingent on observed output of the monitoring technology which is a cornerstone of our solution technique as discussed above. The intuition behind condition (iii) is that incentive compatibility alone requires that the monitoring technology level has to be non-decreasing in type. This clearly makes monitoring technology as a potentially effective signaling device, as long as the client maintains a non-decreasing pricing schedule contingent on observed quality via monitoring system.
Client offers a pricing schedule \( p(\mu) \) to achieve Perfect Bayesian Equilibrium such that the incentive compatibility constraints given in Lemma 3 are satisfied. Therefore, the client has to design the scheme in such a way that (i) it is in the best interest for the vendors to truthfully reveal their type by announcing self-selected monitoring level and (ii) exert quality-improving effort that is optimal from the client’s perspective. Note that our setting is one of adverse selection followed by moral hazard, wherein expected output of the monitoring mechanism \( \mu \) is additive in intrinsic capability and effort. A limited liability rent must be given to the agent to induce effort and adverse selection incentive constraints must also be met to ensure self-selection. Thus presence of moral hazard exacerbates the conflict between the participation and adverse selection (Laffont and Martimort 2002).

Now, we first check if vendor-induced monitoring technology is a good signaling device that can lead to Separating Perfect Bayesian Equilibrium. In order for it to be a signaling device, optimal vendor-induced monitoring technology must be monotonic in vendor’s intrinsic capability type \( \theta \), i.e., \( m_\theta(\theta) > 0 \) or \( m_\theta(\theta) < 0 \) for all \( \theta \). And for those \( \theta \) in \( \Theta \) for which \( m_\theta(\theta) = 0 \), we get pooling equilibrium. This means that the slope of the monitoring schedule \( m(\theta) \) should not change sign in \( \Theta \). The truth-telling incentive compatibility constraint in (3) must be satisfied for monitoring technology to be a perfectly separating self-induced signaling device. Therefore the contract should be such that the best choice of \( \hat{\theta} \) by any \( \theta \) type should such that \( \hat{\theta} = \theta \). This implies that for truth-revealing to be optimal response for any \( \theta \), the following first order condition must be met for all \( \theta \):

\[
p(\theta)m_\theta(\theta) - C_\theta(m(\theta)) = 0
\]

In order for response \( \hat{\theta} = \theta \) to be locally optimal, we check for the local second order condition. But before we analyze monitoring-signaling game we need to check for single-crossing property. The single-crossing property states that the indifference curves of different types in the two-dimensional agent’s-signal-principal’s-action space cross only once. This property ensures that, if the low-type is indifferent between type signal-action pairs, then the high-type strictly prefers to send a higher signal. Therefore, the single-crossing property maps signals to types in such a way that it guarantees that higher types send weakly higher signals in equilibrium. In standard signaling models (ex. Spence’s education signaling model), the single-crossing property is achieved by the negative correlation between marginal costs of signaling and types: the high-type has lower marginal cost of signaling than the low-type does. In our model, however, the costs of signaling (monitoring) are homogenous among the agents. We achieve single-crossing property because there is a positive correlation between the marginal benefits of monitoring and agents’ types. The economic intuition is the same: to keep low-type indifferent when
adding signal (monitoring), he must be given a higher-valued action (price) than the high-type and thus in equilibrium the high-type chooses high signal and receives high price while the low-type chooses low signal and receives low price. Using this property and the condition (iii) of Lemma 3, we have the following proposition.

**PROPOSITION 3:** *Vendor-induced monitoring technology is a signaling mechanism.*

The economic intuition for vendor-induced monitoring technology to be a signaling device is as follows. As discussed before, client has to offer a pricing scheme that rewards more for higher observed output of the monitoring technology. Suppose there are two vendors of type $\theta_1$ and $\theta_2$ such that $\theta_1 > \theta_2$, and both exert same effort $e_1$. Expected payoff for $\theta_1$ is higher than expected payoff for $\theta_2$ for the same level of monitoring $m$ as $\frac{\partial p(\mu)}{\partial \mu} > 0$ and expected output of monitoring mechanism is increasing in type. Also, expected payoff’s expected benefit term $(mp(\mu))$ is linear in $m$ and cost term $C(m)$ is convex in $m$. So for any given effort level, expected payoff of a vendor will be inverted $U$ shaped and payoff maximizing monitoring level $m^*$ for $\theta_1$ will be higher than that of $\theta_2$. This is so because marginal benefit of monitoring is higher for $\theta_1$ than for $\theta_2$ (because for same effort $e$, expected $\mu$ for $\theta_1$ is higher than $\theta_2$), while the marginal cost of monitoring is the same. Vendors’ payoffs as functions of monitoring are shown in Figure 2 below. The optimal monitoring level is monotonically increasing for all $\theta \in [\hat{\theta}, \bar{\theta}]$ and we get separating PBE in this range. For all $\theta \in [\hat{\theta}, \bar{\theta}]$ we get pooling PBE as $m^*(\theta) = 1$ in this range.

![Figure 2: Separating and pooling equilibrium with monitoring as a signaling mechanism](image-url)
4.2 Pricing schedule offered by the client firm

Now, we study the properties of pricing contract \( p(\mu) \) for the client who takes into account the optimal choice of monitoring \( m(\theta) \) and effort \( e(\theta) \) of the vendor and incentive compatibility constraint in (3) and individual rationality constraint in (4) to maximize her own payoff. In order to keep the solutions tractable, as discussed before we have made a simplifying assumption that quality premium due to effort is strictly larger than the capability premium, i.e., \( Y(\bar{\theta} - \theta) \). This assumption ensures that on the optimal contract frontier, the distance between the pricing schedule under the favorable outcome of stochastic effort and that under the unfavorable outcome is independent of the optimal pricing schedule. Now we summarize the properties of the Separating PBE in the following lemma where we assume the functional form of cost of monitoring technology \( C(m) \) and cost of effort \( \psi(e) \).

**LEMMA 4:** Given that \( C(m) = \beta m^2 \) and \( \psi(e) = \alpha e^2 \), the Separating Perfect Bayesian Equilibrium exists under the condition that \( p(\theta) < 2\beta \) and it is characterized by the following properties:

(i). The vendor chooses his optimal monitoring level \( m^*(\theta) = \frac{p(\theta)}{2\beta} \).

(ii). The vendor chooses his optimal effort level \( \hat{e}(\theta) = e^* \).

(iii) The price schedule offered by the client is such that the information rent paid to the \( \theta \) type vendor is \( \int_{\theta}^\theta m(\tau)p_\theta(\tau)d\tau \).

(iv). The client updates her beliefs as \( \sigma(\theta = \hat{\theta} | m = m^*(\hat{\theta})) = 1 \).

![Figure 3: Pricing schedule, expected payments, and vendor’s expected payoff](image)
Figure 4: Pricing schedule, expected payments, and vendor’s expected payoff with top-bunching

Figure 3 and 4 above show the price schedule offered by the client firm and how vendor’s profits change with their type. The client has to pay information rent to the vendors to truthfully reveal their type. As expected the information rent is increasing in type and \( \theta \) gets highest rent and \( \bar{\theta} \) gets zero rent. When bunching PBE is achieved, then in the bunching region, vendor’s information rent is increasing in order to maintain the incentive compatibility for efficient effort, but the rate of increase of information rent is lower as client need not pay increasing information rent to vendor’s to reveal their type. In the bunching region monitoring technology has no signaling power. We observe that when the client sets \( p^{*}(\theta) = 0 \), the \( \bar{\theta} \) type vendor will choose \( m^{*}(\bar{\theta}) = 0 \) and \( \hat{e}(\bar{\theta}) = 0 \), and get zero payoff. The intuition is that the lowest type is indifferent between participation and pursuing his outside options, which is consistent with the individual rationality condition (4).

Figure 2 shows a monitoring-signal frontier for an announced pricing scheme. From part (i) of Lemma 4, we can see that the optimal monitoring level shifts upwards if the new announced pricing scheme \( \hat{p}(\mu) \) dominates \( p(\mu) \), i.e., \( \hat{p}(\mu) > p(\mu) \forall \mu \) as marginal benefit of monitoring increases while the marginal cost remains the same. Client’s optimal pricing scheme is such that the \( \bar{\theta} \) type gets the lowest rent and in order to keep the truth-telling as the optimal signaling mechanism for the vendors, the information rent is increasing in \( \theta \) in the part (iii) of Lemma 4. Since the announced pricing scheme ensures that vendor does not gain by deviating from truth-telling, i.e., more specifically:

\[
\frac{\partial \pi_{Y}(\theta)}{\partial \theta} = m(\theta)p_{y}(\theta).
\]

By observing the monitoring level offered by the vendor, the client updates her beliefs about the intrinsic ability of the vendor with probability 1 as stated in part (iv) of the Lemma 4. Further, we see that client’s payoff is increasing in type \( \theta \) and therefore, the client picks up the highest \( \theta \) who submits the offer. Note that in case of pooling equilibrium at the top, the power of monitoring level as a signal of type
vanishes and the client picks up a vendor from all those that offer \( m = 1 \) randomly as she can do no better.

**PROPOSITION 4:** A Separating Perfect Bayesian Equilibrium involves the client’s trade-off between reducing information rent and inducing efficient effort. The performance-based contract

\[
\left\{ p(\theta), p(\theta) + \frac{Y}{m(\theta)} \right\}
\]

induces efficient effort.

From Proposition 4 and Lemma 4, we know that monitoring-signaling game provides a solution such that the client can announce a pricing scheme that leads to perfectly separating self-selection of monitoring level by vendors as long as \( m < 1 \). So, vendor-induced monitoring solves adverse selection problem of the client. Since we have the setting of adverse selection followed by moral hazard, solving the moral hazard problem involves paying additional information rent to induce vendor to exert efficient level of effort. Thus the client is able to achieve efficient screening under Separating PBE, but not the first-best and she has to pay information rent to the vendors (Figure 5).

![Monitoring technology, payoffs as functions of \( \theta \)](image)

**Figure 5: Monitoring technology, payoffs as functions of \( \theta \)**

### 4.3 Contractual inefficiency when quality premium of effort is small

In order to simplifying our solution, we have made an assumption that \( Y > (\overline{\theta} - \theta) \). In this subsection we relax this assumption and study its impact on the allocative efficiency and information rent to be paid to the vendor. Now it is not possible for the client to offer a contract that incentivizes the vendor to pick up monitoring technology that signals their type and at the same time incentivize them to exert efficient effort. Since , now the client cannot partition the monitoring technology output \( \mu \) space.
into type-effort space, client’s ability to induce efficient effort vanishes. It is interesting to note that the optimal effort exerted by the vendor (which may be higher or lower than the efficient level of effort \( e^* \)) is also increasing in \( \theta \). The economic intuition for this result is that while the marginal benefit of effort to the vendor increases in monitoring level \( m \), the marginal cost of effort remains the same for all vendors. Since \( m \) is monotonic in intrinsic capability, when \( \theta \) increases, so does \( m \) and so do the marginal benefit of effort.

In our model, the client’s objective of inducing efficient level of effort (mitigating the moral hazard) is hampered by the needs of inducing vendors’ truth-revealing on their types (resolving the adverse selection issue). This is consistent with the insight from the economic literature that preventing moral hazard makes the adverse selection problem more severe (Laffont & Martimort 2002). When incentivizing all vendors truthfully signaling their types becomes imperative, inducing efficient level of effort may not be achieved in order to economize on the information rents paid to the high types of vendors. What causes this contracting limitation for the client is the assumption of vendors’ limited liability, i.e. the client is not able to penalize the vendor when the outcome of the stochastic effect of effort turns out to be bad. This positive-or-zero pricing schedule misaligns the incentive for efficient effort and that for minimizing rents when truth-revealing has to be achieved, reducing the efficiency of ex ante contracting (Laffont & Martimort 2002). Therefore when quality premium due to effort is small, client cannot achieve efficient screening and at best she can achieve perfect separation only.

In the next section we study properties of simple contracts, which may not be optimal from the client’s perspective but are easy to implement.

### 5. Implementable Incentive Contracts

In practice we observe easy to implement pricing schemes in outsourcing contracts like fixed-price, cost-plus price, input-based pricing and index pricing (Gurbaxani 2007). Therefore, in this section we study some properties of certain easy to implement contracts and also analyze social welfare implication. We first focus on a linear form of the incentive contracts in next section.

#### 5.1 Linear incentive contract

In this section, we study the properties of an incentive contract in which the client offers a price that is a linear function of output of monitoring technology, \( \mu \). In this contract, the client pays a fraction of verifiable output (Bapna et al. 2010) which results from the monitoring system. It is easy to see that a linear contract can be written as:
When the client offers linear pricing contract then monitoring is an effective signaling device only under the condition that \( b \leq \frac{2\beta}{\bar{\theta}} \). Further, vendor’s optimal monitoring signal is \( m^*(\theta) = \frac{b\theta}{2\beta} \), vendor’s optimal quality-improving effort choice is \( \hat{e}(\theta) = e^* \), and client’s payoff, vendor’s payoff and social welfare are \( \pi_c^*(\theta) = \theta - \frac{b^2\theta^2}{2\beta} \), \( \pi_v^*(\theta) = \frac{b^2\theta^2}{4\beta} + e^*Y - \alpha e^*e^2 \), and \( SW^L = \theta - \frac{b^2\theta^2}{4\beta} + e^*Y - \alpha e^*e^2 \). The second term in the social welfare is the total costs of monitoring.

**PROPOSITION 5:** When client offers a linear pricing contract, monitoring-signaling game results in a top-bunching PBE for all \( \theta \in \left[ \theta_B, \bar{\theta} \right] \), where \( \theta_B = \frac{2\beta}{b} \). All vendors with \( \theta \geq \theta_B \), signal with \( m^*(\theta) = 1 \) and exert effort \( \hat{e}(\theta) = e^* \).

The intuition is that in order to achieve complete sorting among the vendors, marginal information rents being paid to a higher type vendor is higher than the marginal surplus being created by selecting that higher type vendor. Thus, if the client wishes to achieve perfect separation then she has to pay higher pricing scheme \( p(\mu) \) and thus paying higher information rent. This implies higher cost of screening to the client, and she may prefer larger zone of vendor pooling.

Under many scenarios, the linear incentive contract achieves perfect separation among vendors so that it enables the client to select the best vendor. More interestingly, in order to achieve effective sorting among the highest types of vendors, the client has to pay a lower fraction of the observed output and that increases inefficiency in quality-improving effort. Furthermore, the linear contract incentivizes high type vendors to over-invest in monitoring technology so that it is possible that some highest types of vendors invest in monitoring technology at the maximum level and thus sorting among these vendors cannot be achieved. The reason is that the linear contract over-incentivizes the highest types of vendors to invest in monitoring technology which may not be in the best interest of the client. Is there a better contract that achieves efficient signaling? To address this issue, now we turn to a nonlinear contract that provides weaker incentives to high type vendors to achieve high expected quality.

**5.2 Nonlinear incentive contract**
As we see in the previous section, linear contracts may over-incentivize the highest types of vendors to over-invest in monitoring technology. In this subsection, we study a nonlinear contract that may overcome these limitations and still achieve perfect separation.

\[ p(\mu) = \begin{cases} b\sqrt{\theta} & \text{when } \mu = \theta \\ b\sqrt{\theta} + \frac{Y}{m(\theta)} & \text{when } \mu = \theta + Y \\ 0 & \text{when } \mu = \emptyset \end{cases} \]

and as the output of monitoring \( \mu \) increases, the rate of increase of payment to vendor is decreasing. Given that \( C(m) = \beta m^2 \), \( \psi(e) = \alpha e^2 \), and \( 0 < b < 1 \), optimal monitoring level and quality-improving effort chosen by a vendor of type \( \theta \) are:

\[ m^*(\theta) = \frac{\sqrt{\theta}}{2\beta} \]

and \( e(\theta) = e^* \). Monitoring serves as an effective signaling mechanism only when \( b < \frac{2\beta}{\sqrt{\theta}} \). The client’s and vendors’ optimal payoffs are \( \pi_c^*(\theta) = \theta - \frac{b^2\theta}{2\beta} \) and \( \pi_v^*(\theta) = \frac{b^2\theta}{4\beta} + e^*Y - \alpha e^{2} \). The social welfare is

\[ SW^{NL} = \theta - \frac{b^2\theta}{4\beta} + e^*Y - \alpha e^{2} \]

**Proposition 6:** For all \( \theta > \theta_B \), \( m^*(\theta) = 1 \), top-bunching takes place, where \( \theta_B = \frac{4\beta^2}{b^2} \). For all vendors, \( \pi_v^*(\theta > \theta_B) = b\sqrt{\theta} - \beta + e^*Y - \alpha e^{2} \).

We see that the self-selected monitoring level, client payoff and information rent paid to vendor is increasing in type \( \theta \) (as shown in Figure 3). Economic intuition is that the quality of delivered services increases in type and so client provides higher information rent to higher type in order to incentivize him to reveal himself by selecting higher level of monitoring. Since \( m^*(\theta) \) is monotonic in type \( \theta \), if \( m^*(\theta_B) = 1 \), then all \( \theta \in [\theta_B, \theta] \) choose monitoring level \( m^* = 1 \) and bunching takes place (Figure 5).

Interestingly, there is a negative effect of monitoring cost coefficient \( \beta \) on the likelihood of top-bunching: as \( \beta \) increases (monitoring is more costly), the likelihood of top-bunching decreases.

As cost coefficient of monitoring \( \beta \) increases, self-selected monitoring level, optimal effort level, information rent and vendor payoff decreases. The economic intuition is easy to see: marginal benefit of the monitoring remains the same while cost increases and vendors choose lower level of monitoring, this reduces the expected payoff that vendor receives. The client payoff increases as client’s payment of information rent decreases. Interestingly, as the sharing factor \( b \) increases, optimal monitoring level, and information rent to vendor all increase and the region of bunching also increases.
5.3. Analysis of the implementable contracts

In this section, we compare some properties of linear and non-linear contract and analyze changes in client and vendor payoff as cost parameters and sharing factor \( b \) changes. We also look at these contracts from the perspective of social planner in this client-vendor economy. The key results are in the following proposition.

**PROPOSITION 7:** Comparing the properties of the linear and non-linear concave pricing schemes:

(i) Client’s payoffs are decreasing in \( b \) under both contracts; client’s payoff is higher under linear contract when she has high sharing factor; client’s payoff is higher under nonlinear contract when she has low sharing factor. (ii) Vendor’s payoff is increasing in \( b \) under both contracts and. Vendor’s payoff is always higher under linear contract. (iii) Under both contracts, social welfare is increasing and then decreasing in \( b \), and (iv) Linear contract is optimal when client’s sharing factor is low, whereas nonlinear contract is optimal when client’s sharing factor is high.

![Figure 6: Client and vendor payoffs and social welfare as a function of sharing factor \( b \).](image)

The results summarized in Proposition 7 are illustrated in Figure 6 above. When the client adopts a linear compensation contract that pays the vendor a fraction of the observed output from the monitoring system, we find that the highest types of vendors being over-incentivized for investments in monitoring and so that bunching occurs more often. More interestingly, in order to achieve effective sorting among the highest types of vendors, the client has to pay a lower fraction of the observed output. When a nonlinear compensation contract is adopted, bunching on the top is less likely to occur. Therefore, the nonlinear contract performs better than the linear contract in terms of both effectiveness and efficiency.

**Discussion**
When vendors are heterogeneous in intrinsic capabilities that impact the quality of delivered IT services and quality-improving effort is unobservable, identifying the appropriate vendor and inducing efficient effort is of critical importance, given the long-term nature of IT outsourcing contracts. The vendor selection becomes further confounded due to non-verifiability of quality of delivered IT services. In this work, we develop a model of outsourcing contracts with non-verifiable IT services quality, wherein adverse selection is followed by moral hazard, wherein vendors use self-induced monitoring to share their private information on capability to get favorable contract from the client. Predictions of our analytical analysis conform with observed phenomenon where high capability outsourcing vendors offer self-monitoring technology to the clients.

Our model implies that the seemingly-wasteful investments in performance monitoring systems can actually benefit both the vendor and the client in the context of global sourcing of IT services. A vendor may find it optimal to truthfully reveal his private information by investing an appropriate amount in monitoring technology while the client commits to pay according to the verifiable output of the monitoring system. Interestingly, to ensure that the best vendor can be identified, under certain conditions, the client may have to compromise on the effort level the vendor exerts, and thus allocative efficiency may suffer.

One implication of our model to firms selecting outsourcing service providers is that observations of whether or not a service provider voluntarily invests in performance monitoring systems and how advanced or sophisticated the monitoring technology are very important. They may be actually related to the vendor’s service quality-related capabilities. Therefore, by requesting the potential vendors to commit such investments at a level that they should unilaterally decide, the outsourcing firm may gain valuable information on their internal characteristics. Our model also implies that investing in technologies which mitigate non-verifiability of service quality and thus providing the client additional contracting mechanisms may be a strategy that allows some vendors to differentiate themselves in the ever-increasingly competitive market.

We study two forms of incentive contracts: linear pricing and nonlinear pricing contracts to further explore insights into the contracting mechanism. When the client adopts a linear compensation contract that pays the vendor a fraction of the observed output from the monitoring system, we find that the highest types of vendors are over-incentivized for investments in monitoring and so that bunching is more likely to occur. More interestingly, in order to achieve effective sorting among the highest types of vendors, the client has to lower the fraction of the observed output being paid, which would further reduce allocative inefficiency on the dimension of inducing quality-improving effort. When a nonlinear compensation contract is adopted, bunching on the top is less likely to occur and thus the client may be
able to mitigate the allocative inefficiency by paying a higher portion of the discounted observed output from the monitoring system. Therefore, the nonlinear contract performs better than the linear contract in terms of both effectiveness and efficiency.

There are some interesting insights derived from our results. When outsourcing firms can employ additional mechanisms to select the best vendors (Gurbaxani et al. 2011) or potential vendors employ other credible devices to signal their types (Gurbaxani & Zeng 2010), inducing efficient level of effort from the vendor may be imperative and contracting mechanisms to mitigate vendors’ opportunism along effort-exerting should be in place. When effort-related efficiency issue can be well-addressed by potential contractual safeguards such as strategic partnerships (Gurbaxani & Ravindran 2010) or time-and-materials contracts (Bannerjee & Duflo 2000), the client may be in better position to address the vendor-selection issues effectively. We now consider the implications of altering some of our model’s assumptions on our analysis and results.

**Risk Aversion.** Suppose the vendors are risk averse. There are two potential sources of risk for the vendor in our model: monitoring technology may produce “no observation”, i.e., $\phi$ and/or after quality improving effort $\epsilon$ the realized state of nature is unfavorable, i.e., $y = 0$. When the vendors are risk averse, the clients has to increase the expected monetary transfer when favorable outcome of the risky events are realized as our model requires that $p(\phi) = 0$ (to screen out free-riding) and all vendors have to be paid the same price contingent on output when $y = 0$. This means that both pricing schedules $p(\theta)$ and $p(\theta + Y)$ move up, the difference between these two price schedules is larger for any given feasible $\mu$, the vendor’s expected payoff is same, client’s expected payoff is lower and social welfare is lower. Monitoring remains an effective signaling device though the bunching region becomes larger.

**Heterogeneous cost of monitoring.** When the cost of investments in monitoring technology is correlated to vendor’s type, the efficacy of monitoring as a signaling device may be affected. If a high type vendor’s marginal cost of monitoring is lower than that of a low type vendor, single-crossing property would be achieved in the same way as Spence’s education-signaling model, then monitoring serves well as a signaling model. Furthermore, the highest types would be even more incentivized for investments in monitoring and thus the region of top-bunching may be larger. If a high type vendor’s marginal cost of monitoring is higher than that of a low type vendor, higher incentives may be required to achieve monotonic monitoring choices among types. Interestingly, the region of tope-bunching may be smaller.

**Heterogeneous cost of effort.** If a high type vendor has lower marginal cost of exerting effort than a low type vendor does, it may be less costly for the client to induce efficient effort from high types.
And thus the client could pay less rent to achieve efficient screening. If a high type vendor has higher marginal cost of exerting effort, it may be more costly for the client to induce efficient effort from high types. The client may find it optimal to allow higher distortion in effort-inducing from the efficient level in order to economize on information rents.

Future research may study the situation in which multiple vendors are engaged by a client in the context of multi-sourcing (Bapna et al. 2010) when quality is not verifiable and effort is not observable. In this paper we have also linked monitoring with the extent to which quality is verifiable, altering this setup and thus studying the impact of the degree of quality verifiability may be interesting and may provide some more insights.

Reference


Appendix A: Summary of Notations

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>Vendor's capability</td>
<td>$\hat{e}$</td>
<td>Vendor’s optimal choice of effort</td>
</tr>
<tr>
<td>$\bar{\theta}$</td>
<td>Lower bound of capability</td>
<td>$m'$</td>
<td>Vendor’s optimal choice of monitoring</td>
</tr>
<tr>
<td>$\overline{\theta}$</td>
<td>Upper bound of capability</td>
<td>$\alpha$</td>
<td>Parameter in function of cost of effort</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Probability of the vendor being high capability</td>
<td>$\beta$</td>
<td>Parameter in function of cost of monitoring</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Verifiable quality of services</td>
<td>$b$</td>
<td>Parameter in pricing function</td>
</tr>
<tr>
<td>$e$</td>
<td>Vendor’s quality-improving effort</td>
<td>$\theta_h$</td>
<td>Vendor’s type that defines the region of bunching</td>
</tr>
<tr>
<td>$e^*$</td>
<td>Efficient effort</td>
<td>$\sigma$</td>
<td>Client’s updated belief on the vendor’s type</td>
</tr>
<tr>
<td>$p$</td>
<td>Price</td>
<td>$\pi_c$</td>
<td>Client’s expected payoff</td>
</tr>
<tr>
<td>$y$</td>
<td>Effort-related component of quality</td>
<td>$\pi_v$</td>
<td>Client’s expected payoff</td>
</tr>
<tr>
<td>$\psi$</td>
<td>Cost of exerting effort</td>
<td>$\mu$</td>
<td>Output of monitoring system</td>
</tr>
<tr>
<td>$m$</td>
<td>Monitoring level</td>
<td>$f$</td>
<td>probability density function of $\theta$</td>
</tr>
<tr>
<td>$C$</td>
<td>Cost of monitoring</td>
<td>$F$</td>
<td>Cumulative probability function of $\theta$</td>
</tr>
</tbody>
</table>

Appendix B: Proofs of Lemmas and Propositions

Proof of Lemma 1:

The payoff of the vendor of highest available type $\overline{\theta}$ writes

$$
\pi_v(\overline{\theta}) = ep_1^* + (1 - e)p_2^* - c - \psi(e) = e\left[ Y + c + \psi(e^*) - Ye^* \right] + (1 - e)\left[ c + \psi(e^*) - Ye^* \right] - c - \psi(e)
$$

which can be simplified as

$$
\pi_v(\overline{\theta}) = eY - \psi(e) + \psi(e^*) + e^*Y
$$
The first order condition writes \( \dot{e}Y = \psi'(\dot{e}) \) and therefore \( \dot{e} = e^* \) and \( \pi^*(\bar{\theta}) = 0 \) (he is indifferent between accepting the contract and pursuing outside options).

Given the vendor’s optimal choice \( e^* \), the Client’s expected payoff writes

\[
\pi_C(\bar{\theta}) = \bar{\theta} + e^*Y - e^*p_1 - (1 - e^*)p_2 = \bar{\theta} + e^*Y - e^*[Y + c + \psi(e^*) - Ye^*] - (1 - e^*)[c + \psi(e^*) - Ye^*]
\]

which can be simplified as \( \pi_C(\bar{\theta}) = \bar{\theta} + e^*Y - c - \psi(e^*) \) which achieves the optimal outcome.

**Proof of Lemma 2:**
The vendor’s optimization problem is:

\[
\max_{e} pe - \psi(e) - c
\]

And the solution is given as \( \psi'(\dot{e}) = p \) or \( \dot{e} = h(p) \). Incorporating this reaction function of the vendor in choosing quality-improving effort level given an announced price \( p \), client’s optimization problem is:

\[
\max_{p} \bar{\theta} + \dot{e}(Y - p), \quad \text{s.t.} \quad \psi'(\dot{e}) = p
\]

Anticipating the vendor’s optimal choice \( \dot{e} = h(p) \), the above program rewrites as

\[
\max_{p} \bar{\theta} + h(p)(Y - p)
\]

The first order condition writes \( (Y - p^*)h'(p^*) - h(p^*) = 0 \) or \( (Y - p^*)h'(p^*) = h(p^*) \).

**Proof of Lemma 3:**
By definition (1), we have

\[
\pi(\dot{\theta}) = m(\dot{\theta})p(\dot{\theta}) + Ye^* - C(m(\dot{\theta})) - \psi(e^*)
\]

The incentive compatibility constraint (3) implies

\[
\pi(\theta) \geq \pi(\dot{\theta}, \theta) = \pi(\dot{\theta}) + m(\dot{\theta})p(\theta) - m(\dot{\theta})p(\dot{\theta})
\]

Or

\[
\pi(\theta) - \pi(\dot{\theta}) \geq m(\dot{\theta})p(\theta) - m(\dot{\theta})p(\dot{\theta}) \quad (A5)
\]

Similarly,

\[
\pi(\theta) - \pi(\dot{\theta}) \leq m(\theta)p(\theta) - m(\dot{\theta})p(\dot{\theta}) \quad (A6)
\]

(iii) of Lemma 3 holds by assumption.
From (A5) and (A6), we see that \( (m(\theta) - m(\dot{\theta}))(p(\theta) - p(\dot{\theta})) \geq 0 \).

From (iii) of Lemma 3, we have

\[
p(\theta) \geq p(\dot{\theta}), \quad \text{therefore} \quad m(\theta) \geq m(\dot{\theta}).
\]

**Proof of Proposition 3:**

The incentive compatibility constraint \( \pi(\theta, \theta) \geq \max_{\dot{\theta}} \pi(\dot{\theta}, \theta) \forall \theta \) in \([\underline{\theta}, \bar{\theta}]\) implies that the following first-order condition for the optimal response \( \dot{\theta} \) chosen by type \( \theta \) is satisfied:
\[ m_\theta(\hat{\theta})p(\theta) - C_\theta(m(\hat{\theta})) = 0 \]

For the truth to be an optimal response for all \( \theta \), it must be the case that
\[ m_\theta(\theta)p(\theta) - C_\theta(m(\theta)) = 0 \]  \( \text{(A7)} \)

And (A7) must hold for all \( \theta \) in \( \Theta \) since \( \theta \) is unknown to the client.

It is also necessary to satisfy the local second-order condition,
\[ p(\theta)(m_{\theta\theta}(\hat{\theta}) | \hat{\theta} = \theta) - C_{\theta\theta}(m(\hat{\theta})) | \hat{\theta} = \theta \leq 0 \]

Or
\[ p(\theta)m_{\theta\theta}(\theta) - C_{\theta\theta}(m(\theta)) \leq 0 \]  \( \text{(A8)} \)

Differentiating (A7), (A8), can be written more simply as
\[ m_\theta(\theta)p_\theta(\theta) \geq 0 \]  \( \text{(A9)} \)

By assumption, we have
\[ \frac{\partial p(\theta)}{\partial \theta} \geq 0 \]  \( \text{(A10)} \)

(A9) and (A10) imply that \( m_\theta(\theta) \geq 0 \).

**Proof of Lemma 4:**

By solving the first order conditions, the vendor’s optimal choices on effort and monitoring write,
\[ \hat{e}(\theta) = e^*, \quad \hat{m}(\theta) = \frac{p(\theta)}{2\beta}. \]

Dividing (i) of Lemma 3 by \( \theta - \hat{\theta} \) and take the limit as \( \hat{\theta} \to \theta \), we obtain \( \pi_\theta(\theta) = m(\theta)p_\theta(\theta). \)

Integrating the above equation implies that the following equation must hold.
\[ \pi(\theta) = \pi(\hat{\theta}) + \int_\hat{\theta}^\theta m(\tau)p_\theta(\tau)d\tau \]

Imposing that \( \pi(\hat{\theta}) = 0 \), we have \( \pi(\theta) = \int_\hat{\theta}^\theta m(\tau)p_\theta(\tau)d\tau. \)