INSEAD

The Business School for the World®

Faculty & Research Working Paper

Optimal Contracts for Outsourcing of Repair and Restoration Services

Nitish JAIN
Sameer HASIJA
Dana POPESCU
2011/91/TOM

Optimal Contracts for Outsourcing of Repair and Restoration Services

Nitish Jain*

Sameer Hasija**

Dana Popescu***

* PhD Candidate in Technology and Operations Management at INSEAD, 1 Ayer Rajah Avenue, Singapore 138676. Email: nitish.jain@insead.edu

* Assistant Professor of Technology and Operations Management at INSEAD, 1 Ayer Rajah Avenue, Singapore 138676. Email: sameer.hasija@insead.edu

* Assistant Professor of Technology and Operations Management at INSEAD, 1 Ayer Rajah Avenue, Singapore 138676. Email: dana.popescu@insead.edu

Optimal Contracts for Outsourcing of Repair and Restoration Services

Nitish Jain, Sameer Hasija, Dana G. Popescu
Technology and Operations Management, INSEAD, 1 Ayer Rajah Avenue, 138676 Singapore.
{nitish.jain, sameer.hasija, dana.popescu}@insead.edu.

Outsourcing of equipment repair and restoration is commonly practiced by firms across many industries. The operational performance of the equipment is determined by joint decisions of both the firm and the service provider. Although some decisions are verifiable (and thus directly contractible), many decisions in such settings are unverifiable. This naturally creates a double-sided moral hazard environment in which each party has incentives to free ride on the other party's effort. A performance-based contract allows the firm to align the incentives of the service provider but also exposes the service provider to stochastic earnings that, in turn, create disincentives to the provider's optimal decision making. To capture these issues, we develop a novel principal–agent model by integrating elements of the machine repairman model and a stochastic financial distress model within the double-sided moral hazard framework. We apply our model to solve the firm's problem of designing the optimal performance-based contract. We find that the firm can attain the first-best profit by restricting the search space to only two classes of performance-based contract structures: linear and tiered. We show that the linear contract structure has only limited capacity for attaining the first-best outcome, contingent on the exogenous characteristics of the vendor. In contrast, the tiered contract structure allows the client to attain the first-best outcome regardless of vendor characteristics. Our results provide normative insights on the role of contract structures in eliminating any loss due to double-sided moral hazard or the vendor's financial concerns. The results reported here also provide theoretical support for the extensive use of tiered contracts in practice.

Key words: Double-sided moral hazard, Performance-Based Contracts, Financial distress, Machine repairman model

1. Introduction

Over the last two decades, firms have increasingly outsourced repair and restoration services of their equipment to external service providers. The advantages of hiring and training engineers to manage repair in-house are becoming limited for firms as a result of continuous advancement in

equipment technology, decreasing margins due to competition, and the increasing preference for geographical diversification of business units. For example, the repair and restoration outsourcing contract by DBS Bank (Singapore) to IBM exemplifies all of these factors.[1] To keep up with the latest technology, DBS Bank engaged IBM for 10 years to carry out repair and restoration of servers in its two data centers, which are located in Singapore and Hong Kong. DBS Bank expected to gain overall cost savings of about 20% by leveraging the specialized information technology know-how of IBM engineers, who were exclusively dedicated to fulfilling DBS Bank's service requirements.

Firms typically trade off the advantages of outsourcing against the agency problems that naturally arise in such settings. The inability of a firm to contract directly on the outsourced decisions creates moral hazard issues between the firm and the service provider. These issues may result in loss of efficiency of the service supply chain and loss of profit for the firm. The outsourcing of repair and restoration services poses additional challenges because the equipment's operational performance, defined here as availability, is determined by the joint efforts of both the firm and the service provider. In such joint effort settings, each player may have incentives to free ride on the other player's effort, and this may result in suboptimal system outcomes (Holmstrom 1982, Bhattacharyya and Lafontaine 1995).

In our setting, the equipment's operational performance is determined by how frequently the equipment fails (the failure rate) and how quickly it can be restored (the repair rate) when it fails. Whereas various decisions of the service provider (such as hiring and training of engineers) determine the equipment's repair rate, various decisions of the firm (such as training of operators and investments in the operating environment) determine the equipment's failure rate. In the DBS–IBM outsourcing relationship, IBM is responsible for decisions on hiring and training of its engineers, both of which affect the repair rate of servers. DBS Bank is responsible for decisions regarding investments in the operating environment of its data centers, which affect the failure rate of its servers.[2] In a related example, our discussions with the management of Raymond Corporation[3] reveal that operating floor quality at the client's premises was the main factor determining the failure rate of its forklift trucks. Additionally, training of operators and expenditure on routine

checks and maintenance activities can significantly influence the failure rate of certain types of equipment, including medical equipment (Cummins et al. 1991) and forklift trucks.[4] Some of the decisions made by the firm and the service provider may be verifiable, but several are not verifiable in a court of law (e.g, efforts in the training of equipment operators and repair engineers). Direct contracts cannot be written on unverifiable decisions, a fact that creates a double-sided moral hazard problem.

If direct contracts are thus precluded, then the firm must design contracts that measure and appropriately reward or penalize the service provider based on metrics that provide information about its efforts. Such incentive or performance-based contracts are increasingly used in the service outsourcing setting because they mitigate the moral hazard problems that arise in outsourcing and also free the client of the transaction costs associated with closely monitoring the vendor. For example, the US Department of Defense issued a directive in 2003 requiring program managers to develop and implement performance-based contract strategies to optimize systems availability (Kim et al. 2007). Hasija et al. (2008) highlight the growing adoption of different incentive contracts in the call center outsourcing industry.

According to a review report on performance-based contracts by the Government Finance Officials Association,[5] a key emerging challenge for the implementation of performance-based contracts is that such contracts expose the service provider to uneven and unplanned cash flows that may be undesirable. Volatile earnings inhibit small service providers, which typically have limited access to financial resources, from carrying out day-to-day operations and thus may lead them to financial distress. Such financial concerns reduce the incentives of these service providers to make optimal decisions. One way to ensure that a service provider makes optimal decisions is to remit some of the surplus to that provider. The surplus ensures that the service provider's exposure to potential financial distress is within acceptable levels, although this limits the firm's ability to extract first-best optimal profits. Earnings volatility may not jeopardize financing of the larger service providers' but managers of those providers may have their bonuses or continuation with the firm tied to project-specific profits (or losses). Hence earnings volatility may distort the incentives of

these managers to make optimal project-specific investments. As industry experts have noted: "No vendor expects to achieve 100% of all service levels on a regular basis, thus they pad their profit margins to accommodate some performance issues. The most sophisticated vendors run probability models to determine likely impacts to their profitability" (see "Outsourcing Contract Penalties: Do Vendors Respond to the Pain?"[6]).

In this paper, we study contract structures that address two key challenges faced by a firm (the client, "he") that outsources the repair and restoration service of his equipment to a service provider (the vendor, "she"): (i) double-sided moral hazard, and (ii) the vendor's induced exposure to financial distress on account of performance-based contracts. We assume such an outsourcing environment to be exogenous and so do not study the client's decisions concerning what (and whom) to outsource. In many industries, such as banking and healthcare, firms may choose to own and operate their own equipment because of concerns about confidentiality, quality of service, and other factors affecting the firm's competitive advantage and thus may outsource only the repair and restoration services. We focus on static (strategic-level) decisions that influence the repair and failure rates, that are made via long-term commitments, and that cannot be altered over short periods. We abstract all such decisions made by the client as monetary investments in *operating technology* and likewise the decisions made by the vendor as monetary investments in *repair technology.*

We study the contract design problem from the client's perspective and compare the effectiveness of the optimal contract structure against the *first-best* outcome achieved by the client in a centralized setting. Specifically, we ask the following questions: (i) What is the optimal contract structure that should be offered to the vendor? (ii) How much does the client lose (compared with the first-best outcome) under such an optimal contract? (iii) How does the optimal contract structure depend on the vendor's tolerance towards potential financial distress?

In order to answer these questions, we limit our focus to two classes of contract structures: *linear* and *tiered.* Our choice of linear contracts is motivated by the past literature (Bhattacharyya and Lafontaine 1995, Corbett et al. 2005, Roels et al. 2010). Bhattacharyya and Lafontaine (1995)

Performance Measure	Tiered Levels	Penalty as % of Monthly Charges
Average delivery time to restore network services	1 hour in a calendar month	5%
	≥ 2 hours in a calendar month	10%

Note: The above terms are extracted from a sample contract offered by a voice and data network service provider.

Table 1 Example of a Tiered Contract

show that, for double-sided moral hazard problems, a linear contract can replicate any nonlinear optimal contract. In our setting, linear contracts allow the customer to penalize the vendor with a proportional penalty contingent on her realized performance. In practice, however, we find that contracts are designed to levy a limited penalty on the vendor if her performance deteriorates beyond a threshold level. Table 1 shows an example of tiered penalties in the context of a repair and restoration service. Analogously, contracts in the call center outsourcing industry often stipulate service-level agreements (SLAs)—for instance $\Pr(\text{delay} \leq 60 \text{ sec}) \geq 0.8$ for a day—along with an associated (constant) penalty for not meeting the SLA. See Hasija et al. (2008) for more examples of such constant-penalty SLA contracts used in practice. This motivates our choice of tiered contracts.

Even though our analysis is restricted to these two contract structures, we find that they are sufficient for the client to attain the first-best optimal solution with all feasible values of exogenous parameters. We find that, for a vendor with a high tolerance for financial distress and/or a high reservation value, the client can attain the first-best outcome with both linear and tiered contracts. In contrast, for a vendor with a low tolerance for financial distress and/or a low reservation value, we find that linear contracts fail to attain the first-best outcome. It is interesting that the client can continue to attain the first-best outcome with such a vendor by using tiered contracts. This finding runs counter to results in the existing literature that suggest optimal contracts in a double-sided moral hazard environment attain only the second-best outcome (Bhattacharyya and Lafontaine 1995).[7] Intuitively, the limited penalty structure in the tiered contracts may have a countervailing effect: while reducing the vendor's exposure to potential financial distress, it may also reduce her incentives to invest optimally in the repair technology. Yet surprisingly there exists a class of tiered contracts that allow the client to avoid *any* loss due to the vendor's low tolerance for financial

distress, thereby eliminating any need for the client to provide surplus to the vendor.

Even though the double-sided moral hazard setting seems natural to the repair and restoration service outsourcing environment, this paper is, to the best of our knowledge, the first work to model this setting analytically. Our model also captures the real-world managerial issue of vendor concerns about the potential financial distress caused by volatile earnings due to performance-based contracts in a stochastic system. Our work complements earlier studies that stress the importance of carefully designing a contract to achieve high contract efficiency of the supply chain. In our setting we show that the client can attain the first-best profit—despite the double-sided moral hazard and the vendor's financial concerns—by selecting the right contract structure and setting the appropriate parameter values. Our results provide theoretical support for the prevalence of tiered contracts in practice (Hasija et al. 2008). From a modeling perspective, our paper offers a novel approach that combines technical elements of three different fields of research: operations (renewal theory), economics (contract theory), and actuarial science (ruin theory[8]).

The rest of the paper is organized as follows. In Section 2 we review the related literature, and in Section 3 we discuss the model setup and formalize related assumptions. In Section 4, we present the analysis and discuss findings of the client's problem of searching for the optimal contract structure. We conclude in Section 5 by summarizing the main contributions of our work and opportunities of future research.

2. Literature Review

Our work is related to three primary research streams: the operations management literature on service operations outsourcing, the economics and operations management literature on incentive alignment under double-sided moral hazard, and the growing body of literature at the operations–finance interface that studies operational decisions contingent on their impact on financial health of a firm or service providers.

In the service operations outsourcing stream, our work is in the specific substream of after-sale service for equipment. Murthy and Asgharizadeh (1999) is among the early papers to study repair

outsourcing for equipment. They model the repair outsourcing problem between multiple clients and a single vendor as a Stackelberg game in which the vendor is the leader and the clients (equipment operators) are followers. The vendor offers a fixed-price contract to the clients and makes decisions on the number of service channels to set up and the number of customers to serve. Kim et al. (2007) is the first paper to study such after-sale services supply chain contracting issues within the principal–agent framework. They studied a one-to-one relationship between the client and the vendor at the product component level. These authors addressed the single-sided moral hazard problem that arises when the vendor makes two key repair decisions—cost reduction and spare-parts inventory management—on the client's behalf. They show in this case that if both parties are risk neutral then the first-best outcome can be attained with a performance-based contract. If either party is risk averse, then a contract with a fixed payment, a cost-sharing incentive, and a performance-based incentive can attain the second-best outcome. In a related paper, Kim et al. (2009), study the trade-offs in the decisions of a vendor (an original equipment manufacturer) that are made during two different phases of a product's life cycle: development and exploitation. During the *development* phase, the vendor invests in R&D that improves product reliability; during the *exploitation* phase, the investment she makes in spare-parts inventory determines the after-sale repair service level. These two decisions jointly determine the total availability of product in the exploitation phase. The authors find that a performance-based contract is preferable to a material contract and that investment in product reliability is increasing in vendor's ownership of the spare-parts inventory. Bakshi et al. (2011) study the role of performance-based incentives attaining economically efficient outcomes when the failure rate of the equipment is exogenous and the manufacturer has private information on the failure rate. Bakshi et al. (2011) show that when spare part inventory is verifiable performance-based contracts can be used as a signaling mechanism by the informed manufacture leading to a separating equilibrium that yields the first-best outcome.

This paper differs in several aspects from the studies cited previously. First, we focus on the joint role of the client and the vendor in ensuring availability of the equipment. Second, we limit our focus to the exploitation phase of the product life cycle. Third, we abstract from specific

decisions that determine the equipment's failure (e.g., frequency of regular upkeep) and repair rate (e.g., spare-parts inventory level) and focus instead on the investments made in operating and repair technologies by the client and the vendor, respectively. Fourth, we use a double-sided moral hazard framework to study simultaneous decisions by client and vendor that are contingent on the contractual terms offered by the client.

This paper is also related to the double-sided moral hazard literature in economics and operations management. In economics, different contract structures (e.g., revenue sharing, shared savings, buyout agreements, option-based contracts) have been studied for resolving the double-sided moral hazard between two parties. Bhattacharyya and Lafontaine (1995) study a joint production problem with risk-neutral parties and show that a two-part linear contract with a variable revenue-sharing component and a fixed-fee part is optimal and attains the second-best outcome. In a related study, Kim and Wang (1998) show that with a risk-averse agent the linear contract structure is not optimal; furthermore, the optimal nonlinear contract attains only a second-best outcome.

In the operations management literature, there are a few studies that apply the double-sided moral hazard framework to contracting issues. Baiman et al. (2000) study joint efforts of a risk-neutral customer and a supplier to improve product quality and reduce external failures. They examine fixed-price contracts and show that the first-best outcome can be attained under two special conditions: (i) the decisions of the two parties can be directly contracted upon; and (ii) the contract can be based on measures that allow the separability of the profit functions in the respective decisions. Such separation allows for the reduction of a double-sided moral hazard problem to two single-sided moral hazard problems, which in turn can be resolved for risk-neutral parties. Corbett and DeCroix (2001) study the consumption of indirect materials in a supply chain with risk-neutral parties. They show that, under cost functions that are increasing and convex, the customer cannot attain the first-best outcome. In a related study, Corbett et al. (2005) show that if the first-best outcome requires positive efforts from at most one of the two parties, then the customer can attain the first-best outcome. More recently, Roels et al. (2010) have studied contracting issues in the outsourcing of collaborative services. They show that first-best decisions can be attained

by incurring the costs of monitoring and verifying the other party's efforts. However, the profits earned by the parties in this case are less than first best owing to the additional cost of monitoring. In the sequential decision setting of an R&D partnership between a pharmaceutical firm and a biotech firm that are both risk neutral, Bhattacharya et al. (2011) show that the first-best outcome can be attained through renegotiation, late-stage contracting, and option mechanisms. They also demonstrate that if the biotech firm is risk averse, then the first-best outcome can be attained by timing the offer or exercise of the option contract. Our work differs from these studies in that we study *simultaneous* investment decisions by the client and the vendor. We study a linear contract structure that differs from revenue sharing or shared savings because it penalizes the vendor based on her realized performance. We also study the design and performance of optimal contracts that not only provide sufficient incentives to resolve double-sided moral hazard but also ensure that a vendor's exposure to financial distress is lower than her exogenous threshold. We show that, within the two classes of contracts studied in this paper, the client can attain first-best profits by customizing contractual terms to fit the vendor's characteristics.

A relatively new area of research in operations management addresses how a firm's contractual decisions affect the financial health of its suppliers and service providers. In a survey of automakers and their suppliers, Choi and Hartley (1996) find that the financial health of suppliers is an important selection criterion. Swinney and Netessine (2009) is among the first papers to study the relationship between a firm's contractual choices and its endogenous effect on a supplier's bankruptcy. These authors study a buyer's choice between offering a short-term and a long-term contract (in a two-period model) to a supplier facing the risk of bankruptcy. The supplier declares bankruptcy if the random production costs for any period are higher than the contractual payments for that period. Swinney and Netessine (2009) show that the buyer's preference for a long-term contract is increasing in the buyer's switching cost between suppliers. The long-term contract allows the buyer to make an up-front payment and thereby reduce the probability of supplier bankruptcy. In a related paper, Babich (2010) studies the role of subsidies with a cost-plus contract in managing

a supplier's performance. The supplier's bankruptcy is a function of endogenous contractual payments and exogenous cash flows from the supplier's other businesses. The buyer can influence the supplier's bankruptcy risk by making up-front payments (subsidies) in addition to the supplier's up-front cost requirement. Dong and Tomlin (2011) study the joint role of business interruption insurance and operational measures to mitigate losses due to disruptions in the supply chain. We contribute to this literature in several ways. First, we model the financial health of the vendor as being purely endogenous to contractual payments. Second, we capture the vendor's potential financial distress by modeling the stochastic evolution of her cumulative wealth. This allows us to capture the long-term, one-to-one relationship between client and vendor as well as the evolution of vendor's wealth in terms of stochastic earnings (losses) across service calls. In this paper we demonstrate that the buyer can exploit contract features—other than up-front subsidies—that affect the supplier's financial health.

3. Model Setup

We study a profit-maximizing client who outsources the repair and restoration services of his equipment to a vendor. We assume that such an outsourcing environment is exogenous to our problem setting, and we assume a long-term contractual relationship between client and vendor that continues throughout the equipment's life cycle. To retain our central focus on the challenging issues of contracting and to achieve analytical tractability, we assume that the client has only one piece of equipment.

The equipment can be in only one of two states: operational or failed. We refer to the time period between two consecutive failures as *uptime* (U) and to the time spent in restoring the failed equipment to its operational state is referred as *downtime* (D). We also define the combination of a span of uptime and the subsequent downtime as the equipment's *operational cycle* (OC); see Figure 1. In this paper we often refer to the period of realized downtime as a *service call* for clarity of exposition. We assume that the operational performance of the equipment is independent across operational cycles (i.e., that the realized uptimes and downtimes are independent across operational cycles). However, the realized uptime and downtime *within* a cycle may be correlated.

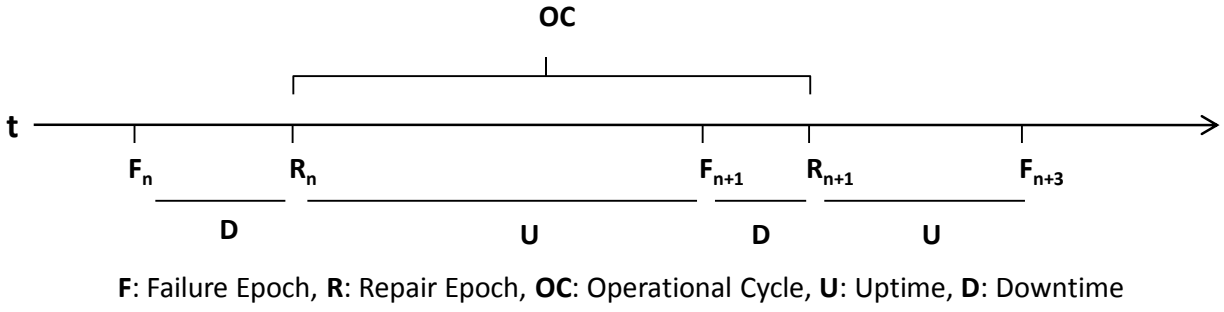


Figure 1 Operational Cycle of the Equipment

The equipment's failure rate λ is determined by static decisions that are made by the customer with a long-term view. For example, a client typically invests in aspects of operating technology—such as the equipment's operating environment—with an eye to the long term. Likewise, committing resources for the training of equipment operators and carrying out routine checks and maintenance activities[9] are made with a long-term view. The repair rate μ of the equipment is similarly determined by static decisions made by the vendor. A vendor typically takes a long term view when making decisions on the hiring and training of engineers. We abstract all such decisions made by the client and the vendor as investments in operating and repair technologies, respectively. In this problem setting we do not study opportunistic dynamic decisions—such as reducing the quality of repair during a particular repair cycle (cf. Murthy 1991), inducing frequent failures, or reducing the care in the usage of the equipment during an operating cycle—which may also influence system performance.

We define a measure of equipment reliability τ that is equal to the mean time between failures $(1/\lambda)$. The client, through his investments, determines the reliability of the equipment within a bounded range $[\underline{\tau}, \overline{\tau}]$; the bounds reflect the reliability of current technologies used in the equipment. The vendor, through her investments, determines the repair rate within a bounded range $[\underline{\mu}, \overline{\mu}]$, which reflects the effectiveness of current repair methodologies. We assume that $\underline{\tau} > 1/\underline{\mu}$; this is consistent with the observed operational behavior of equipment in real-world settings, where the equipment's average uptime $(1/\lambda \equiv \tau)$ is typically greater than the average downtime. We also

assume that the failed equipment is restored to its original state (i.e., to reliability level τ) after every repair. The up-front investments by the customer and the vendor determine (respectively) the equipment's failure and repair rates, which remain constant throughout its (presumably infinite) life cycle. Given these assumptions, we build our model following the conventional machine repairman model, which is commonly used to analyze repair-related problems.

We assume that the client generates revenue from continued usage of the equipment, which is subject to random breakdowns. The client earns a revenue r (per unit of time) during uptime and incurs an opportunity cost o (also per unit of time) during downtime. The opportunity cost captures the loss of customer goodwill, the cost of losing a customer to competition, and in some cases the institutional penalties levied on the client for disruption of service to customers.[10] The vendor earns revenue from the contractual payments made by the client for her repair services.

We model the investment costs in operating and repair technologies as the general cost functions $c_u(\tau)$ and $c_d(\mu)$, respectively. These cost functions are increasing and convex ($c_d'(\mu) > 0,\ c_u''(\tau) > 0,\ c_d'(\mu) > 0,\ c_d''(\mu) > 0$). We normalize the cost functions so that the investment required to achieve the equipment's lowest reliability $\underline{\tau}$ and lowest repair rate $\underline{\mu}$ is zero. The upper bounds on the reliability $\overline{\tau}$ and the repair rate $\overline{\mu}$ represent the theoretical achievable levels, so we assume that the cost of achieving these theoretical levels is infinite. Thus, we have $c_u(\underline{\tau}) = 0$, $\lim_{\tau\to\underline{\tau}} c_u'(\tau) = 0$, $\lim_{\tau\to\overline{\tau}} c_u(\overline{\tau}) = \infty$, $\lim_{\tau\to\overline{\tau}} c_u'(\overline{\tau}) = \infty$, $c_d(\underline{\mu}) = 0$, $\lim_{\mu\to\underline{\mu}} c_d'(\underline{\mu}) = 0$, $\lim_{\mu\to\overline{\mu}} c_d(\overline{\mu}) = \infty$, and $\lim_{\mu\to\overline{\mu}} c_d'(\overline{\mu}) = \infty$. These conditions on the cost structure imply that an interior point solution exists for a profit maximization problem. Note that these cost functions represent the "per unit time" investment made by each party over the contract period, although both parties commit up front to these static investments. The constant investment rates can be mapped to any lump-sump, up-front investment by selecting an appropriate discount rate.

The sequence of events in our setting is as follows. The client offers a take-it-or-leave-it contract to the vendor. Given the contractual terms, the client and the vendor make simultaneous unverifiable investments in the operating and repair technologies, respectively. The decisions made by the players are outcome of a simultaneous investment game. In this investment game, the client's aim

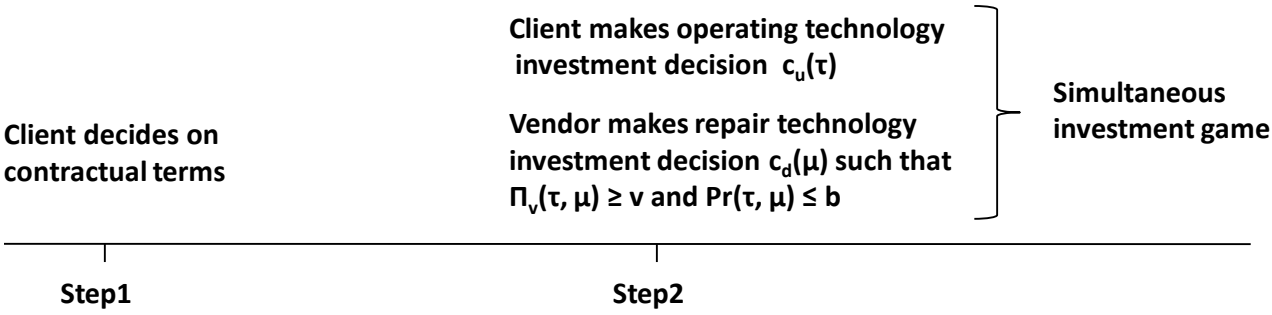


Figure 2 Two-Step Sequential Decision Making

is to maximize his profit rate Π_c and the vendor's aim is to maximize her profit rate Π_v subject to her individual rationality (reservation value) and financial distress constraint. Note that the client will determine the optimal contract by using backward induction constrained by the outcome of this simultaneous investment game. We formalize the setting as a contract design optimization problem in Section 4.2; see equations (8)–(12). Figure 2 captures this two-step decision-making process, which is embedded in a typical double-sided moral hazard problem.

3.1. Performance-Based Contract

We analyze the challenges of outsourcing repair and restoration services from the client's perspective. The client designs and offers a performance-based contract to the vendor with the objective of maximizing his expected profit or, equivalently, the profit rate (i.e., profit earned per unit of time) over the contract period.

Without loss of generality, a performance-based contract can be split into a fixed component and a performance-linked variable component. In our setting, the fixed component (w) captures payments that are made per unit of time, while the variable component $T(D)$ captures the performance-linked compensation to the vendor, where D is the realized downtime (vendor's performance) during a service call. It is intuitive that, in order to align the incentives of the vendor, the client will choose a structure for the optimal $T(D)$ whereby the vendor is penalized for longer downtimes. We restrict our analysis to the following two forms of $T(D)$:

$$T(D) = \begin{cases} f - pD & \text{under a linear contract,} \\ \bar{f} - \bar{p} I_{D \geq \bar{d}} & \text{under a tiered contract,} \end{cases}$$

where f, $\bar{f}$ are the fixed payments made during a service call and p, $\bar{p}$ are the penalty parameters under the two contracts. The linear contract in our paper is similar to the performance-based contracts studied in Kim et al. (2009). The distinction between linear and tiered contracts is the penalty structure levied against the vendor during a service call, which in turn determines the variable component $T(\cdot)$ during that service call. Under a linear contract, the penalty levied in the ith service call is *proportional* to the realized performance D_i. Under a tiered contract,[11] a *constant* penalty is levied whenever the realized performance D_i is greater than the performance threshold $\bar{d}$.

The individual equilibrium profit rate functions of the client (Π_c) and the vendor (Π_v) in their respective efforts are as follows:[12]

$$\Pi_c = \lim_{t \to \infty} \frac{E\left[\sum_{i=1}^{N_t} rU_i - oD_i - T(D_i)\right]}{t} - c_u(\tau) - w, \tag{1}$$

$$\Pi_v = \lim_{t \to \infty} \frac{E\left[\sum_{i=1}^{N_t} T(D_i)\right]}{t} + w - c_d(\mu); \tag{2}$$

here N_t is a random variable that captures the number of operational cycles completed up to time t, and U_i and D_i denote (respectively) the uptime and downtime realized during the ith operational cycle. Equations (1) and (2) imply that the profit rate functions Π_c and Π_v are determined by the joint efforts of the client and the vendor because the realization of N_t is contingent on both τ and μ. This joint dependence creates a double-sided moral hazard that results in incentives for both the vendor and the client to underinvest (free-rider effects—see Holmstrom 1982).

3.2. Model for Induced Financial Distress

The downtime during each service call is stochastic and hence a performance-based contract leads to stochastic earnings for the vendor. These stochastic earnings may deplete the cumulative wealth of the vendor to undesirable levels, exposing her to financial distress. As discussed in Hendricks and Singhal (2005) (p. 36), the probability of financial distress is an important metric for the various

stakeholders of a firm, including investors, managers, customers etc. Consequently the vendor, when making her profit-maximizing investment decision, ensures that her exposure to financial distress does not exceed her tolerance threshold.

In this section we first describe a stochastic wealth model that captures the dynamic wealth accumulation of the vendor. Then we use this stochastic model to compute the vendor's exposure to financial distress. We assume that the vendor has initial wealth w_o that evolves stochastically. During an operational cycle, the vendor earns revenue (cash inflow) in the form of contractual payments made by the client for her repair services; the vendor's cost (cash outflow) consists of the investment rate $c_d(\mu)$ and the performance-linked penalty, if any. We assume that the vendor does not reinvest her accumulated wealth and so earns no interest on it. Given these assumptions, we can write the process of the vendor's wealth evolution under the offered contract $(w, T(D))$ as

$$w_t = w_o + (w - c_d(\mu))t + \sum_{i=1}^{N_t} T(D_i)\,, \tag{3}$$

where w_t denotes the wealth of the vendor at time t. Note that the realization of $T(D_i)$ across operational cycles can be captured by a sequence of random variables that are independent and identically distributed. This implies that the wealth level w_t is a random sum (N_t) of random variables $(T(D_i))$. The distribution of these random variables is determined by (i) the contractual terms offered by the client to the vendor and (ii) the Nash Equilibrium outcome of the simultaneous investment game.

Similar stochastic wealth models are employed extensively in the literature of actuarial science to study the wealth evolution of insurance firms (Asmussen and Albrecher 2000). An insurance firm receives a continuous cash inflow per unit time from the premium it charges to policyholders, and it is subject to stochastic cash outflows when the insured make claims. Note that our vendor's wealth model, equation (3), maps one-to-one to the wealth model of an insurance firm.[13] In our model the transfer payment $T(D)$ for a service call occurs at the completion of an operational cycle.

Our vendor's wealth model, however, differs from wealth models of insurance firms in one important aspect. Our model allows the distribution of $T(D)$ or "claims" to have positive support. The support of $T(D)$ is $(-\infty, f]$, where $T(D) > 0$ implies a cash inflow for the vendor during a service call. The vendor will see such instances when her realized performance is such that the induced penalty levied is less than the fixed payment. In contrast, the distribution of claims has only non-positive support $(-\infty, 0)$; in other words, all claims represent cash outflow for the insurance firm. Following Iglehart (1969), we approximate the vendor's wealth process as a Brownian motion with drift $\mu_{\text{BM}} = w - c_d(\mu) + E[T(\cdot)]\frac{\mu}{1+\tau\mu}$ and variance $\sigma^2_{\text{BM}} = V[T(\cdot)]\frac{\mu}{1+\tau\mu}$.

We define the induced *financial distress level,* contingent on offered contractual terms, as the probability of the vendor's wealth dropping to zero before completion of the contract period (i.e., at some time $\alpha < \infty$). In other words, α denotes the time span between the start of a contractual agreement and the first instance of the vendor's wealth falling to (or below) zero. In the actuarial science literature such a time span α is referred to as the time to "ruin", and the probability of $\alpha < \infty$ is used as the relevant metric for the insurance firm's financial health (Asmussen and Albrecher 2000). Conceptually, the vendor's wealth dropping to zero captures the phenomenon that, as the cash reserves of a firm decrease and approach zero, the firm's capacity to sustain operations becomes limited; that is the firm enters a state of financial distress.

We use the Brownian motion approximation and write the vendor's induced financial distress level as

$$\Pr(\tau, \mu, w, T(D)) = \Pr(\alpha < \infty) = \exp\left\{-2w_o \frac{(w - c_d(\mu))(1+\tau\mu) + E[T(D)]\mu}{V[T(D)]\mu}\right\}. \tag{4}$$

We model the vendor's problem as a profit maximization problem subject to the constraint that the induced financial distress level $\Pr(\tau, \mu, w, T(D))$ does not exceed the distress threshold $\mathbf{b}$. We assume that the vendor's distress threshold $\mathbf{b} \in (0, 1]$ is determined exogenously and reflects the vendor's financial state and/or business priorities. For example, a service provider with a low debt-to-equity ratio will have a relatively high distress threshold compared to a service provider who is highly leveraged (i.e., from excessive debt borrowing) and therefore has limited access to external

financial resources. For a large service provider that commits client-specific investments and builds a dedicated team, this threshold captures the vendor's priority of engaging with businesses that meet certain tolerance threshold criteria for induced financial distress level. Usually such tolerance threshold decisions are made by senior management and line managers use these guidelines to make investment decisions and to accept or reject business proposals.

4. Analysis

To benchmark the performance of the two contract structures, we first determine the optimal centralized investments in the operating and repair technologies. In the rest of this paper, we use an asterisk ($*$) to denote the profit and decisions of the centralized entity.

4.1. Optimal Centralized Investments

Given the model's description in Section 3, we can write the centralized profit rate function Π as,

$$\Pi = \lim_{t\to\infty} \frac{E\left[\sum_{i=1}^{N_t} rU_i - oD_i\right]}{t} - c_u(\tau) - c_d(\mu)\,. \tag{5}$$

The following proposition characterizes the optimal centralized investments in the operating and repair technology—in other words, the investments that maximize Π. Note that investments in operating and repair technology $\{c_u(\tau), c_d(\mu)\}$ uniquely map into the decisions $\{\tau, \mu\}$. Henceforth we formulate the decision spaces of the client and the vendor as $\tau \in [\underline{\tau}, \overline{\tau}]$ and $\mu \in \left[\underline{\mu}, \overline{\mu}\right]$, respectively.

Proposition 1. *The optimal centralized decision $\{\tau^*, \mu^*\}$ is the unique solution to the following system of simultaneous equations:*

$$\frac{(r+o)\mu}{(1+\tau\mu)^2} = c_u'(\tau)\,, \tag{6}$$

$$\frac{(r+o)\tau}{(1+\tau\mu)^2} = c_d'(\mu)\,. \tag{7}$$

Proof. All omitted proofs are given in the Appendix. ■

Proposition 1 characterizes the unique interior solution in the $\mathbb{R}_+^2$ space defined by $(\underline{\tau}, \overline{\tau}) \times (\underline{\mu}, \overline{\mu})$. We denote the centralized optimal profit rate by Π^*, which is achieved by implementing the decision $\{\tau^*, \mu^*\}$. When the client outsources the repair services to the vendor, the implicit decentralized service supply chain can achieve a profit rate of no more than Π^*.

Proposition 2. *Investments in operating technology or repair technology are substitutes for each other.*

Proposition 2 implies that investments in operating and repair technologies are determined by the relative marginal costs of implementing them. More importantly, Proposition 2 indicates that both the client and the vendor have incentives to underinvest in their respective decisions. Since the two investments are substitutable, each party has incentives to free ride on the other party's effort, resulting in some loss of service supply chain profits. In the next section we analyze the decentralized setting.

4.2. The Client's Problem under Decentralized Decision Making

We formulate the client's problem under decentralized decision making—in terms of a principal–agent model featuring double-sided moral hazard (Bhattacharyya and Lafontaine 1995, Roels et al. 2010)—as follows:,

$$\max_{T(\cdot),w} \Pi_c(\tau', \mu', w, T(\cdot)) \tag{8}$$

$$\text{s.t.} \quad \tau' = \arg\max_{\underline{\tau} \le \tau \le \overline{\tau}} \Pi_c(\tau, \mu', w, T(\cdot)), \tag{9}$$

$$\mu' = \arg\max_{\underline{\mu} \le \mu \le \overline{\mu}} \Pi_v(\tau', \mu, w, T(\cdot)) \tag{10}$$

$$\text{s.t.} \quad \Pi(\tau', \mu', w, T(\cdot)) \ge \mathbf{v}, \tag{11}$$

$$\Pr(\tau', \mu', w, T(\cdot)) \le \mathbf{b}. \tag{12}$$

Here Π_c and Π_v are (respectively) the client's and vendor's profit rate functions as defined by equations (1) and (2). To avoid trivial solutions, we assume that $\mathbf{v} \in (0, \Pi^*]$. Also, we model the distribution of equipment uptime and downtime with exponential distributions in order to gain analytical tractability for computing the potential financial distress induced by offered contractual terms.

Equations (9) and (10) represent the simultaneous investment game subject to the vendor's participation constraint (inequality (11)) and her financial distress constraint (inequality (12)).

Observe that the theoretical maximum profit that the client could earn in the decentralized case is attained by a contract that (i) induces the client and the vendor to make investments equal to the optimal centralized decisions (thus rendering moot the free-rider issue), (ii) makes the participation constraint of the vendor "tight", and (iii) satisfies the financial distress constraint for the vendor. If there exists a $\{T(\cdot), w\}$ that satisfies these conditions (i)–(iii), then such a contract is called the *first-best* contract. While the feasible set of $\{T(\cdot), w\}$ in the client's problem is infinite, as we show below, he can attain the first-best outcome within the class of linear and tiered contract structures for all feasible values of exogenous parameters. More specifically, the client can design a first-best contract by limiting his search space to $\{T(\cdot), w\} \in \{\mathcal{L} \cup \mathcal{T}, \mathbb{R}\}$, where $\mathcal{L}$ is the class of linear contracts, $\mathcal{T}$ is the class of tiered contracts, and $\mathbb{R}$ is the set of real numbers. Next we shall analyze each class of contracts in turn.

4.2.1. Linear Contractual Structure Recall that the variable component of a linear contract is

$$T(D) = f - pD, \tag{13}$$

where f is the fixed payment made during the service call and p is the penalty rate. At the end of the service call i, the client levies a penalty (pD_i) that is proportional to the vendor's realized performance D_i for that service call.

Lemma 1. *A linear contract $\{w, f, p\}$ attains the first-best outcome for the client only if*

(a) $p = f \cdot \mu^*$;

(b) $w = c_d(\mu^*) + \boldsymbol{v}$.

Lemma 1 describes the necessary conditions for designing a linear contract that can attain the first-best outcome for the client. By applying Lemma 1, we can reduce the client's problem of searching for a first-best contract (within the class of linear contracts) into a single decision $\{p\} \in \mathbb{R}^+$.

We find that the client can attain first-best outcome with a linear contract if the vendor's exogenous parameters $\{\mathbf{v}, \mathbf{b}\}$ are within a particular range. In the next proposition we characterize the set of parameters for which the linear contract attains the first-best outcome.

Proposition 3.

(a) For a vendor with reservation value $\boldsymbol{v} \in \left[\frac{(r+o)\tau^*\mu^*}{(1+\tau^*\mu^*)(1+2\tau^*\mu^*)}, \Pi^*\right]$*, the client:*

(i) can attain the first-best outcome by designing a unique linear contract

$$\{w, f_1, p_1\} = \left\{c_d(\mu^*) + \boldsymbol{v}, \frac{(r+o)\tau^*}{1+\tau^*\mu^*}, \frac{(r+o)\tau^*\mu^*}{1+\tau^*\mu^*}\right\} \quad \textit{if } \boldsymbol{b} \in \left[\exp\left\{-\frac{2w_o\boldsymbol{v}(1+\tau^*\mu^*)^3}{(r+o)^2\tau^{*2}\mu^*}\right\}, 1\right];$$

(ii) can attain the first-best outcome by designing a unique linear contract

$$\{w, f_2, p_2\} = \left\{c_d(\mu^*) + \boldsymbol{v}, \left(\frac{2w_o\boldsymbol{v}(1+\tau^*\mu^*)}{\log(1/\boldsymbol{b})\mu^*}\right)^{1/2}, \left(\frac{2w_o\boldsymbol{v}(1+\tau^*\mu^*)\mu^*}{\log(1/\boldsymbol{b})}\right)^{1/2}\right\}$$
$$\textit{if } \boldsymbol{b} \in \left(0, \exp\left\{-\frac{2w_o\boldsymbol{v}(1+\tau^*\mu^*)^3}{(r+o)^2\tau^{*2}\mu^*}\right\}\right].$$

(b) For vendors with the reservation value $\boldsymbol{v} \in \left(0, \frac{(r+o)\tau^*\mu^*}{(1+\tau^*\mu^*)(1+2\tau^*\mu^*)}\right)$*, the client:*

(i) can attain the first-best outcome with a unique linear contract

$$\{w, f_1, p_1\} = \left\{c_d(\mu^*) + \boldsymbol{v}, \frac{(r+o)\tau^*}{1+\tau^*\mu^*}, \frac{(r+o)\tau^*\mu^*}{1+\tau^*\mu^*}\right\} \quad \textit{if } \boldsymbol{b} \in \left[\exp\left\{-\frac{2w_o\boldsymbol{v}(1+\tau^*\mu^*)^3}{(r+o)^2\tau^{*2}\mu^*}\right\}, 1\right];$$

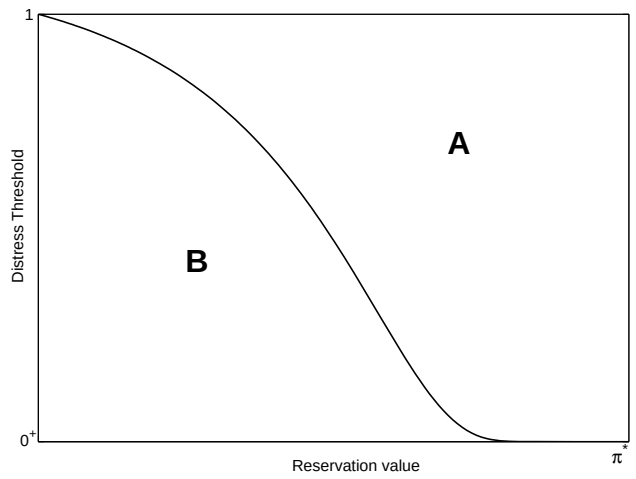
(ii) can attain the first-best outcome with a unique linear contract

$$\{w, f_2, p_2\} = \left\{c_d(\mu^*) + \boldsymbol{v}, \left(\frac{2w_o\boldsymbol{v}(1+\tau^*\mu^*)}{\log(\frac{1}{\boldsymbol{b}})\mu^*}\right)^{1/2}, \left(\frac{2w_o\boldsymbol{v}(1+\tau^*\mu^*)\mu^*}{\log(\frac{1}{\boldsymbol{b}})}\right)^{1/2}\right\}$$
$$\textit{if } \boldsymbol{b} \in \left[\exp\left\{-\frac{2w_o\boldsymbol{v}(1+\tau^*\mu^*)^3\mu^*}{((r+o)\tau^*\mu^* - \boldsymbol{v}(1+\tau^*\mu^*)(1+2\tau^*\mu^*))^2}\right\}, \exp\left\{-\frac{2w_o\boldsymbol{v}(1+\tau^*\mu^*)^3}{(r+o)^2\tau^{*2}\mu^*}\right\}\right);$$

(iii) cannot attain the first-best outcome under a linear contract structure

$$\textit{if } \boldsymbol{b} \in \left(0, \exp\left\{-\frac{2w_o\boldsymbol{v}(1+\tau^*\mu^*)^3\mu^*}{((r+o)\tau^*\mu^* - \boldsymbol{v}(1+\tau^*\mu^*)(1+2\tau^*\mu^*))^2}\right\}\right).$$

Proposition 3 broadly classifies vendors into two groups based on their reservation values $\mathbf{v}$ and distress thresholds $\mathbf{b}$. A vendor belongs to the first group (region A of Figure 3) if she has a high reservation value and/or a high distress threshold. With such a vendor the client can attain the first-best outcome by offering a customized linear contract in which contractual terms are appropriately adjusted according to characteristics $\{\mathbf{v}, \mathbf{b}\}$ of the vendor. In contrast, a vendor belongs to the second group (region B of Figure 3) if she has a low reservation value and/or a low distress threshold. With such a vendor, the client cannot attain the first-best outcome by offering a linear contract.



	Contract Structure			
	Linear		Tiered	
Region	**A**	**B**	**A**	**B**
Contract Efficiency	First Best	No First Best	First Best	

Figure 3 First-Best Region for Linear and Tiered Contract Structures

Proposition 3 shows that linear contracts can attain the first-best outcome only for a subset of parameter values. This result complements the extant literature that studies double-sided moral hazard. In economics, Bhattacharyya and Lafontaine (1995) and Kim and Wang (1998) show that attaining the first-best outcome in a double-sided moral hazard environment is not feasible under general conditions. In the operations management literature, it has been shown that the first-best outcome can be attained under three special conditions: (i) when the decisions of both parties are verifiable, which allows the contract to be based directly upon those decisions (Baiman et al. 2000); (ii) when there exists a contractible performance measure that can reduce the profit functions of the two parties to endogenous functions of their respective decisions (Cooper and Ross 1985, Baiman et al. 2000); (iii) when the first-best outcome require positive effort from at most one of the parties (Corbett et al. 2005). Condition (ii) would be satisfied in our setting by a contract based on a hypothetical performance measure that reduces the profit rate functions to $\Pi_c = g_c(\tau)$ and $\Pi_v = g_v(\mu)$. Such reduction would convert the double-sided moral hazard problem to a single-sided moral hazard problem that, in turn, could be completely resolved. Similarly condition (iii)

implicitly reduces the double-sided moral hazard problem to a single-sided one.

None of these three special conditions apply in our setting. First, neither the client's investments in operating technology nor the vendor's investment in repair technology is verifiable in a court of law; hence no enforceable contract can be based on those investment levels. Second, the linear performance-based contract does not reduce the profit rate functions of the client and the vendor in terms of their respective decisions τ and μ (see equations (1) and (2)). Third, in our setting the optimal decision is a interior point in the *joint* decision space of client and vendor; thus both parties must exert positive efforts in order to attain the first-best outcome. In sum: because these special conditions do not apply in our setting, the results described by Proposition 3 add to the existing operations literature.

To understand what empowers a linear contract to attain the first-best outcome with the vendors in region A, one should understand the interaction between the "fixed fee per service call" component f and the penalty rate p. The term f plays the dual role of carrot and stick for vendor and client, respectively. The fixed component per unit time w covers the investment cost $c_d(\mu^*)$ and the reservation value $\mathbf{v}$, so f provides an opportunity for the vendor to gain additional profits. At the same time, if the client underinvests in τ then the failure frequency will increase, and with each failure the client must pay f in addition to the per unit time payment w. The term p serves as a stick in that it discourages the vendor from underinvesting.

It is straightforward to explain the linear contract structure's inability to attain the first-best outcome in region B. The intuition is that, if the vendor has a high reservation value (or, equivalently, a large share in the supply chain profit) then she has strong incentives to make the optimal decision. However, if the vendor has a low reservation value then a linear penalty exposes her to higher level of financial distress, which in turn may incentivize her to make suboptimal investments. This implies that vendors with low reservation value must trade off between investment decisions and the associated potential financial distress levels. Therefore, unless vendors with low reservation value have a strong appetite for potential financial distress, a linear contract will not attain the first-best outcome.

Next we show that the client can use linear contracts to incentivize a vendor in region B to implement the optimal centralized decision μ^* by paying her a surplus to exceed her reservation value. Contracts that are so devised to attain system-optimal decisions are known as *coordinating* contracts. Of course, even though such contracts attain the optimal system performance, they are not first best for the client because he must now pay the vendor a surplus. Corollary 1 presents the analytical expression for the profit rate earned by a region B vendor, $\Pi_v(> \mathbf{v})$, under the coordinating linear contract. By paying a profit-rate Π_V higher than the vendor's reservation value $\mathbf{v}$, the client reduces the vendor's trade-off between making system-optimal investments and limiting her exposure to financial distress.

Corollary 1. *The client can attain first-best decisions with a vendor* $\{\boldsymbol{v}_B, \boldsymbol{b}_B\}$ *that belongs to region B by paying a her a profit rate*

$$\Pi_v = \left(\frac{-(1+\tau^*\mu^*)\sqrt{2w_o\mu^*} + \sqrt{2w_o(1+\tau^*\mu^*)^2\mu^* + 4\log(1/\boldsymbol{b}_B)(2\tau^*\mu^*+1)(r+o)\tau^*\mu^*}}{2(2\tau^*\mu^*+1)\sqrt{(1+\tau^*\mu^*)\log(1/\boldsymbol{b}_B)}} \right)^2 > \boldsymbol{v}_B$$

The results of Proposition 3 have important implications. Linear contracts are intuitive and easy to implement, yet they are seldom observed in practice. One possible explanation for this is that most vendors may actually belong to region B; in other words vendors typically have a low reservation value and/or a low financial distress threshold. Since the linear contracts do not attain the first-best outcome in region B, there may exist a class of contracts that outperform linear contracts in this region. As we show in the following section, tiered contracts outperform linear contracts in region B. In fact, we find that the tiered contract structure attains the *first-best* outcome for the client in region B *and* in region A.

4.2.2. Tiered Contractual Structure Contracts in repair and restoration settings are seldom linear; instead they levy a limited penalty on the vendor if her performance deteriorates beyond a threshold level (see Table 1). Often such contracts have multiple performance brackets (tiers), each with a corresponding fixed penalty that is levied when the vendor's realized performance falls within that bracket. In this paper (as discussed in Section 3.1) we analyze a simple

one-tier contract structure under which the client proposes to levy a constant penalty $\bar{p}$ only if the vendor's realized performance D deteriorates beyond a threshold level $\bar{d}$ during a service call:

$$T(D) = \bar{f} - \bar{p} I_{D \geq \bar{d}}, \tag{14}$$

here $I_{D\geq\bar{d}}$ is a indicator function on $\bar{d}$ (i.e., $I_{D\geq\bar{d}} = 1$ if $D \geq \bar{d}$ and 0 otherwise), $\bar{f}$ is the fixed fee paid per service call, and $\bar{p}$ is the constant finite penalty levied after service call i if $D_i \geq \bar{d}$.

Proposition 4. *The client can attain the first-best outcome with a tiered contractual structure for all* $\boldsymbol{v} \in (0, \Pi^*]$ *and for all* $\boldsymbol{b} \in (0, 1]$. *Specifically, the client:*

(a) can attain the first-best outcome with a tiered contract

$$\{w, \bar{f}_3, \bar{p}_3, \bar{d}\} = \left\{ c_d(\mu^*) + \boldsymbol{v}, \frac{c_d'(\mu^*)(1+\tau^*\mu^*)}{\bar{d}\mu^*}, \frac{c_d'(\mu^*)(1+\tau^*\mu^*)}{e^{-\mu^*\bar{d}}\bar{d}\mu^*}, \frac{2}{\underline{\mu}} \right\}$$

if $\boldsymbol{b} \in \left[\exp\left\{ -\frac{8w_o \boldsymbol{v}\mu^*}{c_d'(\mu^*)^2(1+\tau^*\mu^*)(e^{2\mu^*/\underline{\mu}}-1)\underline{\mu}^2} \right\}, 1 \right]$;

(b) can attain the first best outcome with a tiered contract

$$\{w, \bar{f}_4, \bar{p}_4, \bar{d}\} = \left\{ c_d(\mu^*) + \boldsymbol{v}, \left(\frac{2w_o \boldsymbol{v}(1+\tau^*\mu^*)}{log(1/\boldsymbol{b})\mu^*\left(e^{\mu^*\bar{d}}-1\right)} \right)^{1/2}, \left(\frac{2w_o \boldsymbol{v}(1+\tau^*\mu^*)}{log(1/\boldsymbol{b})\mu^*\left(e^{-\mu^*\bar{d}}-e^{-2\mu^*\bar{d}}\right)} \right)^{1/2}, \bar{d} \right\}$$

if $\boldsymbol{b} \in \left(0, \exp\left\{ -\frac{8w_o v\mu^*}{c_d'(\mu^*)^2(1+\tau^*\mu^*)(e^{2\mu^*/\underline{\mu}}-1)\underline{\mu}^2} \right\}\right)$, *where* $\bar{d} \geq \frac{2}{\underline{\mu}}$ *and satisfies*

$$\bar{d}\left(1 - \frac{1}{(e^{\mu^*\bar{d}}-1)}\right) \geq \frac{\mu^*}{\boldsymbol{v}(1+\tau^*\mu^*)}(c_d(\mu^*) + \boldsymbol{v} - c_d(\mu))\frac{1}{\mu^2} + c_d'(\mu)\left(\frac{1}{\mu} + \tau^*\right) \quad \forall \mu \in [\underline{\mu}, \mu^*]. \tag{15}$$

Proposition 4 shows that tiered contracts not only outperform linear contracts but also enable the client to attain the first-best outcome. This adds to the literature on double-sided moral hazard by showing that the first-best outcome can be attained for *all* possible parameter values. The result is surprising since, for a vendor in region B, one would not expect that the client could simultaneously incentivize the vendor to make the optimal centralized investment, satisfy her financial distress constraint, and also make her participation constraint tight.

The results in Proposition 4 are particularly interesting because their intuition is not straightforward. Intuitively, the limited penalty structure of tiered contracts might have a countervailing effect: while reducing the vendor's exposure to financial distress, it may also reduce her incentives to invest optimally in the repair technology. In particular, with a tiered contract, no penalty is levied unless the vendor's performance deteriorates beyond a threshold level $\bar{d}$. This contract term

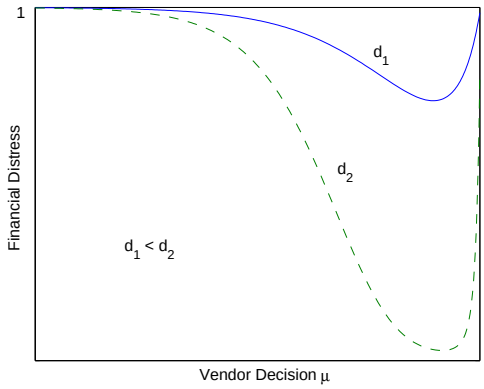


Figure 4 Financial Distress Contingent on $\bar{d}$

reduces the vendor's probability of earning negative profits in consecutive operational cycles. However, it also limits the vendor's downside to underinvesting in repair technology and so may not eliminate the free-rider aspect of the double-sided moral hazard.

The key to understanding the ability of tiered contracts to attain the first-best outcome is to appreciate the flexibility provided by the component $\bar{d}$. This term allows the client to control for the sensitivity of the vendor's decision μ to potential financial distress (Figure 4) while keeping the terms $\bar{f}$ and $\bar{p}$ constant. This extra degree of freedom allows the client to choose values of $\bar{f}$, $\bar{p}$, and $\bar{d}$ so as to balance the countervailing effect of the tiered contract's limited penalty feature and thus attain the first-best outcome. This flexibility is absent in linear contracts because the terms f and p are jointly set to provide sufficient incentives for optimal decision making (see the discussion in Section 4.2.1). This joint setting implicitly determines the sensitivity of the vendor's decision to potential financial distress. The resulting absence of flexibility limits the power of linear contracts to set the required low penalty rates for a vendor with a low distress threshold and/or a low reservation value (region B).

Our results for tiered contracts have important managerial implications. First, we show that such contracts can attain the client's first-best outcome irrespective of the vendor characteristics $\{\mathbf{v}, \mathbf{b}\}$. This result is consistent with the observed prevalence of tiered contracts in practice. Second, our results show how choosing an appropriate contract structure can help the client attain the first-

best outcome—by incentivizing the vendor to make the first-best decision while simultaneously satisfying her financial distress constraint—but without paying the vendor any surplus over her reservation value. Therefore, we show that using appropriate (tiered) contracts eliminates the possibility of losses due to agency issues in our context.

5. Conclusions and Future Research

In this paper we study the contracting issues that arise when a firm outsources the repair and restoration of its equipment to an external service provider. The design of an optimal contract for such outsourcing of repair services is complicated by two primary challenges. First, the contract must resolve the double-sided moral hazard problem that naturally occurs in such a setting. Second, a performance-based contract implicitly exposes the service provider to financial distress due to the inherent stochasticity in the system. Such implicit exposure creates disincentives for the vendor to make system-optimal decisions.

We find that if the vendor has a high reservation value and/or a high financial distress threshold, then the client can attain the first-best outcome with either a linear or a tiered contract structure. But when the vendor has a low reservation value and/or a low distress threshold, only the tiered contract structure empowers the client to attain the first-best outcome in all cases. The dominance of the tiered over the linear contract, irrespective of the vendor's characteristics, could well explain the preference for tiered contracts over linear contracts that is observed in practice.

This work contributes to the operations literature in a number of ways. First, the paper establishes a model that captures important challenges in the repair and restoration service outsourcing industry by combining elements of three different fields of research: the machine repairman model from the operations literature; the double-sided moral hazard framework from the economics literature; and the model of vendor's potential financial distress from the actuarial science literature. Under this model we find that the client's problem of designing an optimal contract can be reduced to a search space of the two classes of contracts that we study. Second, results based on our model show that tiered contracts can attain the client's first-best outcome for all parameter values. Third,

past studies have shown that, in a double-sided moral hazard environment, no contract can attain the first-best outcome. In some exceptional settings where certain special conditions apply, a few studies have shown that the first-best outcome can be attained. Even though none of those special conditions apply in our setting, we show how a firm can design contracts that achieve the first-best outcome. Fourth, our analysis also contributes to the relatively recent body of literature that studies contracting issues and their implications for the service provider's (or supplier's) financial health. Complementing the literature that focuses mostly on the role of up-front payments or subsidies in balancing the supplier's trade-off between system-optimal investments and her financial concerns, we show that contractual features can be leveraged to attain the first-best outcome without also paying any surplus to the vendor.

Our choice of performance-based contracts that penalize the vendor for each occurrence of the downtime are best suited for an environment where the frequency of transfer payments (rewards and/or penalties) between client and vendor are of the same order as the failure rate. In some settings, however, the frequency of transfer payments may be much lower than the failure rate e.g., when the contract penalizes the vendor based on average downtime over a prespecified (finite) period of time. Although we do not explicitly consider such settings, the insights of our paper extend directly to them. Tiered contracts will continue to attain the first-best outcome under all conditions, but there will be a range of parameter values for which linear contracts cannot attain the first-best outcome. This setting becomes even more interesting for future research when one considers the trade-off between lower frequency of transfer payments and the impact on firms of discounting future cash flows.

We focus on a one-to-one contractual relationship between the firm and the service provider. However, in practice we also observe many-to-one relationships between firms and a service provider. In these real-world settings, the service provider's exposure to potential financial distress is naturally lower due to diversification. Note that such a service provider may seek to exploit the "economies of scale" advantages, which could lead to opportunistic decision making by the service provider. It

would be interesting to study the optimal contract design for such many-to-one relationships. We believe this setting is an important topic for future research, but it will require a separate study.

Appendix

We invoke *renewal reward theory* to rewrite the profit rate functions to be used in all subsequent proofs:

$$\Pi = r - \frac{r+o}{1+\tau\mu} - c_u(\tau) - c_d(\mu), \tag{16}$$

$$\Pi_c = r - w - c_u(\tau) - \frac{r+o-E[T]\mu}{1+\tau\mu}, \tag{17}$$

$$\Pi_v = w - c_d(\mu) + \frac{E[T]\mu}{1+\tau\mu}. \tag{18}$$

In the proofs we use $\mathrm{BR}_c(\mu)$ and $\mathrm{BR}_v(\tau)$ to denote best response mappings for (respectively) client and vendor.

Proof of Proposition 1: The profit rate function Π is jointly concave in τ and μ since $r - \frac{r+o}{1+\tau\mu}$ is jointly concave in τ and μ (the Hessian matrix H is negative semi-definite) and since cost functions are convex. We have

$$H\left(r - \frac{r+o}{1+\tau\mu}\right) = \begin{bmatrix} -\frac{2(r+o)\mu^2}{(1+\tau\mu)^3} & \frac{(r+o)(1-\tau\mu)}{(1+\tau\mu)^3} \\ \frac{(r+o)(1-\tau\mu)}{(1+\tau\mu)^3} & -\frac{2(r+o)\tau^2}{(1+\tau\mu)^3} \end{bmatrix}.$$

Note,

$$\begin{aligned} M_1 &= -\frac{2(r+o)\mu^2}{(1+\tau\mu)^3} < 0 \\ M_2 &= \frac{2(r+o)\mu^2}{(1+\tau\mu)^3} \cdot \frac{2(r+o)\tau^2}{(1+\tau\mu)^3} - \frac{(r+o)(1-\tau\mu)}{(1+\tau\mu)^3} \cdot \frac{(r+o)(1-\tau\mu)}{(1+\tau\mu)^3} > 0, \end{aligned}$$

since $\underline{\tau}\underline{\mu} > 1$. Hence, the profit rate function, which is the sum of three concave functions $r - \frac{r+o}{1+\tau\mu}$, $-c_u(\tau)$, and $-c_d(\mu)$, is also concave. Since we have $\lim_{\tau\to\underline{\tau}} c_u'(\tau) = 0$, and $\lim_{\mu\to\underline{\mu}} c_d'(\underline{\mu}) = 0$, the concavity of profit rate function implies that the first-best decision $\{\tau^*, \mu^*\}$ is a unique interior point in $(\underline{\tau}, \overline{\tau}) \times (\underline{\mu}, \overline{\mu})$ and is characterized by the two first-order conditions $\frac{\partial\Pi}{\partial\tau} = 0$ and $\frac{\partial\Pi}{\partial\mu} = 0$, where

$$\frac{\partial\Pi}{\partial\tau} = \frac{(r+o)\mu}{(1+\tau\mu)^2} - c_u'(\tau), \tag{19}$$

$$\frac{\partial\Pi}{\partial\mu} = \frac{(r+o)\tau}{(1+\tau\mu)^2} - c_d'(\mu). \tag{20}$$

■

Proof of Proposition 2:

a. Applying the chain rule to (19) and (20) yields

$$\begin{aligned}\left((r+o)\mu^2+c_u''(\tau)(1+\tau\mu)^3\right)\frac{\partial\tau}{\partial o}+(r+o)(\tau\mu-1)\frac{\partial\mu}{\partial o} &= \mu(1+\tau\mu),\\ (r+o)(\tau\mu-1)\frac{\partial\tau}{\partial o}+\left((r+o)\tau^2+c_d''(\mu)(1+\tau\mu)^3\right)\frac{\partial\mu}{\partial o} &= \tau(1+\tau\mu).\end{aligned}$$

By Cramer's rule, we have

$$\begin{aligned}\frac{\partial\mu}{\partial o} &= \frac{\begin{bmatrix}(r+o)(\tau\mu-1) & \tau(1+\tau\mu)\\ (r+o)\mu^2+c_u''(\tau)(1+\tau\mu)^3 & \mu(1+\tau\mu)\end{bmatrix}}{\begin{bmatrix}(r+o)(\tau\mu-1) & (r+o)\tau^2+c_d''(\mu)(1+\tau\mu)^3\\ (r+o)\mu^2+c_u''(\tau)(1+\tau\mu)^3 & (r+o)(\tau\mu-1)\end{bmatrix}}\\ &= \frac{(1+\tau\mu)\{(r+o)\mu+\tau c_u''(\tau)(1+\tau\mu)^3\}}{(r+o)^2(2\tau\mu-1)+(1+\tau\mu)^3\{(r+o)c_u''(\tau)\tau^2+(r+o)\mu^2c_d''(\mu)+(1+\tau\mu)^3c_u''(\tau)c_d''(\mu)\}}\\ &> 0\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\tau}{\partial o} &= \frac{\begin{bmatrix}\tau(1+\tau\mu) & (r+o)\tau^2+c_d''(\mu)(1+\tau\mu)^3\\ \mu(1+\tau\mu) & (r+o)(\tau\mu-1)\end{bmatrix}}{\begin{bmatrix}(r+o)(\tau\mu-1) & (r+o)\tau^2+c_d''(\mu)(1+\tau\mu)^3\\ (r+o)\mu^2+c_u''(\tau)(1+\tau\mu)^3 & (r+o)(\tau\mu-1)\end{bmatrix}}\\ &= \frac{(1+\tau\mu)\{\tau(r+o)+\mu c_d''(\mu)(1+\tau\mu)^3\}}{(r+o)^2(2\tau\mu-1)+(1+\tau\mu)^3\{(r+o)c_u''(\tau)\tau^2+(r+o)\mu^2c_u''(\mu)+(1+\tau\mu)^3c_u''(\tau)c_d''(\mu)\}}\\ &> 0,\end{aligned}$$

since $\underline{\tau\mu}>1$, $c_u''(\tau)>0$, and $c_d''(\mu)>0$. Similarly, it is easy to see that $\frac{\partial\tau}{\partial r}>0$ and $\frac{\partial\mu}{\partial r}>0$.

b. We have $\frac{\partial^2\Pi}{\partial\tau\partial\mu}=\frac{(r+o)(1-\tau\mu)}{(1+\tau\mu)^3}<0$ since $\underline{\tau\mu}>1$. ■

Proof of Lemma 1: Note that the first-best decisions $\{\tau^*,\mu^*\}$ will correspond to a unique Nash-Equilibrium outcome of the simultaneous investment game *only if* $\mathrm{BR}_c(\mu^*)=\{\tau^*\}$ and $\mathrm{BR}_v(\tau)=\{\mu^*\}$. Furthermore, under the linear contract structure $\{w,f,p\}$ the client's profit rate function is concave

$$\frac{\partial^2\Pi_c(\tau,\mu^*)}{\partial\tau^2} = -2\frac{(r+o)\mu^{*2}}{(1+\tau\mu^*)^3}-c_u''(\tau)-2\frac{(f\mu^*-p)\mu^{*2}}{(1+\tau\mu^*)^3}<0,$$

since $f\mu^*-p=0$. This implies that the client's best-response action is determined by the first-order condition $\frac{\partial\Pi_c(\tau,\mu^*)}{\partial\tau}=0$, where by (17) we have,

$$\frac{\partial\Pi_c(\tau,\mu^*)}{\partial\tau} = \frac{(r+o)\mu^*}{(1+\tau\mu^*)^2}+\frac{(\mu^*f-p)\mu^*}{(1+\tau\mu^*)^2}-c_u'(\tau). \tag{21}$$

Now, (6) and (21) imply that $\mathrm{BR}_c(\mu^*)=\{\tau^*\}$ *only if* $\mu^*f-p=0$. Note that if the client set terms $\{f,p\}$ such that $\mu^*f-p=0$ then the vendor's profit rate function for the investment decisions $\{\tau^*,\mu^*\}$ will reduced to $\Pi_v(\tau^*,\mu^*)=w-c_d(\mu^*)$. Therefore, the client can attain the first-best outcome *only if* the client set $w=c_d(\mu^*)+v$ because for the first-best outcome the reservation value constraint (see inequality 11) has

to be binding (cf. definition of first-best outcome in Section 4). ■

PROOF OF PROPOSITION 3:

a. (i) The proof follows two steps. First, we show that for the proposed contractual terms $\{w, f_1, p_1\}$, which satisfy both the conditions enlisted in Lemma 1, the Nash-Equilibrium (NE) of the simultaneous investment game is $\{\tau^*, \mu^*\}$. Second, given $\{w, f_1, p_1\}$, and the NE $\{\tau^*, \mu^*\}$ the financial distress threshold (FDT) is satisfied. Together this will prove 5a(i).

Consider the modified client profit-maximization problem in which the FDT constraint is excluded. The NE for this modified problem is characterized by a system of first-order equations because for the proposed contractual terms $\{w, f_1, p_1\}$ the profit-rate function for both the client, and the vendor is concave in respective decisions τ and μ. In Lemma 1 we showed concavity of $\Pi_c(\tau, \mu^*)$. For $\Pi_v(\tau^*, \mu)$ note,

$$\frac{\partial^2 \Pi_v(\tau^*, \mu)}{\partial \mu^2} = -2\frac{(f_1 + p_1\tau^*)\tau^*}{(1+\tau^*\mu)^3} - c''(\mu) < 0,$$

since $f_1, p_1 > 0$. From (18) we obtain first-order condition for the vendor as,

$$\frac{\partial \Pi_v(\tau^*, \mu)}{\partial \mu} = \frac{f + p\tau^*}{(1+\tau^*\mu)^2} - c'(\mu) = 0. \tag{22}$$

Using (6), (7), (21), and (22) it is easy to verify that for $\{f_1, p_1\} = \left\{\frac{(r+o)\tau^*}{1+\tau^*\mu^*}, \frac{(r+o)\tau^*\mu^*}{1+\tau^*\mu^*}\right\}$, $\{\tau^*, \mu^*\}$ is a unique solution for the system of first-order equations $\left\{\frac{\partial \Pi_c(\tau,\mu^*)}{\partial \tau} = 0, \frac{\partial \Pi_v(\tau^*,\mu)}{\partial \mu} = 0\right\}$. Now it is remaining to show that given $\{w, f_1, p_1\}$ and the NE outcome $\{\tau^*, \mu^*\}$ the FDT constraint (see inequality 12) is satisfied. For remaining part of this proof we set

$$b_1 = e^{-\frac{2w_o \mathbf{v}(1+\tau^*\mu^*)^3}{(r+o)^2\tau^{*2}\mu^*}}, b_2 = e^{-\frac{2w_o \mathbf{v}(1+\tau^*\mu^*)^3\mu^*}{((r+o)\tau^*\mu^* - \mathbf{v}(1+\tau^*\mu^*)(1+2\tau^*\mu^*))^2}}, \text{ and } v_2 = \frac{(r+o)\tau^*\mu^*}{(1+\tau^*\mu^*)(1+2\tau^*\mu^*)}.$$

Note that under a linear contract $\{w, f, p\}$, the earning per service call for the vendor is $T = f - p.D$. If the vendor makes decision μ, then the vendor's expected earnings per service call and variance in earnings per call are respectively

$$E[T] = f - \frac{p}{\mu}, \quad V[T] = \frac{p^2}{\mu^2}.$$

Moreover, if the client makes decision τ, then the potential financial distress is $e^{-2w_o\frac{(w-c_d(\mu))(1+\tau\mu)+E[T]\mu}{V[T]\mu}}$ or equivalently $e^{-2w_o\frac{\Pi_v(\tau,\mu)(1+\tau\mu)}{V[T]\mu}}$ (from (18) we have $\Pi_v = w - c_d(\mu) + \frac{E[T]\mu}{1+\tau\mu}$). Thus, the FDT constraint can be written as

$$e^{-2w_o\frac{\Pi_v(1+\tau\mu)}{V[T]\mu}} \leq \mathbf{b},$$

or equivalently,

$$\Pi_v(1+\tau\mu)\mu - \frac{p^2}{2w_o}\log\frac{1}{\mathbf{b}} \geq 0. \tag{23}$$

Denote by $h(\tau,\mu\, p)$ the left hand side in inequality (23). Thus,

$$h(\tau,\mu,p) = \Pi_v(\tau,\mu)(1+\tau\mu)\mu - \frac{p^2}{2w_o}\log\frac{1}{\mathbf{b}}. \tag{24}$$

Therefore, in terms of $h(\tau,\mu,p)$ the FDT constraint is $h(\tau,\mu,p)\geq 0$. Note that $h(\tau,\mu,p)$ is increasing in $\mathbf{b}$. This implies that for all $\mathbf{b}\in[b_1,1]$ we have

$$h(\tau^*,\mu^*,p_1)|_{\mathbf{b}} \geq h(\tau^*,\mu^*,p_1)|_{b_1} = \Pi_v(\tau^*,\mu^*)(1+\tau\mu^*)\mu^* - \frac{p_1^2}{2w_o}\log\frac{1}{b_1}.$$

By definition of p_1 and b_1 we have

$$\begin{aligned} h(\tau^*,\mu^*,p_1)|_{b_1} &= \Pi_v(\tau^*,\mu^*)(1+\tau\mu^*)\mu^* - \frac{1}{2w_o}\left(\frac{(r+o)\tau^*\mu^*}{1+\tau^*\mu^*}\right)^2 \frac{2w_o\mathbf{v}(1+\tau^*\mu^*)^3}{(r+o)^2\tau^{*2}\mu^*}, \\ &= 0, \end{aligned}$$

since $\Pi_v(\tau^*,\mu^*)=\mathbf{v}$. ■

a. (ii) We show that for a $b\in[0,b_1]$ the proposed contract $\{w,f_2,p_2\}$ achieves first best. We will prove the following: (1) $h(\tau^*,\mu^*,p_2)\geq 0$, (2) $BR_c(\mu^*)=\{\tau^*\}$ and (3) $BR_v(\tau^*)=\{\mu^*\}$.

First we show that the proposed contract satisfies $h(\tau^*,\mu^*,p_2)=0$. To see this note that $\Pi_V(\tau^*,\mu^*)=\mathbf{v}$, therefore

$$\begin{aligned} h(\tau^*,\mu^*,p_2) &= \mathbf{v}(1+\tau^*\mu^*)\mu^* - p_2^2\frac{1}{2w_o}\log\frac{1}{\mathbf{b}}, \\ &= \mathbf{v}(1+\tau^*\mu^*)\mu^* - \frac{2w_o\mathbf{v}(1+\tau^*\mu^*)\mu^*}{\log\frac{1}{\mathbf{b}}}\frac{1}{2w_o}\log\frac{1}{\mathbf{b}}, \\ &= 0, \end{aligned} \tag{25}$$

by definition of p_2. Additionally, the proposed contract $\{w,f_2,p_2\}$ satisfies the two required conditions mentioned in Lemma 1. This implies $BR_c(\mu^*)=\{\tau^*\}$.

In order to prove that $BR_v(\tau^*)=\{\mu^*\}$, we will first show that the function $\Pi_v(\tau^*,\mu)$ is decreasing on $[\mu^*,\overline{\mu}]$ (Claim 1), and then show that the FDT constraint (i.e., $h(\tau^*,\mu^*,p_2)\geq 0$) is violated on $[\underline{\mu},\mu^*)$ (Claim 2). Therefore, $BR_v(\tau^*)=\{\mu^*\}$.

Claim 1. $\partial_u\Pi_v(\tau^*,\mu)<0$, $\forall\mu\in[\mu^*,\overline{\mu}]$. *Hence* $\Pi_v(\tau^*,\mu^*)>\Pi_v(\tau^*,\mu)$, $\forall\,\mu\in[\mu^*,\overline{\mu}]$

Proof: We show that $\frac{\partial \Pi_v}{\partial \mu}(\tau^*, \mu^*) < 0$. This coupled with the fact that $\partial_{\mu\mu}\Pi_v(\tau^*, \mu) < 0$ implies $\partial_u \Pi_v(\tau^*, \mu) < 0$, $\forall \mu \in [\mu^*, \overline{\mu}]$.

$$\frac{\partial \Pi_v}{\partial \mu}(\tau^*, \mu^*) = \frac{f_2 + p_2\tau^*}{(1+\tau^*\mu^*)^2} - c'(\mu^*).$$

Using (7) and the fact that $f_2 = p_2/\mu^*$ we get

$$\begin{aligned} \frac{\partial \Pi_v}{\partial \mu}(\tau^*, \mu^*) &= \frac{p_2}{\mu^*(1+\tau^*\mu^*)} - \frac{(r+o)\tau^*}{(1+\tau^*\mu^*)^2} \\ &= \left(\frac{2w_o\mathbf{v}}{\mu^*(1+\tau^*\mu^*)}\right)^{1/2} \left(\frac{1}{\sqrt{\log\frac{1}{\mathbf{b}}}} - \frac{1}{\sqrt{\log\frac{1}{b_1}}}\right), \end{aligned} \tag{26}$$

by definition of p_2 and b_1. But $\mathbf{b} < b_1$, hence $\frac{\partial \Pi_v}{\partial \mu}(\tau^*, \mu^*) < 0$. ■

Claim 2. $h(\tau^*, \mu, p_2) < 0, \forall \mu \in [\underline{\mu}, \mu^*)$.

Proof: We show that $\partial_\mu h(\tau^*, \mu, p_2) > 0, \forall \mu \in [\underline{\mu}, \mu^*)$. This together with $h(\tau^*, \mu^*, p_2) = 0$ (see (25) implies $h(\tau^*, \mu, p_2) < 0 \forall \mu \in [\underline{\mu}, \mu^*)$.

To show $\partial_\mu h(\tau^*, \mu, p_2) > \mathbf{0}$, $\forall \mu \in [\underline{\mu}, \mu^*)$, we split $[\underline{\mu}, \mu^*)$ into two regions $R_1 = [\underline{\mu}, \mu')$ and $R_2 = [\mu', \mu^*)$, where μ' is the unique solution of $\partial_\mu \Pi_v(\tau^*, \mu) = 0$. Such a μ' exists since $\partial_\mu \Pi_v(\tau^*, \mu)$ is continuous, $\frac{\partial \Pi_v}{\partial \mu}(\tau^*, \mu^*) < 0$ as shown in Claim 1, and $\frac{\partial \Pi_v}{\partial \mu}(\tau^*, \underline{\mu}) > 0$ since $\lim_{\mu \to \underline{\mu}} c_d'(\underline{\mu}) = 0$. Note that

$$\partial_\mu h(\tau^*, \mu, p_2) = \partial_\mu \Pi_v(\tau^*, \mu)(1+\tau^*\mu)\mu + 2\Pi_v(\tau^*, \mu)\tau^*\mu + \Pi_v(\tau^*, \mu).$$

From the concavity of $\Pi_v(\tau^*, \mu)$, combined with the fact that $\frac{\partial \Pi_v}{\partial \mu}(\tau^*, \underline{\mu}) > 0$ and $\frac{\partial \Pi_v}{\partial \mu}(\tau^*, \mu') = 0$ we get that $\partial_\mu \Pi_v(\tau^*, \mu) > 0 \ \forall \mu \in R_1$. Hence, $\partial_\mu h(\tau^*, \mu, p_2) > 0 \forall \mu \in R_1$.

We now write $\partial_\mu h(\tau^*, \mu, p_2) = \mu g(\mu)$, where

$$g(\mu) = \partial_\mu \Pi_v(\tau^*, \mu)(1+\tau^*\mu) + \Pi_v(\tau^*, \mu)\left(2\tau^* + \frac{1}{\mu}\right). \tag{27}$$

The function $g(\mu)$ is decreasing in R_2. To see this, note that

$$g'(\mu) = \partial_{\mu\mu}\Pi_v(\tau^*, \mu)(1+\tau^*\mu) + \partial_\mu \Pi_v(\tau^*, \mu)\left(2\tau^* + \frac{1}{\mu}\right) - \frac{\Pi_v(\tau^*, \mu)}{\mu^2} < 0,$$

because $\partial_{\mu\mu}\Pi_v(\tau^*, \mu) < 0$ for all $\mu \in [\underline{\mu}, \overline{\mu})$ and $\partial_\mu \Pi_v(\tau^*, \mu) < 0$ for all $\mu \in R_2$. Moreover, $g(\mu^*) > 0$. To see this, note that from (26) and $\Pi_v(\tau^*, \mu^*) = \mathbf{v}$ we have

$$g(\mu^*) = \left(\frac{p_2}{\mu^*(1+\tau^*\mu^*)} - \frac{(r+o)\tau^*}{(1+\tau^*\mu^*)^2}\right)(1+\tau^*\mu^*) + 2\mathbf{v}\tau^* + \frac{\mathbf{v}}{\mu^*}$$

$$= \left(\frac{2w_o\mathbf{v}(1+\tau^*\mu^*)\mu^*}{\log(\frac{1}{\mathbf{b}})}\right)^{1/2}\frac{1}{\mu^*} + \left(\frac{\mathbf{v}(2\tau^*\mu^*+1)(1+\tau^*\mu^*)-(r+o)\tau^*\mu^*}{(1+\tau^*\mu^*)\mu^*}\right) \tag{28}$$
$$> 0$$

since $\mathbf{v} \in \left[\frac{(r+o)\tau^*\mu^*}{(1+\tau^*\mu^*)(1+2\tau^*\mu^*)}, \Pi^*\right]$. Thus, $g(\mu) > 0$ for $\mu \in R_2$. This implies $\partial_\mu h(\tau^*, \mu, p_2) > 0$ for all $\mu \in R_2$. We previously showed $\partial_\mu h(\tau^*, \mu, p_2) > 0 \, \forall \mu \in R_1$, hence $\partial_\mu h(\tau^*, \mu, p_2) > \mathbf{0} \, \forall \mu \in [\underline{\mu}, \mu^*)$. ■

b. (i) Following the steps of proof for a. (i) it is easily verifiable that for the proposed contract $\{w, f_1, p_1\}$ the NE of the simultaneous investment game is $\{\tau^*, \mu^*\}$, the reservation value constraint is binding, and the FDT constraint is redundant. Thus, the proposed contract attains the first-best outcome. ■

b. (ii) Note that contingent on a vendor's exogenous characteristics $\{\mathbf{v}, \mathbf{b}\}$, the proposed contract terms $\{w, f_2, p_2\}$ are same as the one proposed in a(ii). For a given $\{\mathbf{v}, \mathbf{b}\}$, following the steps of proof for a. (ii), it is easy to verify that (1) $BR_c\mu^* = \{\tau^*\}$, (2) $h(\tau^*, \mu^*, p_2) = 0$, and (3) $\partial_u \Pi_v(\tau^*, \mu) < 0$, $\forall \mu \in [\mu^*, \overline{\mu}]$ holds true. Now, to prove the remaining step that $h(\tau^*, \mu, p_2) < 0 \, \forall \, \mu \in [\underline{\mu}, \mu^*)$ (see Claim 2 in a.(ii)) note that the necessary and sufficient condition for $\{w, f_2, p_2\}$ to attain the first-best is $g(\mu^*) \geq 0$ because otherwise Claim 2 would not hold true. We show that for a $\mathbf{v} \in (0, v_2)$, $g(\mu^*)$ is increasing in $\mathbf{b}$ and that $g(\mu^*)|_{b_2} = 0$. Taken together it implies $g(\mu^*) \geq 0 \, \forall \, \mathbf{b} \in [b_2, b_3)$ i.e., the necessary and sufficient condition is satisfied. From (28) we have

$$g(\mu^*, \mathbf{b}) = \left(\frac{2w_o\mathbf{v}(1+\tau^*\mu^*)\mu^*}{\log(\frac{1}{\mathbf{b}})}\right)^{1/2}\frac{1}{\mu^*} + \left(\frac{\mathbf{v}(2\tau^*\mu^*+1)(1+\tau^*\mu^*)-(r+o)\tau^*\mu^*}{(1+\tau^*\mu^*)\mu^*}\right).$$

This implies $\partial_{\mathbf{b}} g(\mu^*, \mathbf{b}) = \left(\frac{2\mathbf{v}(1+\tau^*\mu^*)}{\log(\frac{1}{\mathbf{b}})}\right)^{1/2} \frac{1}{\mu^*} \frac{1}{2\mathbf{b}\log(\frac{1}{\mathbf{b}})} > 0$, and

$$g(\mu^*, b_2) = \left(\frac{2\mathbf{v}(1+\tau^*\mu^*)\mu^*}{\log(\frac{1}{b_2})}\right)^{1/2}\frac{1}{\mu^*} + \left(\frac{\mathbf{v}(2\tau^*\mu^*+1)(1+\tau^*\mu^*)-(r+o)\tau^*\mu^*}{(1+\tau^*\mu^*)\mu^*}\right).$$

By definition of b_2 we get

$$\begin{aligned} g(\mu^*, b_2) &= \left(\frac{2w_o\mathbf{v}(1+\tau^*\mu^*)\mu^*\left((r+o)\tau^*\mu^* - \mathbf{v}(1+\tau^*\mu^*)(1+2\tau^*\mu^*)\right)^2}{2w_o\mathbf{v}(1+\tau^*\mu^*)^3\mu^*}\right)^{1/2}\frac{1}{\mu^*} \\ &\quad + \left(\frac{\mathbf{v}(2\tau^*\mu^*+1)(1+\tau^*\mu^*)-(r+o)\tau^*\mu^*}{(1+\tau^*\mu^*)\mu^*}\right) \\ &= \frac{(r+o)\tau^*\mu^* - \mathbf{v}(1+\tau^*\mu^*)(1+2\tau^*\mu^*)}{(1+\tau^*\mu^*)}\frac{1}{\mu^*} + \left(\frac{\mathbf{v}(2\tau^*\mu^*+1)(1+\tau^*\mu^*)-(r+o)\tau^*\mu^*}{(1+\tau^*\mu^*)\mu^*}\right) \\ &= 0. \end{aligned}$$

■

b. (iii) We show that for $b \in (0, b_2)$, for every contract $\{w, f, p\}$ of the form described in Lemma 1, we have either (1) $BR_v(\tau^*) \neq \mu^*$ or (2) $h(\tau^*, \mu^*, p) < 0$, hence the linear contract cannot attain first-best outcome.

For some $b \in (0, b_2)$ and $p_2 = \frac{2w_o \mathbf{v}(1+\tau^*\mu^*)\mu^*}{\log(\frac{1}{\mathbf{b}})}$ one can easily verify that $h(\tau^*, \mu^*, p_2) = 0$. Further, we have shown in b (ii). that $\partial_{\mathbf{b}} g(\mu^*, \mathbf{b}) > 0$ and $g(\mu^*, b_2) = 0$. This implies for $b \in (0, b_2)$ we have $\frac{\partial h(\tau^*, \mu, p_2)}{\partial \mu} < 0$. From continuity of $h(\tau^*, \mu, p_2)$ we know that (1) $\exists\, \delta_h > 0$ such that $h(\tau^*, \mu^* - \epsilon, p_2) > 0$ for all $\epsilon \in (0, \delta_h)$.

Also from Claim 1 in a. (ii) we know $\frac{\partial \Pi_v}{\partial \mu}(\tau^*, \mu^*) < 0$. From the continuity of $\Pi_v(\tau^*, \mu)$ we know that (2) $\exists\, \delta_\mu > 0$ such that $\Pi_v(\tau^*, \mu^* - \epsilon) > \Pi_v(\tau^*, \mu^*)$ for all $\epsilon \in (0, \delta_\mu)$.

Combining (1) and (2) we get $h(\tau^*, \mu^* - \epsilon, p_2) > 0$ and $\Pi_v(\tau^*, \mu^* - \epsilon) > \Pi_v(\tau^*, \mu^*) \,\forall\, \epsilon \in (0, \min(\delta_h, \delta_\mu)]$. This implies $BR_v(\tau^*) \neq \{\mu^*\}$. Hence, no contract of the form $\{w, f, p_2\}$ can attain the first-best outcome.

For $p > p_2$, we have $h(\tau^*, \mu^*, p) < 0$, Hence no contract of the form $\{w, f, p\}$ with $p > p_2$ can attain the first-best outcome.

Lastly, for $p < p_2$, we have $h(\tau^*, \mu^*, p) > 0$ and $\frac{\partial \Pi_v}{\partial \mu}(\tau^*, \mu^*) < 0$. Again by continuity arguments, we know that $\exists \delta'_h > 0, \delta'_\mu > 0$ such that $h(\tau^*, \mu^* - \epsilon, p_2) > 0$ and $\Pi_v(\tau^*, \mu^* - \epsilon) > \Pi_v(\tau^*, \mu^*) \,\forall\, \epsilon \in (0, \min(\delta'_h, \delta'_\mu)]$. This implies $BR_v(\tau^*) \neq \{\mu^*\}$. Hence, no contract of the form $\{w, f, p\}$ with $p < p_2$ can attain the first-best outcome. ■

Proof of Corollary 1:

Let $\Pi_v^B(\tau^*, \mu^*)$ denote the profit rate paid by the client to a Region B vendor, with exogenous parameters $\{\mathbf{v}_B, \mathbf{b}_B\}$, for inducing first-best investment decisions from her.

With reference to Figure 3 we denote the intersection point of a horizontal line $y = \mathbf{b}_B$ and the region split curve by $\{v^\dagger, \mathbf{b}_B\}$. Note that the vendor $\{\mathbf{v}_B, \mathbf{b}_B\}$ would lie on left hand side of this intersection point. It is easy to see that the client can attain the first-best investments from such a vendor if the client offers her contract terms $\{w, f, p\}$ as suggested by Proposition 3 for a vendor characterized by $\{v^\dagger, \mathbf{b}_B\}$. Intuitively, the client mimics the vendor $\{\mathbf{v}_B, \mathbf{b}_B\}$ by a vendor characterized by $\{v^\dagger, \mathbf{b}_B\}$. This implies that the client would pay $\Pi_v^B(\tau^*, \mu^*) = v^\dagger (> \mathbf{v}_B)$. Following steps of proof for Proposition 3 a (ii). one can verify that the region split curve is defined by $g(\mu^*) = 0$. Therefore, using (28) we compute $v^\dagger$ by

$$\left(\frac{2w_o v^\dagger(1+\tau^*\mu^*)\mu^*}{\log(\frac{1}{\mathbf{b}_B})}\right)^{1/2} \frac{1}{\mu^*} + \left(\frac{v^\dagger(2\tau^*\mu^*+1)(1+\tau^*\mu^*) - (r+o)\tau^*\mu^*}{(1+\tau^*\mu^*)\mu^*}\right) = 0$$

$$\left(\frac{2w_o v^\dagger(1+\tau^*\mu^*)\mu^*}{\log(\frac{1}{\mathbf{b}_B})}\right)^{1/2} + \left(v^\dagger(2\tau^*\mu^*+1) - \frac{(r+o)\tau^*\mu^*}{(1+\tau^*\mu^*)}\right) = 0$$

Solving the above quadratic equation in $v^\dagger$ for the positive root,

$$v^{\dagger} = \left(\frac{-(1+\tau^*\mu^*)\sqrt{2w_o\mu^*} + \sqrt{2w_o(1+\tau^*\mu^*)^2\mu^* + 4\log(\frac{1}{\mathbf{b}_B})(2\tau^*\mu^*+1)(r+o)\tau^*\mu^*}}{2(2\tau^*\mu^*+1)\sqrt{(1+\tau^*\mu^*)\log(\frac{1}{\mathbf{b}_B})}} \right)^2 .$$

■

Proof of Proposition 4:

a. We show that for $\bar{d} = \frac{2}{\underline{\mu}}$ the specified tiered contract attains first-best outcome. For this we must show that: (1) $BR_c(\mu^*) = \{\tau^*\}$, (2) $BR_v(\tau^*) = \{\mu^*\}$, and (3) $\Pi_v(\tau^*, \mu^*) = \mathbf{v}$. Note that (1) and (2) implies that Nash Equilibrium of the simultaneous investment game is $\{\tau^*, \mu^*\}$ and (3) implies that when the optimal investment decisions are made the vendor earns no more than her reservation profit rate $\mathbf{v}$.

Note that, under the tiered contract structure, $E[T] = \bar{f} - \bar{p}e^{-\mu\bar{d}}$ and $V[T] = \bar{p}^2(e^{-\mu\bar{d}} - e^{-2\mu\bar{d}})$. We can rewrite the profit rates and function $h(\tau, \mu, \bar{p})$ as:

$$\Pi_c(\tau, \mu) = r - w - c_u(\tau) - \frac{r + o + (\bar{f} - \bar{p}e^{-\mu\bar{d}})\mu}{(1+\tau\mu)} \tag{29}$$

$$\Pi_v(\tau, \mu) = w - c_d(\mu) + \frac{(\bar{f} - \bar{p}e^{-\mu\bar{d}})\mu}{1+\tau\mu}; \tag{30}$$

$$h(\tau, \mu, \bar{p}) = \Pi_v(\tau, \mu)(1+\tau\mu) - \bar{p}^2(e^{-\mu\bar{d}} - e^{-2\mu\bar{d}})\mu\frac{1}{2w_o}\log\frac{1}{\mathbf{b}}. \tag{31}$$

Let $b_3 = e^{-\frac{8w_o\mathbf{v}\mu^*}{c_d'(\mu^*)^2(1+\tau^*\mu^*)(e^{2\mu^*/\underline{\mu}}-1)\underline{\mu}^2}}$. Observe that the proposed contract parameters $\bar{f}_3$ and $\bar{p}_3$ satisfy the following equality:

$$\bar{f}_3 = \bar{p}_3 e^{-\mu^*\bar{d}}. \tag{32}$$

Under this structural relationship, the profit rate $\Pi_c(\tau, \mu^*)$ is concave in τ:

$$\frac{\partial^2\Pi_c(\tau, \mu^*)}{\partial\tau^2} = -2\frac{(r+o)\mu^{*2}}{(1+\tau\mu^*)^3} - c_u''(\tau) < 0 \tag{33}$$

Also note that

$$\frac{\partial\Pi_c(\tau, \mu^*)}{\partial\tau} = \frac{(r+o)\mu^* + (\bar{f}_3 - \bar{p}_3 e^{-\mu^*\bar{d}})\mu^{*2}}{(1+\tau\mu^*)^2} - c_u'(\tau). \tag{34}$$

and by (6) we have $\frac{\partial\Pi_c}{\partial\tau}(\tau^*, \mu^*) = 0$. Thus, $BR_c(\mu^*) = \{\tau^*\}$.

We now show $BR_v(\tau^*) = \{\mu^*\}$in two steps. First, we show that under the modified unconstrained problem in which FDT constraint is excluded the vendor's best-response action under $\{w, \bar{f}_3, \bar{p}_3, \bar{d}\}$ is determined by the first-order condition. Second, we show that given $\{w, \bar{f}_3, \bar{p}_3, \bar{d}\}$ and the investment decisions $\{\tau^*, \mu^*\}$ the FDT constraint is satisfied.

Note that the profit rate function $\Pi_v(\tau^*,\mu)$ is concave given contract parameters $\{w, f_3, p_3, \bar{d}\}$

$$\frac{\partial^2 \Pi_v}{\partial \mu^2}(\tau^*,\mu^*) = \frac{-2\bar{f}_3\tau^* - \bar{p}_3 e^{-\mu\bar{d}}(\tau^*(\tau^*\mu(\mu\bar{d})^2 - 2) + \bar{d}(\mu\bar{d} - 2) + 2\tau^*\mu\bar{d}(\mu\bar{d} - 1))}{(1+\tau^*\mu)^3} - c_d''(\mu) < 0$$

since $\bar{f}_3 > 0$, $\bar{p}_3 > 0$, $\underline{\tau}\underline{\mu} > 1$, and $\bar{d} = \frac{2}{\underline{\mu}}$.

This implies that best-response action is determined by the first-order condition $\frac{\partial \Pi_v(\tau^*,\mu)}{\partial \mu} = 0$. Note that μ^* is a solution for the first-order condition i.e., $\frac{\partial \Pi_v}{\partial \mu}(\tau^*,\mu^*) = 0$:

$$\frac{\partial \Pi_v(\tau^*,\mu)}{\partial \mu} = \frac{\bar{f}_3 - \bar{p}_3 e^{-\mu\bar{d}} + \bar{p}_3 e^{-\mu\bar{d}}\bar{d}\mu(1+\tau^*\mu)}{(1+\tau^*\mu)^2} - c_d'(\mu), \tag{35}$$

By (32) and definition of $\bar{p}_3$ we get

$$\begin{aligned}
\frac{\partial \Pi_v}{\partial \mu}(\tau^*,\mu^*) &= \frac{\bar{f}_3 - \bar{p}_3 e^{-\mu^*\bar{d}} + \bar{p}_3 e^{-\mu^*\bar{d}}\bar{d}\mu^*(1+\tau^*\mu^*)}{(1+\tau^*\mu^*)^2} - c_d'(\mu^*), \\
&= \frac{\bar{p}_3 e^{-\mu^*\bar{d}}\bar{d}\mu^*}{(1+\tau^*\mu^*)} - c_d'(\mu^*), \\
&= 0.
\end{aligned} \tag{36}$$

Now to show that $h(\tau^*,\mu^*,\bar{p}_3) \geq 0 \, \forall \, \mathbf{b} \in [b_3, 1]$, observe that $h(\tau^*,\mu^*,\bar{p}_3)$ is increasing in $\mathbf{b}$. This implies that

$$\begin{aligned}
h(\tau^*,\mu^*,\bar{p}_3)|_{\mathbf{b}} &\geq h(\tau^*,\mu^*,\bar{p}_3)|_{b_3} \\
&\geq \Pi_v(\tau^*,\mu^*)(1+\tau^*\mu^*) - \bar{p}_3^2(e^{-\mu^*\bar{d}} - e^{-2\mu^*\bar{d}})\mu^* \frac{1}{2w_o}\log\frac{1}{b_3} \\
&\geq 0,
\end{aligned} \tag{37}$$

since $\Pi_v(\tau^*,\mu^*) = \mathbf{v}$ and by definition of b_3 and $\bar{p}_3$. (36) and (37) together implies that $BR_v(\tau^*) = \{\mu^*\}$.

Finally, under the proposed contract parameters the vendor's profit rate with investment decisions $\{\tau^*,\mu^*\}$ is

$$\begin{aligned}
\Pi_v(\tau^*,\mu^*) &= w - c_d(\mu^*) + \frac{(\bar{f}_3 - \bar{p}_3 e^{-\mu^*\bar{d}})\mu^*}{1+\tau^*\mu^*} \\
&= c_d(\mu^*) + \mathbf{v} - c_d(\mu^*) \\
&= \mathbf{v},
\end{aligned}$$

by (32) and definition of w. ■

b. We need to show four parts: (i) there exists $\bar{d} \geq \frac{2}{\underline{\mu}}$ satisfying (15); (ii) for such a $\bar{d}$, the proposed contract $\{w, \bar{f}_4, \bar{p}_4, \bar{d}\}$ leads to $BR_c(\tau^*) = \{\mu^*\}$, (iii) $BR_v(\mu^*) = \{\tau^*\}$and (iv) $\Pi_v(\tau^*,\mu^*) = \mathbf{v}$.

We first show (i): there exists $\bar{d} \geq \frac{2}{\underline{\mu}}$ satisfying (15), where inequality (15) is:

$$\bar{d}\left(1 - \frac{1}{(e^{\mu^*\bar{d}} - 1)}\right) \geq \frac{\mu^*}{\mathbf{v}(1+\tau^*\mu^*)}(c_d(\mu^*) + \mathbf{v} - c_d(\mu))\frac{1}{\mu^2} + c_d'(\mu)\left(\frac{1}{\mu} + \tau^*\right), \forall \mu \in [\underline{\mu}, \mu^*]$$

Proof: Define

$$g_1(\bar{d}) = \bar{d}\left(1 - \frac{1}{e^{\mu^*\bar{d}} - 1}\right),$$
$$g_2(\mu) = \frac{\mu^*}{\mathbf{v}(1+\tau^*\mu^*)}\left(\frac{c_d(\mu^*) + \mathbf{v} - c_d(\mu)}{\mu^2} + c_d'(\mu)\left(\frac{1}{\mu} + \tau^*\right)\right).$$

We show that there exists $\bar{d} \geq \frac{2}{\underline{\mu}}$ such that

$$g_1(\bar{d}) \geq g_2(\mu) \quad \forall \mu \in [\underline{\mu}, \mu^*].$$

Functions $g_1$and g_2 have the following properties:

Property (P1): g_1 is monotonically increasing in $\bar{d}$ and $\lim_{\bar{d}\to\infty} g_1(\bar{d}) = \infty$. (Trivial).

Property (P2): g_2 is bounded on $[\underline{\mu}, \mu^*]$. (Trivial)

Since $g_2(\mu)$ is bounded on $[\underline{\mu}, \mu^*]$ and $g_1(\bar{d})$ is monotonically increasing in $\bar{d}$ with $\lim_{\bar{d}\to\infty} g_1(\bar{d}) = \infty$, there exists a $\bar{d} \geq \frac{2}{\underline{\mu}}$ such that $g_1(\bar{d}) \geq g_2(\mu), \quad \forall \mu \in [\underline{\mu}, \mu^*]$. ■

We now prove (ii) that $BR_c(\mu^*) = \{\tau^*\}$. Note that the contract terms $\{\bar{f}_4, \bar{p}_4, \bar{d}\}$ again satisfy the structural relationship described in (32) i.e.,

$$\bar{f}_4 = \bar{p}_4 e^{-\mu^*\bar{d}}. \tag{38}$$

Using this structural relationship and following the steps of proof for part (a) we can easily verify that the client's profit rate function $\Pi_c(\tau, \mu^*)$ is concave and τ^* satisfy the first-order condition i.e., $\frac{\partial \Pi_c}{\partial \tau}(\tau^*, \mu^*) = 0$. This proves (ii) that $BR_c(\mu^*) = \{\tau^*\}$. Similarly, it is easily verifiable that with this structural property and $w = c_d(\mu^*) + \mathbf{v}$ the vendor's profit rate with investment decisions (τ^*, μ^*) is $\mathbf{v}$ i.e., $\Pi_v(\tau^*, \mu^*) = \mathbf{v}$. Hence, (iv) holds true.

To prove the remaining part (iii) that $BR_v(\tau^*) = \{\mu^*\}$ we follow the same steps as in Proposition 5 (a.ii and b.ii). We first show that the FDT constraint is satisfied under the proposed contract when the investment decisions are $\{\tau^*, \mu^*\}$, follow it by showing that the vendor's profit rate function $\Pi_v(\tau^*, \mu)$ is decreasing on $[\mu^*, \overline{\mu}]$ (Claim 3), and finally showing that the FDT constraint is violated on $[\underline{\mu}, \mu^*)$ (Claim 4). Therefore, $BR_v(\tau^*) = \{\mu^*\}$. Using the definition of $\bar{p}_4$, $\Pi_v(\tau^*, \mu^*) = \mathbf{v}$, and (31) we compute

$$\begin{aligned} h(\tau^*, \mu^*, \bar{p}_4) &= \Pi_v(\tau^*, \mu^*)(1 + \tau^*\mu^*) - \frac{2w_o\mathbf{v}(1+\tau^*\mu^*)}{\log(1/\mathbf{b})\mu^*(e^{-\mu^*\bar{d}} - e^{-2\mu^*\bar{d}})}(e^{-\mu^*\bar{d}} - e^{-2\mu^*\bar{d}})\mu^* \frac{1}{2w_o}\log\frac{1}{\mathbf{b}}, \\ &= 0, \end{aligned} \tag{39}$$

since $\Pi_v(\tau^*, \mu^*) = \mathbf{v}$.

Claim 3. $\partial_u \Pi_v(\tau^*,\mu) < 0, \forall \mu \in [\mu^*, \overline{\mu}]$. *Hence* $\Pi_v(\tau^*,\mu^*) > \Pi_v(\tau^*,\mu), \forall \mu \in [\mu^*, \overline{\mu}]$.

Proof: We show that under $\{w, \bar{f}_4, \bar{p}_4, \bar{d}\}$ the vendor's profit rate function is concave i.e., $\partial_{\mu\mu}\Pi_v(\tau^*,\mu) < 0$. Further we will show that $\frac{\partial \Pi_v}{\partial \mu}(\tau^*,\mu^*) < 0$. Therefore, $\partial_u \Pi_v(\tau^*,\mu) < 0, \forall \mu \in [\mu^*, \overline{\mu}]$. Note,

$$\frac{\partial^2 \Pi_v}{\partial \mu^2}(\tau^*,\mu^*) = \frac{-2\bar{f}_4\tau^* - \bar{p}_4 e^{-\mu\bar{d}}(\tau^*(\tau^*\mu(\mu\bar{d})^2 - 2) + \bar{d}(\mu\bar{d} - 2) + 2\tau^*\mu\bar{d}(\mu\bar{d} - 1))}{(1+\tau^*\mu)^3} - c_d''(\mu) < 0$$

since $\bar{f}_4 > 0$, $\bar{p}_4 > 0$, $\underline{\tau}\underline{\mu} > 1$, and $\bar{d} \geq \frac{2}{\underline{\mu}}$. Further,

$$\frac{\partial \Pi_v(\tau^*,\mu)}{\partial \mu} = \frac{\bar{f}_4 - \bar{p}_4 e^{-\mu\bar{d}} + \bar{p}_4 e^{-\mu\bar{d}}\bar{d}\mu(1+\tau^*\mu)}{(1+\tau^*\mu)^2} - c_d'(\mu) \tag{40}$$

$$\frac{\partial \Pi_v}{\partial \mu}(\tau^*,\mu^*) = \frac{\bar{f}_4 - \bar{p}_4 e^{-\mu^*\bar{d}} + \bar{p}_4 e^{-\mu^*\bar{d}}\bar{d}\mu^*(1+\tau^*\mu^*)}{(1+\tau^*\mu^*)^2} - c_d'(\mu^*). \tag{41}$$

By (38) and definition of $\bar{p}_4$ we get

$$\begin{aligned}
\frac{\partial \Pi_v}{\partial \mu}(\tau^*,\mu^*) &= \frac{\bar{p}_4 e^{-\mu^*\bar{d}}\bar{d}\mu^*}{(1+\tau^*\mu^*)} - c_d'(\mu^*) \\
&= \left(\frac{2w_o\mathbf{v}(1+\tau^*\mu^*)}{\log(1/\mathbf{b})\mu^*(e^{-\mu^*\bar{d}} - e^{-2\mu^*\bar{d}})}\right)^{1/2} \frac{e^{-\mu^*\bar{d}}\bar{d}\mu^*}{(1+\tau^*\mu^*)} - c_d'(\mu^*) \\
&= \left(\frac{2w_o\mathbf{v}\bar{d}^2\mu^*}{\log(1/\mathbf{b})(e^{\mu^*\bar{d}} - 1)(1+\tau^*\mu^*)}\right)^{1/2} - c_d'(\mu^*).
\end{aligned} \tag{42}$$

Set $g_3(\mathbf{b}) = \frac{2w_o\mathbf{v}\bar{d}^2\mu^*}{\log(1/\mathbf{b})(e^{\mu^*\bar{d}}-1)(1+\tau^*\mu^*)}$. Note that $g_3(\mathbf{b})$ is increasing in $\mathbf{b}$. This implies $g_3(\mathbf{b}) < g_3(b_3) \quad \forall \mathbf{b} \in (0, b_3)$. By definition of b_3 we get

$$\begin{aligned}
g_3(b_3) &= \frac{c_d'(\mu^*)^2(1+\tau^*\mu^*)(e^{2\mu^*/\underline{\mu}} - 1)\underline{\mu}^2 2w_o\mathbf{v}\bar{d}^2\mu^*}{8w_o\mathbf{v}\mu^*(e^{\mu^*\bar{d}} - 1)(1+\tau^*\mu^*)} \\
&= c_d'(\mu^*)^2 \frac{(e^{2\mu^*/\underline{\mu}} - 1)\underline{\mu}^2}{4} \frac{\bar{d}^2}{e^{\mu^*\bar{d}} - 1}.
\end{aligned} \tag{43}$$

Define a function $g_4(\bar{d}) = \frac{(\mu^*\bar{d})^2}{e^{\mu^*\bar{d}}-1}$. Note that $g_4(\bar{d})$ is decreasing on $\bar{d} \in [2/\underline{\mu}, \infty)$ since we have

$$\frac{\partial g_4(\bar{d})}{\partial \bar{d}} = \frac{\bar{d}\mu^{*2}e^{\mu^*\bar{d}}(2 - \bar{d}\mu^*) - 2\bar{d}\mu^{*2}}{(e^{\mu^* d} - 1)^2} < 0, \quad \forall \bar{d} \in [2/\underline{\mu}, \infty).$$

This implies,

$$\begin{aligned}
g_4(\bar{d}) &\leq g_4(\frac{2}{\underline{\mu}}), \quad \forall \bar{d} \in [2/\underline{\mu}, \infty), \\
\frac{(\mu^*\bar{d})^2}{e^{\mu^*\bar{d}} - 1} &\leq \frac{\left(\mu^*\frac{2}{\underline{\mu}}\right)^2}{e^{\mu^* 2/\underline{\mu}} - 1}, \\
\frac{\left(e^{\mu^* 2/\underline{\mu}} - 1\right)\underline{\mu}^2}{4} \frac{\bar{d}^2}{e^{\mu^*\bar{d}} - 1} &\leq 1.
\end{aligned} \tag{44}$$

since $\mu^* > \underline{\mu}$ and so $e^{\mu^* 2/\underline{\mu}} > 1$. Using (42), (43), and (44) we get $\frac{\partial \Pi_v}{\partial \mu}(\tau^*,\mu^*) < 0$. ■

Claim 4. $h(\tau^*,\mu,\bar{p}_4) < 0, \forall \mu \in [\underline{\mu}, \mu^*)$.

Proof: We rewrite the FDT constraint $h(\tau,\mu,\bar{p}) \geq 0$ as $\mu g_5(\tau,\mu,\bar{p}) \geq 0$, where

$$g_5(\tau,\mu,\bar{p}) = \Pi_v(\tau,\mu)(\frac{1}{\mu}+\tau) - \bar{p}^2(e^{-\mu\bar{d}} - e^{-2\mu\bar{d}})\frac{1}{2w_o}\log\frac{1}{\mathbf{b}}.$$

Note that since $\underline{\mu} > 0$ there is one-to-one mapping between the FDT constraint $h(\tau,\mu,\bar{p}) \geq 0$ and a modified FDT constraint defined on $g_5(\tau,\mu,\bar{p})$ as $g_5(\tau,\mu,\bar{p}) \geq 0$. For the remaining part of proof we will work with this modified constraint. We will show that $g_5(\tau^*,\mu^*,\bar{p}_4) = 0$ and $\partial_\mu g_5(\tau^*,\mu,\bar{p}_4) > \mathbf{0}$, $\forall\,\mu \in [\underline{\mu},\mu^*)$. Therefore, $g_5(\tau^*,\mu,\bar{p}_4) < 0$, $\forall\,\mu \in [\underline{\mu},\mu^*)$ i.e., the FDT constraint is violated on $\mu \in [\underline{\mu},\mu^*)$. Note that $g_5(\tau^*,\mu^*,\bar{p}_4) = 0$ is a direct outcome of $h(\tau^*,\mu^*,\bar{p}_4) = 0$ (see (39)) and $\mu^* \neq 0$. Now,

$$\frac{\partial g_5(\tau^*,\mu,\bar{p}_4)}{\partial\mu} = \partial_\mu\Pi_v(\tau^*,\mu)(\frac{1}{\mu}+\tau^*) - \Pi_v(\tau^*,\mu)\frac{1}{\mu^2} + \bar{p}_4^2(\bar{d}e^{-\mu\bar{d}} - 2\bar{d}e^{-2\mu\bar{d}})\frac{1}{2w_o}\log\frac{1}{\mathbf{b}}.$$

By (40) and (30) we get

$$\begin{aligned}
\frac{\partial g_5(\tau^*,\mu,\bar{p}_4)}{\partial\mu} &= \left(\frac{\bar{f}_4 - \bar{p}_4 e^{-\mu\bar{d}} + \bar{p}_4 e^{-\mu\bar{d}}\bar{d}\mu(1+\tau^*\mu)}{(1+\tau^*\mu)^2} - c_d'(\mu)\right)(\frac{1}{\mu}+\tau^*) - \left(w - c_d(\mu) + \frac{(\bar{f}_4 - \bar{p}_4 e^{-\mu\bar{d}})\mu}{1+\tau^*\mu}\right)\frac{1}{\mu^2} \\
&\quad + \bar{p}_4^2(\bar{d}e^{-\mu\bar{d}} - 2\bar{d}e^{-2\mu\bar{d}})\frac{1}{2w_o}\log\frac{1}{\mathbf{b}} \\
&= \frac{\bar{f}_4 - \bar{p}_4 e^{-\mu\bar{d}}}{(1+\tau^*\mu)\mu} + \bar{p}_4 e^{-\mu\bar{d}}\bar{d} - c_d'(\mu)(\frac{1}{\mu}+\tau^*) - \frac{w - c_d(\mu)}{\mu^2} - \frac{\bar{f}_4 - \bar{p}_4 e^{-\mu\bar{d}})\mu}{(1+\tau^*\mu)\mu} \\
&\quad + \bar{p}_4^2\bar{d}(e^{-\mu\bar{d}} - 2e^{-2\mu\bar{d}})\frac{1}{2w_o}\log\frac{1}{\mathbf{b}} \\
&= \bar{p}_4 e^{-\mu\bar{d}}\bar{d} - c_d'(\mu)(\frac{1}{\mu}+\tau^*) - \frac{c_d(\mu^*) + \mathbf{v} - c_d(\mu)}{\mu^2} + \bar{p}_4^2\bar{d}(e^{-\mu\bar{d}} - 2e^{-2\mu\bar{d}})\frac{1}{2w_o}\log\frac{1}{\mathbf{b}},
\end{aligned}$$

by definition of w. Since $\bar{p}_4 > 0$ and $\bar{d} > 0$ we get

$$\begin{aligned}
\frac{\partial g_5(\tau^*,\mu,\bar{p}_4)}{\partial\mu} &> -c_d'(\mu)(\frac{1}{\mu}+\tau^*) - \frac{c_d(\mu^*) + \mathbf{v} - c_d(\mu)}{\mu^2} + \bar{p}_4^2\bar{d}(e^{-\mu\bar{d}} - 2e^{-2\mu\bar{d}})\frac{1}{2w_o}\log\frac{1}{\mathbf{b}} \\
&= -c_d'(\mu)(\frac{1}{\mu}+\tau^*) - \frac{c_d(\mu^*) + \mathbf{v} - c_d(\mu)}{\mu^2} + \frac{\mathbf{v}(1+\tau^*\mu^*)\bar{d}(e^{-\mu\bar{d}} - 2e^{-2\mu\bar{d}})}{\mu^*(e^{-\mu^*\bar{d}} - e^{-2\mu^*\bar{d}})},
\end{aligned}$$

by definition of $\bar{p}_4$. Since $e^{-\mu\bar{d}} - 2e^{-2\mu\bar{d}}$ is decreasing in μ for all $\bar{d} \in [2/\underline{\mu},\infty]$ we get

$$\begin{aligned}
\frac{\partial g_5(\tau^*,\mu,\bar{p}_4)}{\partial\mu} &> -c_d'(\mu)(\frac{1}{\mu}+\tau^*) - \frac{c_d(\mu^*) + \mathbf{v} - c_d(\mu)}{\mu^2} + \frac{\mathbf{v}(1+\tau^*\mu^*)\bar{d}(e^{-\mu^*\bar{d}} - 2e^{-2\mu^*\bar{d}})}{\mu^*(e^{-\mu^*\bar{d}} - e^{-2\mu^*\bar{d}})} \\
&= -c_d'(\mu)(\frac{1}{\mu}+\tau^*) - \frac{c_d(\mu^*) + \mathbf{v} - c_d(\mu)}{\mu^2} + \frac{\mathbf{v}(1+\tau^*\mu^*)}{\mu^*}\bar{d}\left(1 - \frac{1}{e^{\mu^*\bar{d}} - 1}\right) \\
&> 0, \quad \forall\,\mu \in [\underline{\mu},\mu^*],
\end{aligned}$$

by (15). ■

Endnotes

[1]www-03.ibm.com/press/us/en/pressrelease/435.wss

[2]In a controlled experiment conducted by a data center client of Sun Microsystems, a 14% marginal increase in the failure rate of servers was observed when the operating conditions deviated by a unit temperature from the recommended range. www.datacenterknowledge.com/archives/2011/02/25/whats-next-hotter-servers-with-gas-pedals/

[3]Raymond Corporation offers a range of products and after-sale services, including field service support for repair of forklift trucks.

[4]Estimates of the US National Institute for Occupational Safety and Health indicate that about 80% of forklift repair and maintenance costs are due to operator abuse of the equipment. www.esc.org/newsletter/May_Newsletter__rev__.pdf

[5]GFOA, 2010 annual conference report entitled "Getting the Most Out of Outsourcing: Performance Based Contracting."

[6]Best Practices in Outsourcing and Vendor Management. http://360vendormanagement.com/2008/04/03/outsourcing-contract-penalties-do-vendors-respond-to-the-pain

[7]A few exceptions (see, e.g., Baiman et al. 2000, Corbett et al. 2005) have shown the first-best outcome to be attainable (in the double-sided moral hazard setting) under special conditions; these works are discussed in Section 2 and Section 4.2.1.

[8]Ruin theory provides well-established models that study the financial risks of insurance firms in settings of stochastic wealth evolution.

[9]In some cases the client may choose to outsource scheduled preventive maintenance as well. We do not capture this in our model explicitly, but note that typically outsourced scheduled preventive maintenance "activities" do not induce any moral hazard. Such activities can be viewed as well-documented procedural work, such as periodic oil replacement, that can be enforced through a

direct contract (e.g., a time-and-materials contract).

[10]In 2010, DBS Bank was assessed a fine of 230 million Singapore dollars by the Central Bank of Singapore for disruption pf the bank's ATM, online, and mobile services for more than seven hours.

[11]Often such contracts have multiple performance brackets (tiers), but in this paper we use a simple single-tiered structure for tractability. However, the insights of our analysis are robust to a more general tiered contract structure.

[12]Following the literature on service contracting, we use the equilibrium profit rate for our analysis (cf. Hasija et al. 2008, Ren and Zhou 2008).

[13]The wealth evolution model of an insurance firm is $X(t) = u + at - \sum_{i=1}^{M_t} x_i$, where u is the initial wealth of the firm, a is the per unit time premium earned by the firm, x_i is the stochastic cash outflow due to the ith claim, and M_t is the stochastic number of claims up to time t.

References

S. Asmussen and H. Albrecher. *Ruin probabilities*, volume 2001. World Scientific Singapore, 2000.

V. Babich. Independence of capacity ordering and financial subsidies to risky suppliers. *Manufacturing & Service Operations Management*, 12(4):583, 2010.

S. Baiman, P. Fischer, and M. Rajan. Information, contracting, and quality costs. *Management Science*, 46: 776–789, 2000.

N. Bakshi, S. Kim, and N. Savva. Performance contracting in service-support supply chains with an informed manufacturer. Ann Arbor, Michigan, June 2011. MSOM Annual Conference.

S. Bhattacharya, V. Gaba, and S. Hasija. Coordinating bilateral investments in a healthcare R&D partnership. Technical report, Working paper, 2011.

S. Bhattacharyya and F. Lafontaine. Double-Sided Moral Hazard and the Nature of Share Contracts. *The Rand Journal of Economics*, 26:761–781, 1995.

T. Choi and J. Hartley. An exploration of supplier selection practices across the supply chain. *Journal of operations management*, 14(4):333–343, 1996. ISSN 0272-6963.

R. Cooper and T. Ross. Product warranties and double moral hazard. *The Rand Journal of Economics*, 16 (1):103–113, 1985. ISSN 0741-6261.

C. Corbett and G. DeCroix. Shared-savings contracts for indirect materials in supply chains: Channel profits and environmental impacts. *Management Science*, pages 881–893, 2001. ISSN 0025-1909.

C. J. Corbett, G. A. DeCroix, and A. Y. Ha. Optimal shared-savings contracts in supply chains: Linear contracts and double moral hazard. *European Journal of Operational*, 163:653–667, 2005.

R. Cummins, K. Chesemore, and R. White. Defibrillator failures. Causes of problems and recommendations for improvement. Defibrillator Working Group. *The Journal of the American Medical Association*, 264(8):1019–25, 1991.

L. Dong and B. Tomlin. Managing disruption risk: The interplay between operations and insurance. Technical report, Working paper, 2011.

S. Hasija, E. J. Pinker, and R. A. Shumsky. Call center outsourcing contracts under information asymmetry. *Management Science*, 54(4):793 – 807, 2008. ISSN 00251909.

K. Hendricks and V. Singhal. An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. *Production and Operations Management*, 14(1): 35–52, 2005.

B. Holmstrom. Moral hazard in teams. *The Bell Journal of Economics*, pages 324–340, 1982.

D. Iglehart. Diffusion approximations in collective risk theory. *Journal of Applied Probability*, 6(2):285–292, 1969. ISSN 0021-9002.

S. Kim, M. Cohen, and S. Netessine. Performance contracting in after-sales service supply chains. *Management Science*, 53(12):1843–1858, 2007. ISSN 0025-1909.

S. Kim, M. Cohen, and S. Netessine. Reliability or inventory? Analysis of product support contracts in the defense industry. Technical report, Working paper, 2009.

S. Kim and S. Wang. Linear contracts and the double moral-hazard. *Journal of Economic Theory*, 82(2): 342–378, 1998.

D. Murthy. A note on minimal repair. *Reliability, IEEE Transactions on*, 40(2):245–246, 1991.

D. Murthy and E. Asgharizadeh. Optimal decision making in a maintenance service operation. *European Journal of Operational Research*, 116(2):259–273, 1999. ISSN 0377-2217.

Z. Ren and Y. Zhou. Call center outsourcing: Coordinating staffing level and service quality. *Management Science*, 54(2):369–383, 2008.

G. Roels, U. Karmarkar, and S. Carr. Contracting for collaborative services. *Management Science*, 56(5): 849–863, 2010. ISSN 0025-1909.

R. Swinney and S. Netessine. Long-term contracts under the threat of supplier default. *Manufacturing & Service Operations Management*, 11(1):109–127, 2009. ISSN 1526-5498.

Europe Campus
Boulevard de Constance
77305 Fontainebleau Cedex, France
Tel: +33 (0)1 60 72 40 00
Fax: +33 (0)1 60 74 55 00/01

Asia Campus
1 Ayer Rajah Avenue, Singapore 138676
Tel: +65 67 99 53 88
Fax: +65 67 99 53 99

Abu Dhabi Campus
Muroor Road - Street No 4
P.O. Box 48049
Abu Dhabi, United Arab Emirates
Tel: +971 2 651 5200
Fax: +971 2 443 9461

www.insead.edu

INSEAD

The Business School
for the World®

Printed by INSEAD