Performance Contracting in After-Sales Service Supply Chains[*]

Sang-Hyun Kim, Morris A. Cohen, and Serguei Netessine[†]

The Wharton School, University of Pennsylvania

Philadelphia, PA 19104

January 11, 2006

Abstract

Performance-based contracting is reshaping service support supply chains in capital intensive industries such as aerospace and defense. Known as "power by the hour" in the private sector and as performance-based logistics (PBL) in defense contracting, it aims to replace traditionally used fixed-price and cost-plus contracts in order to improve product availability and reduce the cost of ownership by tying a supplier's compensation to the output value of the product generated by the customer (buyer).

To analyze implications of performance-based relationships, we introduce a multitask principal-agent model to support resource allocation and use it to analyze commonly observed contracts. In our model the prime (principal) faces a product availability requirement dictated by its customer for the "uptime" of the end product. The prime then offers contracts contingent on availability to n suppliers (agents) of the key subsystems used in the product, who in turn exert cost reduction efforts and set spare parts inventory investment levels. We show that the first-best solution can be achieved if channel members are risk-neutral. When channel members are risk-averse, we find that the second-best contract combines a fixed payment, a cost-sharing incentive and a performance incentive. Furthermore, we show how these contracts evolve over the product deployment life cycle as product use and support cost risks change. We show, in particular, that when the prime is less (more) risk-averse than the suppliers, the performance incentive increases (decreases) while the cost sharing incentive decreases (increases) with time. Finally, we illustrate the application of our model to a problem based on aircraft maintenance data and show how the allocation of performance requirements and contractual terms change under various environmental assumptions.

1   Introduction

Support and maintenance services continue to constitute a significant part of the U.S. economy, often generating twice as much profit as do sales of original products. For example, a study by Accenture (see [1]) found that $9B in after-sales revenues produced $2B in profits for General Motors, which is a much higher rate of profit than its $150B in car sales generated over the same time period. According to the same study, after-sales services and parts contribute only 25% of revenues across all manufacturing companies but are often responsible for 40-50% of profits.

Since maintenance services are often provided and consumed by two different organizations (i.e., the OEM and the customer), the issue of contracting between them becomes important. While contracts for maintenance services of simpler products (electronics, automobiles) often involve fixed payments for warranties, there are many instances of complex systems that require more sophisticated relationships between service buyers and suppliers. For example, in capital-intensive industries such as aerospace and defense, it is very hard to guarantee product availability due to significant uncertainties in product reliability and usage as well as inherent product complexity, resulting in large risks to both the customer and service provider. Therefore maintenance support in these industries is typically conducted using fixed-price or cost-plus contracts: under the former, the buyer of support services pays a fixed fee to the supplier to purchase necessary parts and support services, whereas under the latter the supplier repairs the product and charges full cost plus a premium to the buyer. Studies in the defense industry (see http://www.pblprograms.com) estimate that 80% of current maintenance contracts are cost-plus, and the remaining 20% are fixed-price.

Through our work with major defense contractors we observe a major shift in the world of support and maintenance logistics for complex systems over the past few years. Performance-based contracting, a novel approach for the sustainment business, is replacing traditional service procurement practices. This approach is often referred to as "power by the hour" or performance-based logistics (PBL) in, respectively, the airline and defense industries. The idea behind it is quite simple: one buys the results of product use (e.g., value creation), not the service parts or repair services required to restore or maintain a product. The premise behind performance-based contracting is elaborated in the official Department of Defense (DoD) guidelines[1]:

> The essence of Performance Based Logistics is buying performance outcomes, not the individual parts and repair actions... Instead of buying set levels of spares, repairs, tools, and data, the new focus is on buying a predetermined level of availability to meet the [buyer's]

---

[1]Excerpt from Defense Acquisition Guidebook Section 5.3 (http://akss.dau.mil/dag).

objectives.

Performance-based contracts originally were implemented in commercial settings predominantly for avionics products. For example, engine manufacturers General Electric, Pratt & Whitney, and Rolls Royce all have performance-based contracts with commercial airlines in which their compensation is tied to product availability (hours flown). Recently, the U.S. Department of Defense has initiated the implementation of pilot PBL programs in the military; in 2005 there were 92 such programs, compared with 57 programs in 2002 (see http://www.pblprograms.com). Among frequently cited PBL success stories are avionics for the H-60 helicopter contract that brought logistic response time (LRT) down from 52.7 days pre-PBL to 8 days after PBL implementation and the F/A-18 Hornet aircraft contract with a pre-PBL LRT of 42.6 days and a post-PBL LRT of 2 days, among many others. Inspired by such notable success of PBL contracts, on August 16, 2004, the DoD issued Memorandum 5000.1, which "requires program managers to develop and implement PBL strategies that optimize total system availability," thus mandating all future maintenance contracts to be based on performance.

A critical element of performance-based contracting is the clear separation between the buyer's expectations of service (the performance goal) and the supplier's implementation (how it is achieved): in other words, "The contract explicitly identifies what is required, but the contractor determines how to fulfill the requirement" (Macfarlan and Mansir [16]). As a consequence, PBL contracting should promote new and improved ways to manage spare parts inventory, negotiate contracts, and make resource allocation decisions. For example, under the traditional cost-plus contract, the supplier of a service must truthfully report its detailed cost structure to the buyer in order to estimate which exact expenses are eligible for reimbursement. Under a PBL arrangement, the supplier does not have to support cost sharing at this level of detail. Moreover, the product buyer no longer directly manages or possibly even owns resources such as the inventory of spares and thus is not concerned with specifics such as inventory stocking, as long as the availability target is met. Finally, in the long run suppliers may find it in their interest to invest in designing and producing more reliable products (i.e., with lower part failure rates) and/or more efficient repair and logistics capabilities.

Not surprisingly, such a radical change in the approach to contracting with the DoD has caused controversy among suppliers of maintenance services. For example, among 128 suppliers whose bids were solicited on a PBL support contract, only 5 responded positively, and the rest simply responded "not interested" (see http://www.pblprograms.com). As chief logistician of Northrop Gruman (one of the major DoD contractors) put it, "after nearly five years since its inception, PBL still is generating a great deal of discussion and will engender a major cultural and responsibility change at the supplier level" (see Phillips [19]). Moreover, the purported benefits of PBL arrangements came under the

scrutiny of the Government Accountability Office, which recently concluded that "DoD program offices could not demonstrate that they have achieved cost savings or performance improvements through the use of performance based logistics arrangements."[2]

The preceding discussion underscores the urgency and lack of understanding of PBL arrangements. The academic literature, however, offers little guidance with respect to how such contracts should be executed. In this paper we aim to take a first step towards filling this void by proposing a model of contractual relationships that arise in practice when procuring repair and maintenance services in a performance-based environment. We embed a standard single-location spare parts inventory management problem into a moral hazard model with one principal (representing the prime supplier of the product or the end customer), and multiple interdependent agents (representing suppliers of the key product subsystems), in which each agent (supplier) performs two tasks: inventory management of spares and cost reduction activities. We use this model to analyze three types of contracts (and any combination thereof) that are commonly encountered in aerospace and defense procurement and high technology industries: fixed-price, cost-plus and PBL. In analyzing these contracts we pursue the following goals: (1) what is the optimal combination of contractual levers that achieves the best possible outcome for the buyer? (2) how should performance requirements for the final product be allocated to suppliers? and (3) how should the risk associated with the maintenance of complex equipment be shared among channel members?

We show that, in the absence of incentive problems (i.e., if suppliers' decisions are observable and contractible), the contract that achieves the first-best solution is a nonperformance arrangement that combines partial cost reimbursement with a fixed payment. If supplier actions are unobservable and the parties are risk-neutral, we show that the first-best solution can still be achieved using a contract that combines a performance incentive with a fixed payment (but no cost sharing). However, when even one of the parties is risk-averse, the first-best solution cannot be achieved. We show that in this case "pure" fixed-price, cost-plus or performance-based contracts (or any pair-wise combinations of them) are not suitable because they do not provide the necessary incentives. Thus, we show that the second-best contract involves all three elements: a combination of a fixed payment, a cost sharing payment and a performance-based payment. For any such contract we show that each supplier's problem is well-behaved (quasi-concave) under suitable parameter restrictions and we find analytically optimal decisions for all suppliers for any given contract proposed by the buyer. Unfortunately, the buyer's problem neither is well-behaved nor admits to tractable analytical solutions (the latter is true even in the centralized supply chain). Using a combination of analytical results for special cases and numerical

---

[2]See http://www.gao.gov/cgi-bin/getrpt?GAO-05-966.

4

analysis performed on a data set that is representative of a supply chain supporting a fleet of military airplanes, we obtain insights into the structure of the optimal contract. In particular, we study the sensitivity of the optimal contract to an operating characteristic (i.e., cost uncertainty) and infer that, when the principal is less (more) risk-averse than the suppliers, the performance incentive increases (decreases), whereas the cost sharing incentive decreases (increases) as time progresses. Finally, we analyze the impact of problem parameters on contractual terms, performance, and profitability.

To the best of our knowledge, this paper represents the first attempt to embed the after-sales service supply chain model into the principal-agent framework in which channel members behave in a self-interested manner. Our results are consistent with the observed practice of using multiple contract types whose mix evolves over time. Finally the model framework introduced here can be implemented in conjunction with more detailed supply chain models to support contract negotiations and long-term strategy analysis. The rest of the paper is organized as follows. After a brief review of related literature in Section 2, we present modeling assumptions and notations in Section 3, followed by the formulation of the principal-agent model. In the same section we analyze the first-best solution as well as derive solutions for the general second-best case. In Section 4 we analyze special cases, beginning with the risk-neutrality assumption, then an assumption of partial observability of suppliers' actions, and finally a situation with one supplier. The section concludes with a numerical analysis of the practical set of data. Finally, in Section 5 we discuss managerial implications of our study.

2   Literature Review

Two distinct models blend together in our paper: a classic inventory allocation model for repairable items, well known in operations management, and the moral hazard model that has been an area of active research in economics. The theory of repairable parts inventory management dates back to the 1960s when Feeney and Sherbrooke [11] introduced a stochastic model of the repairable inventory problem whose steady-state solution relies on the application of Palm's Theorem. Sherbrooke's MET-RIC model (Sherbrooke [23]) introduced a heuristic optimization algorithm for allocating inventory resources for the multi-echelon, multi-indentured version of the problem. Subsequent models have led to notable success in enabling the management of multimillion-dollar service parts inventory resources in both commercial and government applications (e.g., see Cohen et al. [8] for a discussion of a successful application of multi-echelon optimization by IBM's service support division). Research in this area has largely focused on improving computational efficiency and incorporating more realistic assumptions, such as allowing for capacitated supply or nonstationary demand processes. For a recent comprehensive account of developments in this field, see Muckstadt [18], who reviews the underlying

theory, Sherbrooke [24], which focuses on aerospace and defense industry applications, and Cohen et al. [7], which introduces a modeling framework that has been used to guide the development of state-of-the art software solutions in various industries. In brief, repairable inventory models are typically concerned with finding the optimal (cost-minimizing) inventory stocking targets for each product component subject to an overall service constraint. Service (performance) requirements can be defined in terms of either item fill rates or end product availability (i.e., system "uptime"). The latter is preferred in our context of performance contracting because there is a one-to-one mapping between inventory investment and item backorders: the latter drive overall product delay, which in turn drives product uptime.

It is important to note that while the one-location variant of the availability problem is well behaved (i.e., convex), extensions that include multiple echelons/locations, material classes/indentures and alternative sourcing options (new buy, repair, internal transfer) lead to large-scale nonlinear, non-convex, stochastic optimization problems with millions of decision variables and thousands of constraints. Current state-of-the-art solutions are based on variants of the greedy heuristic introduced by Sherbrooke.

Numerous papers study the principal-agent models, and comprehensive reviews can be found in Bolton and Dewatripont [3]. The building block for our paper is the moral hazard model in which actions of agents (suppliers) are unobservable to the principal (buyer). Moreover, our model includes elements of multitasking (Holmström and Milgrom [13]), because two decision variables for suppliers, the cost reduction effort and the inventory position, interact with each other. An additional complication is the presence of multiple agents whose contracts are interdependent due to the performance constraint that the principal faces. The mainstream interest in the principal-agent theory is in designing optimal *nonlinear* contracts. Despite the theoretical appeal, the predominant form of contracts observed in practice is linear. Holmström and Milgrom [12] addressed this discrepancy by explaining that a sequence of repeated observations of performance outcomes necessitates a simple linear relationship between the aggregate performance and the aggregate payment, effectively collapsing the multi-period dynamic contracting problem into a single-period problem. In this paper we take a descriptive approach and assume that the linear contract form is exogenously specified, which is consistent with our observations of industrial practices in the defense and other industries. The economics literature that studies contracting for defense procurement takes the same approach when analyzing incentives. For example, Scherer [22] discusses linear (cost-plus and fixed-price) contracts as well as the impact of risk aversion in defense contracting. Similar to Scherer's work, we allow for risk aversion and study cost-plus and fixed-price contracts in the context of maintenance and compare them with performance contracts. Cummins [10] studies risk sharing and the role of risk aversion in defense contracts.

Incentive alignment in supply chains through contracts has been a topic of great interest in oper-

ations management over the past decade (see Cachon [4] for a comprehensive survey). Recently, the role of information asymmetry has received considerable attention both in the adverse selection setting (representative articles include Corbett [9], Iyer et al. [14], Lutze and Ozer [15] and Su and Zenios [25]) and in the moral hazard setting (for example, see Plambeck and Zenios [21], Chen [5] and Plambeck and Taylor [20]). The current paper addresses the growing interest in this area.

As is evident from our survey, although there is voluminous literature on service supply chain parts-inventory management, to date this stream of research has been confined to single-firm models and hence does not address issues that arise in decentralized supply chains in practice. Furthermore, although extensive literature in economics aims to model contractual relationships among different parties, it does not address the complexities of repair and maintenance contracting environments. To our knowledge, our paper is the first to put a repairable parts model into the decentralized framework and to study the issue of contracting in after-sales service supply chains.

3   Model

3.1   Modeling Assumptions

The principal, henceforth called the prime, is the prime supplier of N assembled products ("systems," which can be airplanes, computers, manufacturing equipment, etc.) to customers such as airlines, branches of the military or industrial companies. Each system is composed of n distinct major parts ("subsystems" which, in the case of an airplane, can represent avionics, landing gear, weapons systems, etc.), each produced and maintained by a unique supplier. We use subscript 0 to denote the prime and subscript i for subsystem supplier i, $i = 1, 2, ..., n$. Failure of subsystem i occurs at a Poisson rate λ_i. Each supplier maintains an inventory of spares and a repair facility. A failed unit is immediately replaced by an operating unit (if it is available) from the supplier's inventory with transportation lead times assumed to be negligible. If a replacement is unavailable, a backorder occurs, and the system becomes inoperable. As a result, downtime in any subsystem leads to downtime of the entire system. Upon failure, the defective unit immediately goes into the repair facility, modeled as an M/G/∞ queue (i.e., we assume ample repair capacity). It takes on average L_i time units to repair and ship the subsystem, and once the task is completed the subsystem is placed in the supplier's inventory, i.e., we assume one-for-one replenishment to a target stock level. Therefore, we have a closed-loop cycle for the repair process. As the subsystems are typically very expensive and their lifetimes are very long, we assume that no subsystem is discarded during the entire support period. We also ignore the breakdown of subsystems into multiple components which typically include line replaceable units (LRUs) as well as lower indenture level components, i.e., each subsystem is treated as a single composite item. Figure
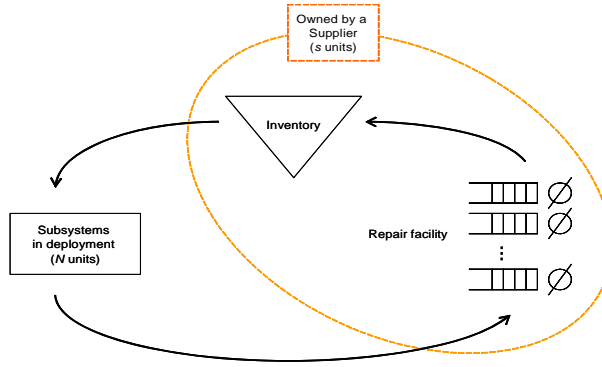
Figure 1: Closed loop cycle for repairable items.

1 illustrates this process. One of supplier i's decisions is the target stocking level for subsystem spare parts inventory s_i which determines the fill rate as well as the expected number of backorders. Evidently, there is a total of $N + s_i$ units of subsystem i in the supply chain, but only s_i of them are owned by the supplier.[3]

The number of backorders, B_i, is a random variable that is observed continuously. B_i and s_i are related to each other through $B_i = (O_i - s_i)^+$, where O_i is a stationary random variable representing the pipeline (on-order) inventory. Palm's Theorem states that O_i is Poisson-distributed, with the mean $\mu_i = \lambda_i L_i$ (see Feeney and Sherbrooke [11]). Although this observation leads to closed-form expressions for system performance metrics, it turns out that working with integer-valued random and decision variables complicates our analysis significantly, as additional complexity results from the game-theoretic situation considerations associated with the various contracting options. In particular, conducting comparative statics to gain insights into firms' behavior is prohibitively complex. For this reason, we depart from the usual discrete Poisson process assumption and model O_i, B_i, and s_i as continuous variables. This approach is reasonable in our context since each unit of a supplier's inventory represents a composite of the various LRUs and the components associated with their particular subsystem. To this end, we let O_i be distributed continuously with cdf F_i and pdf f_i, which have nonnegative support $[0, \infty)$ with $F_i(0) \geq 0$. μ_i, the mean of O_i, is determined by the mean failure rate times the average repair lead time, consistent with the Palm Theorem result noted above. The distribution of B_i, denoted by $G_i(\cdot \mid s_i)$, is related to $F_i(\cdot)$ through $G_i(x \mid s_i) = F_i(x + s_i)$, which is obtained from $P(B_i \leq x \mid s_i) = P(O_i \leq x + s_i)$. Furthermore,

$$E\left[B_i \mid s_i\right] = \int_0^\infty \left[1 - G_i(x \mid s_i)\right] dx = \int_0^\infty \left[1 - F_i(x + s_i)\right] dx$$

---

[3]We also note that an important consideration in implementing PBL relationships in practice concerns defining asset ownership and managerial controls across the three echelons (customer, prime and subsystem supplier). We have observed that many possibilities are being experimented with in practice.

so that we obtain

$$\partial E\left[B_i \mid s_i\right] / \partial s_i = -1 + F_i\left(s_i\right) \leq 0, \qquad \partial^2 E\left[B_i \mid s_i\right] / \partial s_i^2 = f_i\left(s_i\right) \geq 0. \tag{1}$$

Hence we see that expected backorder is decreasing and convex in s_i.

The performance metric of our problem is availability, which is defined as the fraction of time a subsystem is operational. Availability is a random variable related to the backorder through $A_i = 1 - B_i/N$.[4] Since the ratio $E\left[B_i \mid s_i\right]/N$ is typically very small, the expected availability can be approximated as follows:

$$E[A_0 \mid s_1, s_2, ...s_n] = \prod_{i=1}^{n} E[A_i \mid s_i] = \prod_{i=1}^{n} \left(1 - E\left[B_i \mid s_i\right] / N\right) \simeq 1 - \sum_{i=1}^{n} E\left[B_i \mid s_i\right] / N.$$

We see that the relation $A_i = 1 - B_i/N$ can be applied to the system $(i = 0)$ as well if we define $B_0 \equiv \sum_{i=1}^{n} B_i$, i.e., the overall system backorder is the sum of subsystem backorders. With this definition, the system availability requirement $E[A_0 \mid s_1, s_2, ...s_n] \geq \widehat{A}_0$ (e.g., "expected system availability has to exceed 95%") is equivalent to the system backorder constraint $E\left[B_0 \mid s_1, s_2, ...s_n\right] = \sum_{i=1}^{n} E\left[B_i \mid s_i\right] \leq \widehat{B}_0$. We call $\widehat{B}_0$ the system backorder target. The additive separability of B_0 in terms of $\{B_i\}_{i=1,...n}$ relies on the assumption that the probability of two or more subsystems being backordered at a given point in time is negligible, thus implying that there is no ambiguity in assigning accountability for system downtime to a specific supplier and that all system failures are caused by single subsystem failure. We note that the sufficient condition for this assumption to hold is $Cov[B_i, B_j] = 0$, $i \neq j$. We also assume that $\sum_{i=0}^{n} \mu_i > \widehat{B}_0$ in order to rule out the trivial case where $s_1 = s_2 = ... = s_n = 0$ is optimal.

We assume that supplier i's total cost to maintain its subsystem, C_i, has fixed and variable components with an additive stochastic term ε_i, and it can be reduced by supplier's effort a_i so that $C_i = c_i s_i + K_i - a_i + \varepsilon_i$. Thus, the degree of cost uncertainty is assumed to be beyond the supplier's control (e.g., establishing a maintenance network in a different country where the fleet of airplanes is deployed involves large uncertainties beyond the control of the supplier). We note that the inherent uncertainty associated with subsystem failure is captured in the inventory demand process and is reflected in the inventory on-order distribution. Without loss of generality, we normalize the fixed cost to zero, $K_i = 0$ for all i. We assume that cost uncertainties ε_i have zero mean and a finite variance and are uncorrelated, i.e., $Cov[\varepsilon_i, \varepsilon_j] = 0$ for $i \neq j$. Furthermore, we assume that $Cov[\varepsilon_i, B_i] = Cov[\varepsilon_i, B_j] = 0$

---

[4]This relationship is based on the ergodic property of the stationary random variable O_i, i.e., the long-run sample path frequency distribution is identical to the probability distribution. In other words, the fraction of time we observe O_i to be less than or equal to x has the same probability as $F_i(x \mid s_i)$.

holds for all i, j, i.e., the two sources of randomness, product performance/failure and maintenance cost variation, are assumed to be independent. By exerting effort a_i the supplier incurs monetary disutility $\psi_i(a_i)$, which is convex increasing ($\psi_i'(a_i) > 0$, $\psi_i''(a_i) > 0$), with $\psi_i(0) = 0$. In the sequel, we assume quadratic functional form $\psi_i(a_i) = k_i a_i^2 / 2$ with $k_i > 0$. This assumption does not fundamentally change the insights of our model, while generating compact expressions,[5] and for this reason it is commonly used in the literature (see, for example, Chen [5]).

The prime supplier's objective is to maximize her expected utility function subject to the system availability requirement, or equivalently the backorder requirement. Her utility is a function of her total expenditure only, although it is possible to incorporate a fixed revenue term to capture the expected profit associated with delivering the fleet availability to the end customer. This expenditure is a sum of transfers to the suppliers, each of which is comprised of (1) a fixed payment, (2) reimbursement for the supplier's cost and (3) a backorder-contingent incentive payment. In the performance-based contracting environment, neither the details of supplier cost nor how he meets performance objectives is revealed to the prime. Instead, each supplier is compensated for his total realized cost C_i and his realized backorder B_i. The fact that both contractible variables are random raises the issue of incentives. Since C_i and B_i are functions of the supplier's cost reduction effort a_i and base stock level decision s_i, the supplier can partially control the performance related to his subsystem and his compensation by setting a_i and s_i. However, stochasticity means he may choose (a_i, s_i) that are not optimal from the prime's point of view. For example, an opportunistic supplier may choose to minimize his own disutility of efforts by "shirking" (i.e., choosing low a_i and s_i), hoping that a fortuitous state of the world is realized. The prime's task is then to provide appropriate incentives through contract terms that would induce the supplier to perform the desired action.

Specifically, the contract that the prime offers to supplier i has the form

$$T_i(C_i, B_i) = w_i + \alpha_i C_i - v_i B_i, \tag{2}$$

where w_i, α_i, and v_i are the contract parameters determined by the prime; w_i is the fixed payment, α_i is the prime's share of the supplier's costs, and v_i is the penalty rate for each backorder incurred by the supplier. With $v_i = 0$ and $\alpha_i = 0$, we obtain a fixed-price (FP) contract whereas with $\alpha_i = 1$ and $v_i = 0$ we obtain a cost-plus (C+) contract with full reimbursement.

The crucial distinction between the supplier's actions a_i and s_i is the way each variable contributes to the performance outcomes; the backorder is influenced by s_i only, since $B_i = (O_i - s_i)^+$, whereas the total cost is affected by both decision variables, $C_i = c_i s_i - a_i + \varepsilon_i$. This interaction creates

---

[5]It turns out the condition $\psi_i'''(a_i) \geq 0$ is sufficient to ensure solution uniqueness in later analysis.

asymmetry in how the suppliers' actions influence outcomes B_i and C_i. Raising a_i reduces the total cost but has no impact on availability, which is driven by component reliability and customer usage patterns. That said, raising s_i improves availability but incurs a higher cost. The latter is the classical cost-availability trade-off seen in the repairable parts inventory theory. We note that an alternative formulation, whereby supplier effort impacts product reliability and/or repair capabilities (thereby impacting λ_i and L_i), is not considered here and will be the subject of a follow-up paper.

We assume that all members of the supply chain are risk-averse with expected mean-variance utility

$$E[U_i(X)] = E[X] - r_i Var[X]/2. \tag{3}$$

The constant r_i is the risk aversion factor, representing the inherent attitude towards uncertainty. Risk aversion is common among defense contractors, for example, because of great uncertainties that pervade product development, production, and maintenance (see Scherer [22] for discussion and references). The larger the value of r_i, the more risk-averse a firm is, whereas risk neutrality is a special case with $r_i = 0$. This form of utility function is widely used in finance as a basis of mean-variance portfolio theory; see Markowitz [17]. This form of utility function is exact for the constant absolute risk aversion utility function ($U_i(X) = -e^{-rX}$) with a Normally distributed error term ($X \sim \text{Normal}(\mu, \sigma^2)$). For other distributions and utility functions expression (3) is merely an approximation obtained by expanding $U_i(X)$ around $E[X]$ in a Taylor series up to the second-order term. This form of utility function has been widely used in recent operations management literature because of its tractability (Chen and Federgruen [6], Van Mieghem [26]). In our setting risk aversion constants $\{r_i\}$ are likely to be quite small for all supply chain members because they are mostly multinational, multibillion-dollar corporations (see Cummins [10] for evidence that risk aversion is negatively associated with firm size in defense contracting). Thus, the mean-variance approximation does not cause significant distortions and provides a "good recommendation" on how to quantify a firm's risk aversion even if its utility function is unknown (Van Mieghem [26]).

Our modeling setup and assumptions require some discussion. In trying to come up with a realistic model of the maintenance relationship, we do not venture beyond a single-period steady-state setup with moral hazard, since there are enough complexities and richness in our model to merit close inspection before extending the findings to other settings. Indeed, even the centralized problem in the absence of agency issues is quite complex, so we use the most basic version of it (i.e., unlimited repair capacity, stationary failures, etc.), and relaxing any of these assumptions is likely to obscure our findings regarding contracting issues. A single-period, steady-state assumption is a plausible simplification of reality, whereas more complex multi-echelon, multi-indenture models are applied in practice to

determine actual spare part deployments. Due to uncertainties in fleet deployment schedules and future support budgets, the DoD is unwilling to sign long-term contracts (i.e., for the life of the program), and instead typically contracts on a shorter-term basis with annual adjustments. Suppliers typically conduct multi-period budget planning using a single-period steady-state model on a rolling-horizon basis. Although pre-contractual bargaining or renegotiation may exist in practical situations, we do not formally model them and assume that the prime offers take-it-or-leave-it contracts to the suppliers.[6] With respect to the moral hazard elements of the model, we assume that neither inventories of spare subsystem parts nor efforts are observable but that backorders and total costs are observable and contractible. Observability of backorders is a natural assumption because the system becomes non-operational during the backorder, as is clearly observable. Furthermore, our assumption with respect to observability of the total cost implies that sharing of information regarding costs happens on the aggregate level (i.e., for the program budget) rather than at a detailed level (i.e., for subsystem spare parts). This assumption is in the spirit of the PBL arrangement (see above) in which the buyer is not concerned with verifying details about how availability was achieved (e.g., through a combination of cost-reduction efforts and subsystem spare parts investment), which is typical in related papers (see, e.g., Bajari and Tadelis [2]). Of course, it is often hard to measure or define product-specific maintenance costs and some degree of arbitrariness is inevitable in this process. Furthermore, we assume that there is uncertainty with respect to the total cost for maintaining each subsystem but not with respect to its unit cost. This is a plausible assumption in the maintenance industry where the unit cost of spares is easy to estimate based on maintenance and procurement information but where fixed costs for support (e.g., warehouses, overhead, etc.) are largely uncertain because these costs depend greatly on where and how systems are deployed, which is usually unknown a priori. In our discussions with companies involved in such contracts we found that the uncertainty with respect to fixed costs is of greater importance during the maintenance stage whereas the uncertainty with respect to unit cost might be more important during the product acquisition stage, which we do not model.

Under the assumptions we have laid out so far, supplier i who is given a contract $T_i(C_i, B_i)$ has the following expected utility:

$$
\begin{aligned}
E\left[U_i\left(T_i(C_i, B_i) - C_i - \psi_i(a_i)\right) \mid a_i, s_i\right] &= w_i - (1 - \alpha_i)(c_i s_i - a_i) - v_i E\left[B_i \mid s_i\right] - k_i a_i^2/2 \\
&\quad - r_i(1 - \alpha_i)^2 Var[\varepsilon_i]/2 - r_i v_i^2 Var[B_i \mid s_i]/2.
\end{aligned}
\tag{4}
$$

Similarly, the prime's expected utility $E[U_0]$ is

---

[6]We note that our model could be used in a "what-if" manner to support analysis of such negotiations.

$$E\left[U_0\left(-\sum_{i=1}^n T_i(C_i, B_i)\right) \mid \{a_i, s_i\}\right] = -\sum_{i=1}^n \left(w_i + \alpha_i(c_i s_i - a_i) - v_i E\left[B_i \mid s_i\right]\right) \quad (5)$$
$$+ r_0 \alpha_i^2 Var[\varepsilon_i]/2 + r_0 v_i^2 Var[B_i \mid s_i]/2)$$

where we have assumed independence across subsystems, i.e., $Cov[\varepsilon_i, \varepsilon_j] = Cov[B_i, B_j] = Cov[\varepsilon_i, B_i] = Cov[\varepsilon_i, B_j] = 0$ for $i \neq j$. Our final assumption is that each supplier has reservation utility (in expectation) $\underline{U}_i$ which we normalize to zero, $\underline{U}_i = 0$. The sequence of events (which is standard for moral hazard problems) is as follows: (1) the prime offers the suppliers take-it-or-leave-it contracts, (2) the suppliers take cost reduction measures and set the base stock levels of their spares inventory, (3) costs and backorders are realized, and (4) suppliers are compensated according to the contract terms.

3.2 First-Best Solution: Complete Observability of Suppliers' Actions

In this section we analyze the problem under the assumption that suppliers' actions $\{a_i, s_i\}$ are both observable and contractible, a situation often referred to as the first-best solution because the prime avoids incentive problems by dictating $\{a_i, s_i\}$. This is the benchmark case against which we can evaluate the efficiency of other contracts. The prime's problem is

$$(\mathcal{A}_{FB}) \quad \max_{\{w_i, \alpha_i, v_i, a_i, s_i\}} \quad E\left[U_0\left(-\sum_{i=1}^n T_i(C_i, B_i)\right) \mid \{a_i, s_i\}\right],$$

$$\text{s.t.} \quad \sum_{i=1}^n E\left[B_i \mid s_i\right] \leq \widehat{B}_0, \quad (\text{AR})$$

$$E\left[U_i\left(T_i(C_i, B_i) - C_i - \psi_i(a_i)\right) \mid a_i, s_i\right] \geq 0. \quad (\text{IR}_i)$$

The expected utility expressions are given by (4) and (5). (AR) is the system availability requirement constraint expressed in terms of backorders, and (IR$_i$) is the individual rationality constraint that ensures supplier i's participation. This program can be solved in two steps. First we determine the optimal w_i for each choice of (a_i, s_i) satisfying (IR$_i$), thus making w_i a function of those variables, $w_i(a_i, s_i)$. As is typical in such problems, the prime sets fixed payments $\{w_i\}$ in order to extract all of the surplus from the suppliers. Then we optimize over all (a_i, s_i) that satisfy (AR). The following proposition specifies the first-best solution.

Proposition 1 *When the suppliers' decisions are observable and contractible, the optimal contract specifies the following supplier decisions (a_i, s_i):*

$$a_i = 1/k_i, \quad (6)$$

$$s_i(\theta) = F_i^{-1}\left(\max\left(1 - c_i/\theta, 0\right)\right), \tag{7}$$

$$\sum_{i=1}^{n} E\left[B_i \mid s_i(\theta)\right] = \widehat{B}_0. \tag{8}$$

The solution $\{a_i^{FB}\}$, θ^{FB} and $\{s_i^{FB}\} = \{s_i(\theta^{FB})\}$ is unique and is obtained by offering a non-performance-based, risk-sharing contract such that $v_i = 0$ and

$$\alpha_i^{FB} = r_i/\left(r_0 + r_i\right). \tag{9}$$

Supplier i's expected utility is zero, whereas the prime's expected utility is $\sum_{i=0}^{n}\left(-c_i s_i^{FB} + \frac{1}{2k_i} - \frac{1}{2}\frac{r_0 r_i Var[\varepsilon_i]}{r_0 + r_i}\right)$.

We note that $\{s_i^{FB}\}$ and θ^{FB} are determined simultaneously from equations (7) and (8). They can be found using a greedy algorithm similar to the one used in calculating optimal inventory stocking levels for each part in classical service parts problems (Sherbrooke [24]). The optimal risk-sharing rule (9) is a modified version of the Borch rule (see Bolton and Dewatripont [3]). It is useful to consider extreme cases. If $r_i = 0$, i.e., if supplier i is risk-neutral, $\alpha_i = 0$, corresponding to the FP contract; since the prime is risk-averse whereas the supplier is not, the prime transfers all risks to the supplier. At the opposite end, consider $r_0 = 0$, i.e., the prime is risk-neutral. In this case $\alpha_i = 1$, meaning that the C+ contract is used. Although it may sound counterintuitive that the C+ contract achieves the first-best solution, we should recall that incentives are not an issue in the current setting because the suppliers' actions are observable and contractible. The role of the C+ contract is merely to mitigate the suppliers' reluctance to participate in the support relationship (the IR constraint), which requires an extra payment by the prime. When both r_0 and r_i are positive, the prime and the supplier i share the cost-related risk according to (9), i.e., based on the value of the supplier's risk aversion relative to that of the prime.

We now focus on the prime's expected utility in which there are three terms for each supplier. The first term $(-c_i s_i^{FB})$ is the cost of s_i^{FB} units in the supplier's inventory. The second term $1/2k_i$ is the net savings due to the supplier's cost reduction efforts. The last term $\frac{1}{2}\frac{r_0 r_i}{r_0 + r_i}Var[\varepsilon_i]$ can be interpreted as the joint risk premium between supplier i and the prime, which is positive only if they are both risk-averse. If $r_0 > 0$ and $r_i = 0$, the prime can protect herself perfectly from the cost-related risk by offering an FP contract $(\alpha = 0)$, and the supplier, who is risk neutral, absorbs all risks. If, on the other hand, $r_0 = 0$ and $r_i > 0$, a C+ contract is used to facilitate each supplier's participation while the risk-neutral prime absorbs all risks. When both parties are risk-averse, there is a trade-off between the prime's desire to protect herself (represented by the term $r_0 \alpha_i^2 Var[\varepsilon_i]/2$ in (5)) by decreasing α_i and reducing each supplier's risk premium $(r_i(1 - \alpha_i)^2 Var[\varepsilon_i]/2$ in (4)) by increasing α_i, resulting in an inefficiency. The importance of risk allocation among the prime and the suppliers in our model

14

is consistent with observations in the related literature that "in recent years, defense contracting has come to be seen as a problem of optimal risk sharing" (see Cummins [10]).

Unlike cost-related risk, performance risk poses no trade-off between the prime and the suppliers: it can be taken away by setting $v_i = 0$. In other words, all parties mutually benefit without the performance clause in the first-best contract; if $v_i > 0$, a risk-averse supplier demands a premium due to the possible penalty associated with the stochastic realization of backorders, so a risk-averse prime faces income fluctuations. Both concerns disappear when $v_i = 0$ without incurring extra cost because the observability of the suppliers' actions $\{s_i\}$ implies that the actions can be perfectly enforced even without performance incentives. Thus, the prime's attitudes toward cost and performance uncertainties are different. This key observation will continue to hold even when the suppliers' actions are unobservable.

3.3 Private Actions: The Suppliers' Problem

We now turn to the situation in which suppliers' actions are unobservable to the prime – which is to be expected in a PBL environment. Given the contract parameters (w_i, α_i, v_i), supplier i chooses (a_i, s_i) that maximize his expected utility (4). That is, he solves

$$\max_{a_i, s_i} w_i - (1 - \alpha_i)(c_i s_i - a_i) - v_i E\left[B_i \mid s_i\right] - k_i a_i^2/2 - r_i(1 - \alpha_i)^2 Var[\varepsilon_i]/2 - r_i v_i^2 Var[B_i \mid s_i]/2.$$

A distinctive feature of this problem is that $Var[B_i \mid s_i]$ is a function of the decision variable s_i. This is a departure from the common assumption found of most moral hazard models in economics that only the mean of the performance measure is affected by the decision variable. In our model the dependence of $Var[B_i \mid s_i]$ on s_i is unavoidable. As will become clear, this feature complicates the analysis significantly and at the same time creates new dynamics. It turns out that the supplier's problem is generally not quasiconcave in s_i, but unimodality can be guaranteed under a mild parametric assumption.

Proposition 2 *Suppose $\alpha_i < 1$ and $v_i[1 - F(0)] \geq (1 - \alpha_i)c_i$. In this scenario there is a unique interior solution to the supplier's problem in which supplier i chooses optimal a_i^* and s_i^* such that*

$$a_i^* = (1 - \alpha_i)/k_i, \tag{10}$$

$$v_i[1 - F_i(s_i^*)] + r_i v_i^2 F_i(s_i^*) E[B_i \mid s_i^*] = (1 - \alpha_i)c_i. \tag{11}$$

Since α_i and v_i are determined by the prime, the condition in Proposition 2 has to be checked against the optimal solution. We have verified through numerical examples that the condition is mild in the sense that it is violated only under extreme parameter settings (e.g., when the prime's risk aversion measure r_0 is orders of magnitude greater than that of the supplier, r_i). We henceforth assume

that the condition is always satisfied. From the Proposition we obtain the following result, which offers an intuition into the impact of contract parameters on optimal decisions.

Corollary 1 *Suppose conditions in Proposition 2 hold. Then*

(i) $\partial s_i^*/\partial r_i > 0$, $\partial a_i^*/\partial r_i = 0$.

(ii) $\partial s_i^*/\partial \alpha_i > 0$, $\partial a_i^*/\partial \alpha_i < 0$.

(iii) $\partial s_i^*/\partial v_i > 0$, $\partial a_i^*/\partial v_i = 0$.

From (i) we see that the more risk-averse the supplier, the greater the optimal inventory position he chooses. By investing in more spares, the supplier cuts down not only the number of expected backorders but also the likelihood of backorders (in short, he increases the fill rate), reducing the variance associated with backorders. Hence, a risk-averse supplier is inclined to increase s_i to protect himself from performance uncertainty.[7] To put it another way, there exists a preventive measure by the supplier to avoid the risk of backorders (increase s_i) but not the cost risk, because the optimal cost reduction effort is unaffected by the degree of risk aversion, which is apparent from (10).[8]

Parts (ii) and (iii) in Corollary 1 explain optimal supplier responses to the contract terms α_i and v_i which have some intuitive properties. If the prime increases the reimbursement ratio α_i, the supplier becomes less concerned with cost overruns and hence does not exert as much cost reduction effort as he might otherwise ($\partial a_i^*/\partial \alpha_i < 0$). At the same time, his perceived *effective* unit cost of inventory ($(1-\alpha_i)c_i$ on the right-hand side of (11)) decreases, making it desirable to stock more inventory. With respect to the backorder penalty v_i, the larger v_i means a stronger incentive to decrease backorders so that s_i^* increases. However, the performance penalty does not affect a_i^*, as it serves only as an incentive to reduce backorders, not costs. It is important to note that this predicted behavior is, in part, a consequence of our model assumptions in which supplier effort affects only cost and that cost uncertainties and product performance (reliability) risk are unrelated (i.e., independent).

3.4 Private Actions: The Prime's Problem

Anticipating that the suppliers will respond by choosing $\{a_i, s_i\}$ according to (10) and (11), the prime selects contract terms $\{w_i, \alpha_i, v_i\}$ that achieve minimal total disutility subject to the backorder constraint. With the right incentives, each supplier will voluntarily choose (a_i, s_i) that match the prime's expectation, even though there is no way to verify the suppliers' decisions directly. This voluntary

---

[7]This result is the opposite of the conclusion in the risk-averse newsvendor model (Chen and Federgruen [6]), which predicts that the optimal inventory level decreases with the degree of risk aversion.

[8]This result is due to the assumption that the stochastic term ε_i enters additively into the supplier's total cost $C_i = c_i s_i - a_i + \varepsilon_i$; the effort reduces the mean of C_i but not the variance. Under this standard assumption the supplier has no control over the variability of cost, so his attitude toward risk does not factor into the decision about a_i^*.

action is expressed in terms of incentive compatibility (IC) constraints, which are added to the prime's problem formulation as follows.

$$(\mathcal{A}_{SB}) \quad \max_{\{w_i, \alpha_i, v_i\}} \quad E\left[U_0\left(-\sum_{i=1}^n T_i(C_i, B_i)\right) \mid \{a_i^*, s_i^*\}\right],$$

$$\text{s.t.} \quad \sum_{i=1}^n E\left[B_i \mid s_i^*\right] \leq \widehat{B}_0, \tag{AR}$$

$$E\left[U_i\left(T_i(C_i, B_i) - C_i - \psi_i(a_i)\right) \mid a_i^*, s_i^*\right] \geq 0, \tag{IR$_i$}$$

$$(a_i^*, s_i^*) \in \arg\max E\left[U_i\left(T_i(C_i, B_i) - C_i - \psi_i(a_i)\right) \mid a_i, s_i\right]. \tag{IC$_i$}$$

Similar to the first-best case, it can be demonstrated that (IR$_i$) constraints bind at the equilibrium so that we can simplify the problem by solving for a value of w_i that leaves suppliers with zero profits. Using the Lagrange multiplier θ for the backorder constraint, we can write n individual Lagrangian functions. Moreover, it is convenient to convert the Lagrangian into a function of (α_i, s_i, θ) rather than a function of (α_i, v_i, θ) using the monotonicity result $\partial s_i^*/\partial v_i > 0$ from Corollary 1. Using (10), we obtain

$$\mathcal{L}_i(\alpha_i, s_i, \theta) = c_i s_i + \theta E\left[B_i \mid s_i\right] - (1-\alpha_i)/k_i + (1-\alpha_i)^2/(2k_i) + \left(r_0 \alpha_i^2 + r_i(1-\alpha_i)^2\right) Var[\varepsilon_i]/2$$

$$+ (r_0 + r_i)\left[v_i(\alpha_i, s_i)\right]^2 Var[B_i \mid s_i]/2, \tag{12}$$

whereby

$$v_i(\alpha_i, s_i) = \begin{cases} \frac{(1-\alpha_i)c_i}{1-F_i(s_i)} & \text{if} \quad r_i = 0, \\ \frac{1-F_i(s_i)}{2r_i F_i(s_i) E[B_i \mid s_i]}\left(-1 + \sqrt{1 + \frac{4r_i c_i(1-\alpha_i)F_i(s_i)E[B_i \mid s_i]}{[1-F_i(s_i)]^2}}\right) & \text{if} \quad r_i > 0, \end{cases} \tag{13}$$

from (11). We readily notice that the optimal performance incentive $v_i(\alpha_i, s_i)$ is a decreasing function of α_i; in order to have the supplier choose s_i, the prime may decrease v_i while increasing α_i, or vice versa. Thus, v_i, the incentive to increase the stocking level, and $1 - \alpha_i$, the incentive to reduce costs, are complements. This observation plays a key role in a later analysis and will be discussed further.

We denote the optimal solution pairs with superscripts SB, $\{\alpha_i^{SB}, s_i^{SB}\}$. Unfortunately, (12) is not generally quasiconvex and hence is not necessarily unimodal. The analytical specification of s_i^{SB} is intractable even with α_i fixed, requiring numerical analysis. To circumvent this difficulty and gain additional insights, in the next section we focus on several special cases and later analyze the original problem numerically.

Contract type	No performance-based compensation ($v = 0$)	Performance-based compensation ($v > 0$)
Pure performance ($\alpha = w = 0$)		The customer is unable to extract all supplier surplus.
Fixed price ($\alpha = 0$)	While achieving the first-best cost reduction effort a^{FB}, the supplier is incentivized to reduce s as much as possible.	First-best can be achieved with the appropriate choice of w and v under risk neutrality. Second-best is not achieved under risk aversion ($\alpha > 0$ in general).
Cost plus ($\alpha = 1$)	The supplier exerts zero cost reduction effort ($a = 0$) and is indifferent toward s.	The supplier exerts zero cost reduction effort ($a = 0$) and tries to increase s as much as possible.

Table 1: Incentive effects of various contract combinations in the presence of performance constraint.

3.5 Cost Plus (C+) vs. Fixed Price (FP) vs. Performance Contracts

Before delving into the analysis of optimal contracts, we pause here to evaluate the effectiveness of the most widely used contract forms, C+ ($\alpha_i = 1$, $v_i = 0$) and FP ($\alpha_i = v_i = 0$), and compare them with the performance contracts ($v_i > 0$). Let us first consider the traditional cost reimbursement contracts, C+ and FP. Consistent with other literature analyzing and comparing these contracts (see Scherer [22]), our model indicates that they are polar opposites when it comes to providing cost reduction incentives. With the FP contract a supplier becomes the residual claimant and hence it is in his interest to reduce costs as much as possible. In terms of the risk, the FP contract gives perfect insurance to the prime because the supplier bears all cost-related risks (i.e., cost under- or overruns). In contrast, the C+ contract shifts all risks to the prime, as she has to reimburse whatever the realized cost may be. At the same time, the C+ contract provides no incentive for the supplier to reduce costs.[9]

Despite the prevalence of C+ and FP contracts in practice, they do not induce the desired supplier behavior when a performance constraint has to be taken into account and the prime cannot observe suppliers' actions. This becomes clear after inspecting the supplier's utility function (4). With the FP contract, it is in the supplier's interest to reduce not only the effort a_i but also the inventory s_i as much as possible, thus violating the minimum availability desired by the prime. A C+ contract, on the other hand, has the effect of making the supplier indifferent to the choice of s_i. Clearly, inducing proper actions requires performance incentives. The simplest contract in this category (the "pure performance contract") has $\alpha_i = w_i = 0$ and $v_i > 0$. Indeed, such a contract can induce the supplier to choose the optimal inventory level s_i but the prime is unable to extract profits. Interestingly (to be demonstrated in the following section) a contract with $w_i > 0$ and $v_i > 0$ can achieve the first-best solution, but only

---

[9]For this reason, contracts with $0 \leq \alpha_i < 1$ are often called incentive contracts (see Scherer [22]), a term that refers to incentives to reduce costs rather than performance incentives.

if all parties are risk-neutral, because proper risk sharing requires $\alpha_i > 0$. Thus, the optimal contract will have all three components: a fixed payment, a cost-sharing clause, and a performance incentive. Table 1 summarizes supplier behavior under all of these contract combinations.

4 Analysis

4.1 Risk-Neutral Firms

Many difficulties associated with the analysis disappear if all suppliers and the prime are risk-neutral, which may be the case in practice if the prime and the suppliers are all very large, well-diversified corporations. In this case, as we show below, even when actions are unobservable, the first-best solution is achieved with a contract that is a simple combination of FP and performance-based components. This solution highlights the performance allocation aspect of our problem at the expense of ignoring the issue of risk-sharing.

Proposition 3 *With $r_0 = r_1 = ... = r_n = 0$, the first-best solution is achieved if and only if*

(i) $\alpha_1 = \alpha_2 = ... = \alpha_n = 0$,

(ii) $w_i = c_i s_i^{FB} + \theta^{FB} E[B_i \,|\, s_i^{FB}] - 1/2k_i$, and

(iii) $v_1 = v_2 = ... = v_n = \theta^{FB}$.

The supplier i's expected utility is zero while the prime's expected utility is $\sum_{i=0}^{n} \left(-c_i s_i^{FB} + 1/2k_i \right)$.

The preceding result is not entirely new: it is often the case in other principal-agent models that the first-best solution is achieved with an FP/performance contract between two risk-neutral firms when there is only one effort variable (for example, see Bolton and Dewatripont [3]). It turns out that having two effort variables a_i and s_i as well as multiple suppliers leads to the same result. To see why, note that the prime can (1) choose α_i to induce $a_i^* = a_i^{FB}$ because the supplier's response a_i^* is a function of α_i only and (2) given this α_i, choose v_i to induce $s_i^* = s_i^{FB}$ without incurring any inefficiencies associated with risk sharing. Thus, α_i and v_i under risk neutrality serve only as incentives and not as instruments for providing insurance/sharing risk.

There is, however, a major deviation from the classical analysis involving just one supplier. It is captured in part (iii), which can be interpreted to mean that every backorder from heterogeneous subsystems has equal importance regardless of the subsystem unit price c_i so performance incentives are equal across suppliers. In our additively separable backorder model $(B_0 = \sum_{i=0}^{n} B_i)$ this makes intuitive sense, because the prime does not discriminate between a backorder of a \$1,000 item and that of a \$10 item: i.e., each item contributes equally to the downtime of the system. However, it would be erroneous to conclude that item unit cost $\{c_i\}$ has no effect on determining the backorder incentive

θ^{FB} because it determines θ^{FB} indirectly through the (AR) constraint. The fact that penalty rates are linked across suppliers continues to hold in the risk-averse case, although the equality as in (iii) can no longer be sustained because of suppliers' varying attitudes toward risk. The policy implication of this result is to treat all suppliers equally with respect to the performance incentive.

4.2 Risk-Averse Firms: Cases with Partial Observability

As the next step in gaining insights we now analyze the problem under a simplifying assumption that either $\{s_i\}$ or $\{a_i\}$ are observable and contractible, but not both. As will become evident shortly, these special cases serve as bounds on the optimal contract parameters under conditions of complete unobservability and hence are useful in understanding the structure of the problem. We shall first consider the case when $\{s_i\}$ are observable but $\{a_i\}$ are not. This may happen if suppliers utilize consignment inventory management for all subsystems (which is sometimes the case in practice) so that inventories are visible to the prime. As s_i can now be dictated by the prime, there is no need to the provide the performance incentive v_i, i.e., the optimal contract has $v_i = 0$ for all i. The prime's problem $(\mathcal{A}'_{SB})$ then becomes

$$(\mathcal{A}'_{SO}) \quad \min_{\{\alpha_i, s_i\}} \quad \sum_{i=1}^{n} \left(c_i s_i - (1 - \alpha_i)/k_i + (1 - \alpha_i)^2/(2k_i) + \left(r_0 \alpha_i^2 + r_i (1 - \alpha_i)^2 \right) Var[\varepsilon_i]/2 \right),$$
$$\text{s.t.} \quad \sum_{i=1}^{n} E\left[B_i \,|\, s_i \right] \leq \widehat{B}_0.$$

The optimal contract (denoted by the superscript SO) is as follows.

Proposition 4 *When $\{s_i\}$ of all suppliers are observable to the prime but $\{a_i\}$ are not, it is optimal to specify the contract terms according to*
   (i) $\alpha_i^{SO} = k_i r_i / \left(1/Var[\varepsilon_i] + k_i(r_0 + r_i) \right) < \alpha_i^{FB}$,
   (ii) $w_i^{SO} = (1 - \alpha_i^{SO}) c_i s_i^{FB} - (1 - \alpha_i^{SO})^2/(2k_i) + r_i(1 - \alpha_i^{SO})^2 Var[\varepsilon_i]/2$, and
   (iii) $v_i^{SO} = 0$.
$s_i^{SO} = s_i^{FB}$ is imposed on supplier i while the contract terms induce the cost reduction effort $a_i^{SO} = (1 + k_i r_0 Var[\varepsilon_i]) / \left(k_i + k_i^2(r_0 + r_i) Var[\varepsilon_i] \right)$.

Even though one of the supplier's actions is observable to the prime, we see that the first-best solution cannot be achieved and hence there are inefficiencies due to incentive issues. Namely, there is less cost sharing than is first-best optimal because the prime has to give more incentive to reduce costs than would have been the case after offering α_i^{FB}, so the prime transfers more risk to suppliers by cutting down the cost reimbursement α_i. We see that α_i^{SO} exhibits intuitive properties: as $Var[\varepsilon_i]$ approaches infinity, α_i^{SO} increases asymptotically to the first-best optimal risk sharing ratio α_i^{FB} because the

supplier's effort a_i is buried in huge uncertainty and there is no need to provide an extra cost reduction incentive above the level required by the first-best solution. It is also clear that α_i^{SO} moves toward zero (toward an FP contract) as $Var[\varepsilon_i]$ decreases. The relative risk aversion ratio r_0/r_i is another major determinant of α_i^{SO} which is similar to the first-best case: if the ratio is small, α_i^{SO} is on the C+ side (closer to 1) whereas a large ratio implies that α_i^{SO} is on the FP side (closer to 0). In other words, the more risk-averse a firm is compared to its counterpart (on a relative scale that depends on the parameter values), the more it is protected from risk.

We note that, quite interestingly, α_i^{SO} can be derived from an alternative assumption that (a_i, s_i) are both unobservable but $Var[\varepsilon_i] \gg Var[B_i \,|\, s_i]$. In this scenario performance uncertainty is negligible compared to cost uncertainty, so the focus of incentives is on driving down costs, not on improving performance. The situation is different with regard to $\{v_i\}$, however, since the prime needs to utilize the performance penalty rate $(v_i > 0)$ in order to have the suppliers choose the desired inventory positions $\{s_i\}$.

The other possibility is when $\{a_i\}$ of all suppliers are observable but $\{s_i\}$ are not. We denote the optimal solution in this case with the superscript AO. The prime's problem becomes

$$(\mathcal{A}'_{AO}) \quad \min_{\{a_i, \alpha_i, s_i\}} \quad \sum_{i=1}^{n} \left(c_i s_i - a_i + k_i a_i^2/2 + \left(r_0 \alpha_i^2 + r_i (1 - \alpha_i)^2 \right) Var[\varepsilon_i]/2 + (r_0 + r_i) \left[v_i(\alpha_i, s_i) \right]^2 Var[B_i \,|\, s_i]/2 \right)$$

$$\text{s.t.} \quad \sum_{i=1}^{n} E\left[B_i \,|\, s_i \right] \leq \widehat{B}_0$$

We note that the link between a_i and α_i is decoupled but that the link between (α_i, s_i) and v_i remains. It is clear that $a_i^{AO} = a_i^{FB}$ as in (6) but that tractable expressions for α_i^{AO} and s_i^{AO} do not exist. Despite this shortcoming, α_i^{AO} can be evaluated analytically in the special case with only one supplier, a scenario which we present next.

4.3 Single Risk-Averse Supplier

In this subsection we assume that there is only one supplier, so we drop the subscript i. Not only is such a firm-to-firm setting consistent with a majority of supply chain contracting models in the literature, but it is also one of the commonly observed arrangements found in PBL practice. For example, a setting in which manufacturing of a single key component is outsourced or one where a military customer contracts directly with a subsystem supplier fits this description (e.g., the Navy's PBL contract with Michelin for tires or commercial airline "power by the hour" contracts with engine manufacturers like GE and Rolls Royce). In addition, there are instances in which an intermediary acts as a wholesaler of a subsystem. If the intermediary has a performance contract with the subsystem provider, he becomes the "prime" in our model. As we will shortly see through numerical experiments,

insights from this simpler model continue to hold for the general assembly structure with multiple suppliers.

With a single supplier, it seems natural for the prime to set incentives in a way that $E\left[B \mid s^{SB}\right] = \widehat{B}_0$ holds. In particular, this would be the case if the prime's disutility was increasing monotonically in s, which is an intuitive property. Unfortunately, this intuition is not entirely correct. As noted in the previous section, the analysis of risk-averse firms is complicated by the non-quasiconvexity of the performance risk premium term $(r_0 + r)\,v(\alpha, s)^2 Var[B \mid s]$. For a fixed α, numerical plotting shows that this term may exhibit *quasiconcavity* in s (as opposed to the desired quasiconvexity), implying that the Lagrangian (12) can be bimodal. Thus, the prime may prefer to have more inventory than follows from $E\left[B \mid s^{SB}\right] = \widehat{B}_0$. This, however, happens only in extreme cases when the prime is several orders of magnitude more risk-averse than the supplier and therefore wants to protect herself from performance risk with a very large inventory. In most of our numerical examples with a wide range of parameter combinations the prime's objective function was, indeed, increasing monotonically in s. Therefore, we will henceforth assume that the problem parameters are such that the backorder constraint is binding, which effectively requires the optimal inventory position s^{SB} to satisfy $E\left[B \mid s^{SB}\right] = \widehat{B}_0$. Given that v is completely determined by α and s according to (13), the only variable to be determined is the cost-sharing parameter α so that our problem is simplified to a one-dimensional optimization.

Lemma 1 *The prime's Lagrangian (12) is convex in α when s is fixed.*

It follows that there is a unique α^{SB} that minimizes the prime's disutility. There exists a closed-form solution, but it is quite complex (the first order condition for α is a cubic equation; see proof in the Appendix), and inspection alone does not provide ready insights. Instead, we employ implicit differentiation to gain a better understanding. Namely, we focus on understanding how the parameters of the contract change when cost uncertainty $Var[\varepsilon]$ changes. There are several motivations behind this analysis. First, cost uncertainty is of primary importance in practice because it is often harder to estimate than performance uncertainty. Second, there are significant changes in cost uncertainty over the product life cycle (while performance uncertainty is relatively more stable) and therefore there is a need to understand how contractual terms would change in response. Finally, as will be seen shortly, by varying the cost uncertainty we are able to obtain insights that sometimes differ fundamentally from insights in the classical literature on moral hazard problems with multitasking.

Proposition 5 *Suppose $r_0, r > 0$ and that s^{SB} is fixed by the backorder constraint $E\left[B \mid s^{SB}\right] = \widehat{B}_0$. Then $\alpha^{SO} < \alpha^{SB} < \alpha^{AO}$ and $v^{SB} > v^{AO} > v^{SO} = 0$. Further, let $\widetilde{\ell}(r_0, r) = \partial\mathcal{L}/\partial\alpha|_{\alpha=\alpha^{FB}}$ where $\mathcal{L}$ is the prime's Lagrangian defined in (12). Function $\widetilde{\ell}(r_0, r)$ increases in the ratio r/r_0 and crosses zero*
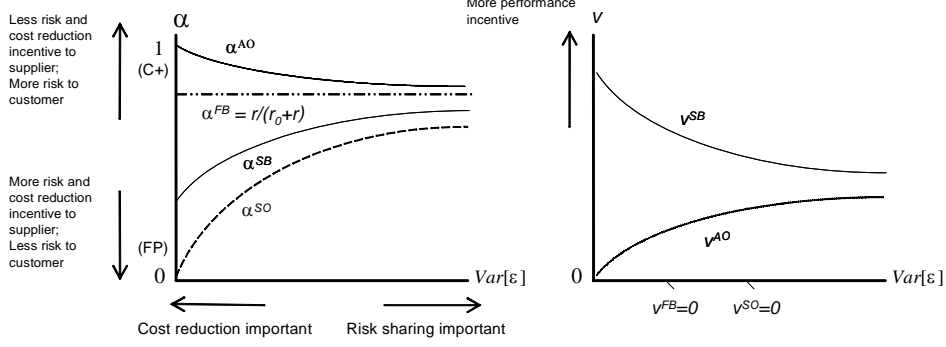
Figure 2: $\widetilde{\ell}(r_0, r) > 0$, the supplier is relatively more risk-averse than the prime.

exactly once. The optimal contract parameters α^{SB} and v^{SB} are related to α^{FB} and v^{FB} as follows.

(i) *If $\widetilde{\ell}(r_0, r) > 0$, $\alpha^{SB} < \alpha^{FB}$, $d\alpha^{SB}/d(Var[\varepsilon]) > 0$, and $dv^{SB}/d(Var[\varepsilon]) < 0$.*

(ii) *If $\widetilde{\ell}(r_0, r) = 0$, $\alpha^{SB} = \alpha^{FB}$, $v^{SB} = v^{FB}$, and $d\alpha^{SB}/d(Var[\varepsilon]) = dv^{SB}/d(Var[\varepsilon]) = 0$.*

(iii) *If $\widetilde{\ell}(r_0, r) < 0$, $\alpha^{SB} > \alpha^{FB}$, $d\alpha^{SB}/d(Var[\varepsilon]) < 0$, and $dv^{SB}/d(Var[\varepsilon]) > 0$.*

First, we note that the optimal cost sharing ratio α^{SB} is bounded above by α^{AO}, the optimal ratio when the cost reduction effort a is observable. In the current case the effort is not observable and therefore the prime has to reduce α to provide more incentives to reduce costs. The side effect is that the supplier's effective unit cost $(1-\alpha)c$ increases, thus requiring a higher performance incentive v to achieve the desired inventory position. Therefore, $v^{SB} > v^{AO}$. Second, we note that α^{SB} is bounded below by α^{SO}, which we derived by assuming that the inventory position s is observable. When inventory is not observable, the prime needs to provide a better performance incentive, $v^{SB} > v^{SO} = 0$, but doing so exposes both the prime and the supplier to performance risk thus creating inefficiency that can be mitigated by increasing α. Higher α reduces the effective unit cost $(1-\alpha)c$ for the supplier and allows him to achieve the inventory position s^{SB} with a smaller v. Hence, increasing α above α^{SO} achieves the optimal solution.

A comparison of the second-best solution with the first-best solution is more complex. It is instrumental to consider two cases based on the relative risk aversion of the prime and the supplier separately. Since function $\widetilde{\ell}(r_0, r)$ increases in the ratio r/r_0 and crosses zero exactly once, the condition $\widetilde{\ell}(r_0, r) > 0$ in (i) can be interpreted as $r \succ r_0$, where the symbol "$\succ$" means that the supplier is relatively more risk-averse than the prime. Similarly, $\widetilde{\ell}(r_0, r) < 0$ can be interpreted as $r \prec r_0$, whereby the prime is relatively more risk-averse than the supplier. We first consider the former situation (which may arise if the prime is a bigger and more diversified company than the supplier). We believe that this case is more natural in practice. Figure 2 illustrates the results in (i). We make the following observations from these figures. First, $\alpha^{SB} < \alpha^{FB}$, and the unobservability of effort and inventory results

in less cost reimbursement than under the first-best solution. Second, α^{SB} increases with $Var[\varepsilon]$ and asymptotically approaches α^{FB}. With a large cost uncertainty, the risk-averse supplier is reluctant to participate in the trade, so the prime has to provide insurance by reimbursing a large proportion of the supplier's costs. Thus the supplier has less incentive to make efforts to reduce costs. On the other hand, when $Var[\varepsilon]$ is small, providing cost-reduction incentives becomes more important. Third, the gap between α^{SB} and α^{SO} decreases in $Var[\varepsilon]$. This gap can be interpreted as the additional inefficiency due to performance risk. When cost uncertainty is large, the performance uncertainty $Var[B \,|\, s^{SB}]$ is negligible and the gap between SB and SO disappears. The gap between α^{SB} and α^{AO} is interpreted similarly. Finally, v^{SB} decreases with $Var[\varepsilon]$, asymptotically approaching $v(\alpha^{FB}, s^{FB})$. With higher cost uncertainty, the performance incentive is lowered.

Overall, we observe that α^{SB} and v^{SB} move in the opposite directions as $Var[\varepsilon]$ increases because the prime increases α to mitigate the supplier's risk (we recall that the supplier is more risk-averse than the prime in the current setting), and as a result, the supplier's effective unit cost $(1 - \alpha)c$ is reduced, making it less expensive to stock inventory and allowing for a smaller incentive v. Therefore increasing $1 - \alpha$ has the same effect on inventory as increasing v; these two incentives are *complements* with respect to s. This conclusion is similar to the one presented in Holmström and Milgrom's [13] original multitask principal-agent model in which increasing variability in one output leads to weaker incentives for all outputs. Yet, the mechanism by which we arrive at our conclusion is different. Specifically, in Holmström and Milgrom [13], raising one effort raises the marginal disutility of raising another effort, which is not the case in our model (whereby the supplier's disutilities $(1 - \alpha)cs$ and $ka^2/2$ are independent of each other). Holmström and Milgrom show that, if an agent has a strong incentive to perform one task because the result of the other task is difficult to measure (and hence there is a little incentive to perform it), the agent's attention is disproportionately directed toward the former task, since he finds it more costly to exert both efforts. In this case the best course of action for the prime is to reduce the incentive to perform the former task as well. Another important assumption in their model is that the outcomes are affected by exactly one effort each, so there is a one-to-one correspondence between an incentive and an effort. In contrast, our model has an outcome C that is a function of both a and s via $C = cs - a + \varepsilon$. Increasing the cost reimbursement ratio α because of large cost uncertainty $Var[\varepsilon]$ produces conflicting reactions by the supplier with respect to a and s, the former decreasing while the latter increases as described above.

The model closest to ours is found in Bolton and Dewatripont ([3], pp. 223-8) where there is *direct conflict between the tasks,* because exerting one effort positively affects one outcome but negatively affects the other. Despite the similarity in structure, there is no direct correspondence between our results, because Bolton and Dewatripont compare two tasks performed by one agent and two tasks
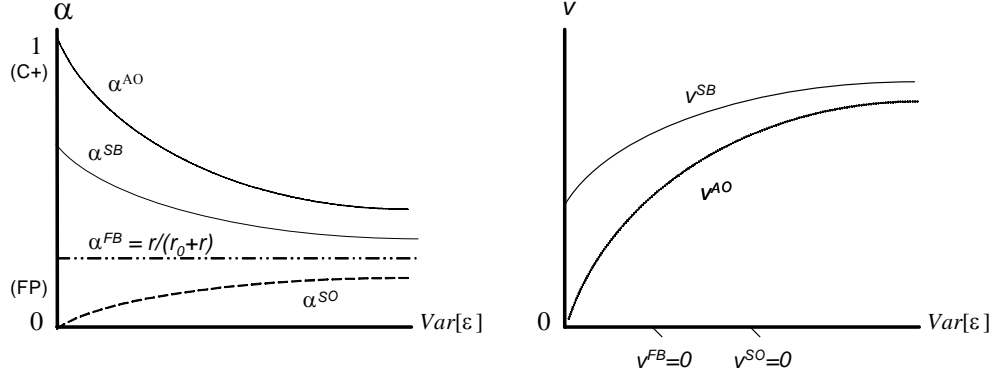
Figure 3: $\widetilde{\ell}(r_0, r) < 0$, the prime is relatively more risk-averse than the supplier.

independently assigned to two agents, a trade-off we do not consider.

Next, we consider the case in which the prime is relatively more risk-averse than the supplier, $r \prec r_0$ (case (iii) in Proposition 5). Figure 3 is an analog of Figure 2. Compared to the previous discussion, α^{SB} and v^{SB} exhibit exactly opposite behavior. Now $\alpha^{SB} > \alpha^{FB}$ and α^{SB} decreases in $Var[\varepsilon]$ while v^{SB} increases in $Var[\varepsilon]$. This fundamental difference arises because, unlike in the previous case where insurance was more important for the supplier, it is now the prime who needs protection from risk. With large cost uncertainty the prime is protected by choosing small α, thereby transferring most of the risk to the supplier. A nonintuitive consequence of this outcome is that the supplier is incentivized more to reduce his cost and increase his stocking level when the cost uncertainty is great. Therefore, the prime's concern for her own risk protection reverses contractual terms and comparative statics. The complementarity between $1 - \alpha^{SB}$ and v^{SB} still remains, however: as $1 - \alpha^{SB}$ increases, so does v^{SB}. We note that results when the prime is more risk-averse than the supplier are somewhat contrary to what we have come to expect from the existing literature on multitasking where the prime is often assumed to be risk-neutral.

4.4   Multiple Risk-Averse Suppliers

In this section we present a numerical analysis of the problem with multiple suppliers. We illustrate our findings via two examples. First, we consider two suppliers that differ by at most one of the parameters $\{r_i, Var[\varepsilon_i]\}$. This example isolates the trade-off between incentives and risk. The second example is based on actual maintenance data from a fleet of military fighter aircraft. This second data set illustrates how our model can be applied in practice to support long-term strategic planning and contract negotiations.

4.4.1 Example 1: Two Symmetric Suppliers

In this example we assume that all parameter values are symmetric across the suppliers except for either $\{r_i\}$ or $\{Var[\varepsilon_i]\}$. Default values are $\mu_i = \sigma_i^2 = 10$, $c_i = 1$, $k_i = 0.2$, $r_1 = 0.1$, $Var[\varepsilon_1] = 10$, and $\widehat{B}_0 = 4$ and a normal distribution of on-order inventory is chosen in keeping with our continuous approximation for the underlying inventory model. We vary supplier 2's risk aversion and cost uncertainty r_2 and $Var[\varepsilon_2]$ and the prime's risk aversion r_0 in order to observe their effects on $(\alpha_i^{SB}, v_i^{SB})$ and (a_i^{SB}, s_i^{SB}). We note that the first-best inventory positions are $s_1^{FB} = s_2^{FB} = 8.725$. Table 4 (see Appendix) summarizes the results of varying r_2 and r_0.

We observe only minimal changes in s_1^{SB} and s_2^{SB} as parameters change, with the greatest change occurring when r_0 is large $((s_2^{SB} - s_1^{SB})/s_1^{FB} = 0.094 = 9.4\%$ distortion when $r_0 = r_2 = 1$, $r_1 = 0.1)$. In contrast, α_2^{SB} changes widely (e.g., with $r_0 = 0.1$, α_2^{SB} increases from 0.184 to 0.726 as r_2 increases from 0.01 to 1). We also see that α_1^{SB} is essentially unaffected by r_2, the risk aversion of the other supplier. We infer from these observations that suppliers' efforts $\{\alpha_i\}$ are more flexible in optimizing contracts because they are not subject to an externality such as the overall backorder constraint, which limits the ranges of $\{s_i^{SB}\}$. We also confirm that $\alpha_2^{SB} > \alpha_2^{FB}$ for a relatively large ratio r_0/r_2, while the opposite is true for a small ratio r_0/r_2, just as predicted by Proposition 5 but for a single supplier. Furthermore, we notice that α_2^{SB} increases monotonically in r_2 for small r_0 $(= 0.01)$, but we do not observe the same monotonicity when r_0 is large $(= 1)$: α_2^{SB} initially decreases from 0.436 to 0.430 but then increases to 0.539. The explanation is as follows: when r_0 is small, increasing α tends to reduce both the cost and performance premiums (see (12) and (13)). However, when r_0 is large, tension exists between the two risk premium terms; although increasing α reduces the performance risk for both the prime and the suppliers, it exposes the prime to the risk of greater cost. These two opposing forces break down the monotonicity.

Next, Table 5 (see Appendix) illustrates the effect of varying $Var[\varepsilon_2]$. Once again we observe that s_2^{SB} is not very sensitive to changes in $Var[\varepsilon_2]$, but that α_2^{SB} is. From the table, we see that α_2^{SB} moves toward α_2^{FB} as $Var[\varepsilon_2]$ increases (regardless of the value of r_0), confirming the prediction from Proposition 5 for the single supplier case. In addition, numbers in the table demonstrate that $d\alpha_2^{SB}/dr_0 > 0$ for small $Var[\varepsilon_2]$ whereas $d\alpha_2^{SB}/dr_0 < 0$ for large $Var[\varepsilon_2]$. (This can be proven analytically in the single supplier case, but we omit the derivation.) In other words, when cost uncertainty is relatively low, the more risk-averse prime moves toward a C+ contract and takes up a larger portion of the cost-related risk, a nonintuitive result. What actually happens is that the performance premium is more important in this situation, outweighing the concern for cost-related risks. Clearly, in the presence of performance risk, the intuition regarding cost sharing is not always straightforward.

Finally, Table 6 (see Appendix) shows how the optimal contract parameters and the suppliers' actions vary as the overall backorder constraint changes. In this example, suppliers 1 and 2 are asymmetric only in their attitude toward risk: $r_1 = 0.1$ and $r_2 = 1$, while $r_0 = 0.5$. As expected, v_1^{SB} and v_2^{SB} decrease as $\widehat{B}_0$ increases, since a less stringent backorder constraint allows for smaller inventories and hence reduces the need for performance incentives. Changes in α_1^{SB} and α_2^{SB} are relatively small. We see that distortion in $\{s_i^{SB}\}$ becomes larger as the constraint is relaxed (measured by the quantity $(s_2^{SB} - s_1^{SB})/s_1^{FB}$, it grows from 2.94% at $\widehat{B}_0 = 1$ to 9.75% at $\widehat{B}_0 = 7$). Intuitively, this happens because the less stringent backorder constraint results in a larger range in which inventories can be adjusted without violating the constraint. However, the magnitude of the distortion is still small, confirming our previous observation that the presence of the backorder constraint limits the prime's contract parameter choices (α and v) such that they induce the base stock levels to be close to the first-best values $\{s_i^{FB}\}$.

4.4.2 Example 2: Actual Data for a Fleet of Military Aircraft

Our second numerical example is based on a real-life maintenance data for a fleet of military fighter aircraft. A total of $N = 156$ aircraft are deployed in the fleet. We obtained data on unit costs, daily failure rates and repair lead times for a representative collection of 50 line replaceable units (LRUs). To utilize our model we aggregate data into five subsystem groups: avionics (a), engines (e), landing gear (l), mechanical (m), and weapons (w), based on descriptions of each part. We employ the following technique to obtain costs, failure rates and lead times for these subsystems. First, we assign each part to one of the groups, and compute the subsystem's mean inventory on-order as $\mu_i = \sum_{j=1}^{n_i} \lambda_j L_j$, where $i = \{$a,e,l,m,w$\}$ and $n_i =$ the number of parts within subsystem i. Thus, we treat each subsystem as a "kit" which is replaced whenever any part within it fails. Since, in practice, only failed items would be replaced in response to a subsystem failure, summing the unit costs of the parts to obtain the unit cost of the subsystem (i.e., $c_i = \sum_{j=1}^{n_i} \lambda_j c_j$) would undoubtedly overestimate the total capital invested in each subsystem. Therefore, we introduce a correction by computing the effective unit cost that is weighted by the demand rate, $c_i = \left(\sum_{j=1}^{n_i} \lambda_j c_j \right) / \sum_{j=1}^{n_i} \lambda_j$. In other words, we give less weight to the value of parts that rarely fail and give more weight to the value of parts that frequently fail. Alternatively we could view the problem where the cost of a failing part kit is equal to the demand-weighted average cost, i.e., weighted over all parts in the kit.

The first row of Table 2 lists computed values of $\{c_i\}$ using our aggregation technique. To verify that this approach is reasonable, we compare the amounts of investment in each subsystem $\left(\sum_{j=1}^{n_i} c_j s_j \right)$ predicted by our model with two benchmarks. The results are shown in Table 2. The first benchmark, which we call the "2-echelon, disaggregated" model, has the solution computed using a proprietary

Subsystem	avionics (a)	engine (e)	landing gear (l)	mechanical (m)	weapons (w)
c_i (in \$1,000)	148.6	9.2	27.5	12.1	87.7
k_i ($\times 10^{-3}$)	0.336	5.435	1.818	4.132	0.570
μ_i	128.20	19.36	13.72	54.61	66.61
Total investment (\$M) with availability target of 95%					
2-echelon, disaggregated	21.461	0.381	1.012	1.437	8.305
1-echelon, disaggregated	19.641	0.295	0.772	0.973	7.592
1-echelon, aggregated	18.872	0.258	0.495	0.811	6.139

Table 2: Choices of contract parameters.

commercial algorithm from MCA Solutions, Inc. that optimizes over multiple echelons and indentures and considers each part-location directly.[10] The second benchmark is called the "1-echelon, disaggregated" model; its solution is computed using a classical greedy algorithm, but each LRU is treated as a separate part. The difference between these two benchmarks is that the former exploits the multi-echelon nature of the model, whereas the latter does not. Finally, we have the "1-echelon, aggregated" model, which uses our aggregation technique. We observe that the aggregated model results in an investment dollar amount that is quite close to those amounts generated by the disaggregated models, especially in terms of the relative investment distribution among the subsystems. We note that the aggregated model underestimates the investments required due to the benefit of the risk pooling associated with treating parts as a kit in one location (the same logic applies when the "1-echelon, disaggregated" model is compared to the "2-echelon, disaggregated" model). This comparison indicates that our aggregation methodology is quite reasonable for the purpose of predicting how a collection of first-tier suppliers would allocate inventory investment and assume risk in response to contract terms proposed by the prime. In particular, our model could be used in the context of a hierarchical solution whereby the model introduced in this paper would predict how risk would be allocated between the prime and the suppliers as well as the relative investments that each supplier would make (in total) to provide the prime with the required system availability, given the terms of a proposed contract. The actual deployments across multiple parts and locations could then be determined in a second phase where commercial software such as MCA's SPO system could be used to optimize the allocation of each supplier's total inventory investment subject to their overall subsystem backorder constraint. Various ownership and control structures could also be accommodated by running the analysis separately for each structural alternative of the service supply chain. This approach is consistent with our observations of the current approach that primes, customers, and subsystem suppliers are taking as they engage in performance-based contract negotiations.

Table 2 summarizes the computed values of $\{k_i\}$, along with $\{c_i\}$ and the subsystem's mean in-

---

[10] http://www.mcasolutions.com/

i	$\sqrt{Var[\varepsilon_i]}/E[\varepsilon_i] = 0.05$					$\sqrt{Var[\varepsilon_i]}/E[\varepsilon_i] = 0.2$				
	a	e	l	m	w	a	e	l	m	w
α_i^{SB}	0.566	0.087	0.407	0.267	0.462	0.630	0.096	0.846	0.293	0.724
v_i^{SB}	65.20	213.61	57.56	134.90	64.04	58.15	209.90	24.31	118.60	41.34
a_i^{SB}	1,285	168	326	201	943	1,099	166	84	171	484
s_i^{SB}	126.31	27.26	18.29	66.98	69.90	126.20	27.24	18.39	66.95	70.05
$\widehat{A}_i$	96.46%	99.96%	99.88%	99.91%	98.80%	96.42%	99.96%	99.88%	99.91%	98.83%
IIR_i	18,769	250	503	810	6,129	18,753	250	505	810	6,143
NCR_i	1,007.46	91.31	229.51	117.63	689.88	895.88	91.16	78.13	110.60	417.76
CRP_i	271.98	0.21	61.04	1.98	169.81	4,252.59	3.35	225.61	25.72	1,795.11
PRP_i	467.17	7.76	17.94	20.96	176.08	375.26	7.58	3.02	16.40	71.33

Table 3: Optimal contract terms and suppliers' actions. The dollar figures are in thousands. IIR stands for investment in resources and is equal to $c_i s_i^{SB}$. NCR is $-a_i^{SB} + \frac{1}{2}k_i(a_i^{SB})^2$, the net cost reduction. CRP is the residual cost risk premium, $\frac{1}{2}(r_0(\alpha_i^{SB})^2 + r_i(1 - \alpha_i^{SB})^2)Var[\varepsilon_i]$, and PRP is the residual performance risk premium, $\frac{1}{2}(r_0 + r_i)(v_i^{SB})^2 Var[B_i|s_i^{SB}]$.

ventory on-order $\{\mu_i\}$ inferred from the data. To determine values of parameters $\{k_i\}$ and $\{Var[\varepsilon_i]\}$, we use the following approach: for each subsystem, we assume that the fixed cost is 100 times higher than the (effective) unit cost c_i and that the maximum dollar amount of cost reduction $a_i^{FB} = 1/k_i$ is 20% of the fixed cost. For the sake of simplicity, we also assume that the coefficient of variation $\sqrt{Var[\varepsilon_i]}/E[\varepsilon_i]$ is the same across suppliers. We infer the risk aversion coefficient for each subsystem provider from the market capitalization of a representative manufacturer of such a subsystem. For example, if Boeing is chosen as the prime and GE as the engine manufacturer, we calculate the risk aversion ratio of $r_0/r_e \simeq 7$ since GE's market capitalization is roughly 7 times that of Boeing (see justification for using company size as a proxy for risk aversion in Cummins [10]). This approach is, of course, quite simplistic, but it fits our aim to illustrate the model (true risk aversion can be estimated empirically). Using this methodology we choose $r_a = 1.79r_0$, $r_e = 0.15r_0$, $r_l = 11.76$, $r_m = r_0$, and $r_w = 3.33$ and we select $r_0 = 0.15$ while the availability target is set at 95%. The optimal contract terms and the suppliers' actions are presented in Table 3.

We consider two scenarios: with small and high cost uncertainty (as captured by the coefficient of variation $\sqrt{Var[\varepsilon_i]}/E[\varepsilon_i]$). As can be seen in Table 3, the avionics system drives the results because it is the most expensive and the subsystem that fails most frequently; due to high unit cost, it is assigned the lowest availability target (96.46% when the coefficient of variation is 0.05) and it also has the largest allocation in terms of the inventory investment. In the case with high cost uncertainty, the cost premium is higher than the performance premium for most suppliers, whereas the performance premium becomes more salient when cost uncertainty is small. Consistent with our results for a single supplier, we observe that $\{\alpha_i^{SB}\}$ increases and $\{v_i^{SB}\}$ decreases with $\sqrt{Var[\varepsilon_i]}/E[\varepsilon_i]$. Finally, we note that the optimal inventories $\{s_i^{SB}\}$ as well as total investment amounts are very close to the values computed by the standard greedy algorithm and values of $\{s_i^{SB}\}$ are quite insensitive to changes in

$Var[\varepsilon_i]$.

5   Conclusion

The goal of this paper is to introduce contracting considerations into the management of after-sales service supply chains. We do so by blending the classical problem of managing the inventory of repairable service parts with a multitask principal-agent model. Furthermore, we use this novel model to analyze incentives provided by three commonly used contracting arrangements, fixed-price, cost-plus and performance-based (FP, C+ and PBL). By doing so, we analyze at least two practically important issues of contracting in service supply chains – performance requirement allocation and risk sharing – when a single customer (either the prime supplier or the end customer) is contracting with a collection of first-tier suppliers of the major subsystems used by an end product/system. When performance is defined as overall system availability, the answer to the former can be found from the solution of the classic multi-echelon, multi-indenture service part resource allocation problem, à la MCA's SPO system. Our innovation is in explicitly modeling decentralized decision making and considering how firms behave when they face uncertainties arising from both support costs and product performance. The notion of risk sharing found in the principal-agent literature is incorporated in our model, providing insights into what types of contract should be used under various operating environments. Specifically, we have discovered that incentive terms in the contract exhibit complementarity, i.e., incentives for both cost reduction and high availability move in the same direction as the operating environment (characterized by cost uncertainty) changes.

Furthermore, our analysis allows us to make normative predictions with respect to how contracts are likely to evolve over the product life cycle. Given our assumption that supplier effort reduces maintenance costs but does not improve product performance reliability or repair capabilities, our model is consistent with the observation that performance uncertainty is relatively stable throughout the sustainment process, whereas cost uncertainty is likely to be reduced over time by learning about costs through the deployment of a larger fleet of systems. Thus, if a series of performance contracts are signed over the product lifetime, our analysis indicates that the cost reimbursement ratio α will decrease (increase) over time if the supplier is relatively more (less) risk-averse than the prime. For the performance incentive v the direction is reversed. It would be interesting to verify this result empirically. From our conversations with executives at a major defense supplier we found that, indeed, typically contracts evolve over time from being more C+-oriented (large α) to being more FP-oriented (smaller α). Since larger, more diversified primes are more common in practice, we believe that suppliers are often more risk-averse in these settings, as independently confirmed by executives we worked with.

Thus, there is encouraging anecdotal evidence that is consistent with our analysis.

We find that, in the presence of great residual uncertainty associated with performance, cost sharing is still an effective tool even if the cost uncertainty is small. That is, the combination FP/performance-based contract is not optimal in such instances (notice the gap between zero and α^{SB} at $Var[\varepsilon] = 0$ in Figure 2), because the cost reimbursement α can be used as a risk protection mechanism even for the risk borne by the performance. Adjusting inventory s for this purpose has limited effect because its primary role is with respect to the availability requirement. Hence, some degree of cost sharing is recommended in the performance contracting environment even when the cost uncertainty is low. Our numerical study shows that the optimal inventory position profile $\{s_i^{SB}\}$ is quite insensitive to changes in risk-related parameters such as r_0, r_i, and $Var[\varepsilon_i]$. This happens because the presence of a stringent backorder constraint limits the range in which $\{s_i^{SB}\}$ can be varied once parameters $\{c_i\}$ and $\{\mu_i\}$ are fixed. Moreover, all analytical results that we obtain for simplified cases continue to hold under more general conditions.

Performance-based contracting in service supply chains offers fertile ground for research where economics and classical inventory theory converge naturally. Not only does it pose theoretically challenging questions but also insights gained from the analysis are of great interest to practitioners who are currently undergoing major business process changes due to the move towards PBL contracting. Our paper analyzes several major issues in performance contracting, but many open questions remain. Follow-up studies may address such topics as the free-riding problem arising from overlapping downtimes across parts; gaming among suppliers and the consequences to realized performance; long-term, strategic product reliability investment vs. intermediate-term, tactical inventory decisions; investment in enhanced repair and logistics capabilities that would reduce lead times; alternative ownership and management scenarios; and many more. We are currently working on some of these issues. Finally, empirical verification of the insights gained from this paper will lead to more effective implementation of contract design and aid contract negotiations.

References

[1] Accenture. 2003. Service Management: Building Profits after the Sale. *Supply Chain Management Perspectives.*

[2] Bajari, P. and S. Tadelis. 2001. Incentives versus transaction costs: a theory of procurement contracts. *RAND Journal of Economics*, 32(3), 387-407.

[3] Bolton, P. and M. Dewatripont. 2005. *Contract Theory.* MIT Press.

[4] Cachon, G. 2003. Supply chain coordination with contracts. *Handbooks in Operations Research and Management Science: Supply Chain Management.* Edited by Steve Graves and Ton de Kok. North Holland.

[5] Chen, F. 2005. Salesforce Incentives, Market Information, and Production/Inventory Planning. *Management Science*, 51(1), 60-75.

[6] Chen, F. and A. Federgruen. 2000. Mean-Variance Analysis of Basic Inventory Models. Working paper, Columbia University.

[7] Cohen, M. A., V. Agrawal, and N. Agrawal. 2006. Achieving Breakthrough Service Delivery Through Dynamic Asset Deployment Strategies. *Interfaces*. Forthcoming.

[8] Cohen, M. A., P. Kamesam, P. Kleindorfer, H. Lee, A. Tekerian. 1990. OPTIMIZER: A Multi-Echelon Inventory System for Service Logistics Management. *Interfaces,* 20(1), 65-82.

[9] Corbett, C. J. 2001. Stochastic Inventory Systems in a Supply Chain with Asymmetric Information: Cycle Stocks, Safety Stocks, and Consignment Stock. *Operations Research*, 49(4), 487-500.

[10] Cummins, J. M. 1977. Incentive Contracting for National Defense: A Problem of Optimal Risk Sharing. *Bell Journal of Economics*, 8, 168-85.

[11] Feeney, G. J. and C. C. Sherbrooke. 1966. The (s-1, s) Inventory Policy under Compound Poisson Demand. *Management Science*, 12(5), 391-411.

[12] Holmström, B. and P. Milgrom. 1987. Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica*, 55, 303-28.

[13] Holmström, B. and P. Milgrom. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization*, 7, 24-52.

[14] Iyer, A., L. B. Schwarz and S. A. Zenios. 2005. A Principal-Agent Model for Product Specification and Production. *Management Science*, 51, 106-119.

[15] Lutze, H. S. and O. Ozer. 2005. Promised Leadtime Contracts and Renegotiation Incentives Under Asymmetric Information. Working Paper, Stanford University.

[16] Macfarlan, W. G. and B. Mansir. 2004. Supporting the Warfighter through Performance-Based Contracting. *Defense Standardization Program Journal*, July/September. http://www.dsp.dla.mil/newsletters/std-news.html.

[17] Markowitz, H. 1952. Portfolio Selection. *Journal of Finance*, 7, 77-91.

[18] Muckstadt, J. A. 2005. *Analysis and Algorithms for Service Parts Supply Chains.* Springer.

[19] Phillips, E. H. 2005. Performance-Based Logistics Takes Center Stage in F-35 Program. *Aviation Week & Space Technology*. October 31, 52.

[20] Plambeck, E. L. and T. A. Taylor. 2004. Partnership in a Dynamic Production System with Unobservable Actions and Noncontractible Output. Working paper, Columbia University.

[21] Plambeck, E. L. and S. A. Zenios. 2000. Performance-Based Incentives in a Dynamic Principal-Agent Model. *Manufacturing and Service Operations Management*, 2(3), 240-63.

[22] Scherer, F. M. 1964. The Theory of Contractual Incentives for Cost Reduction. *The Quarterly Journal of Economics*, 78(2), 257-80.

[23] Sherbrooke, C. C. 1968. METRIC: A Multi-Echelon Technique for Recoverable Item Control. *Operations Research*, 16, 122-41.

[24] Sherbrooke, C. C. 1992. *Optimal Inventory Modeling of Systems: Multi-Echelon Techniques.* John Wiley and Sons.

[25] Su, X. and S. A. Zenios. 2005. Recipient Choice Can Address the Efficiency-Equity Trade-off in Kidney Transplantation: A Mechanism Design Model. Working paper, Stanford University.

[26] Van Mieghem, J. A. 2004. Risk-Averse Newsvendor Networks: Resource Sharing, Substitution, and Operational Hedging. Working paper, Northwestern University.

Appendix

Proof of Proposition 1. We first prove that at the equilibrium, all (IR_i) constraints are binding, i.e., $w_i - (1-\alpha_i)(c_i s_i - a_i) - v_i E\left[B_i\,|\,s_i\right] - k_i a_i^2/2 - r_i(1-\alpha_i)^2 Var[\varepsilon_i]/2 - r_i v_i^2 Var[B_i\,|\,s_i]/2 = 0$ for all i. Suppose otherwise, i.e., that there exists j such that $w_j - (1-\alpha_j)(c_j s_j - a_j) - v_j E\left[B_j\,|\,s_j\right] - k_j a_j^2/2 - r_j(1-\alpha_j)^2 Var[\varepsilon_j]/2 - r_j v_j^2 Var[B_j\,|\,s_j]/2 > 0$. By reducing w_j by ϵ, the prime's utility (5) is increased by ϵ while the (AR) constraint is unaffected. This result allows us to transform $(\mathcal{A}_{FB})$ into

$$(\mathcal{A}'_{FB}) \quad \min_{\{\alpha_i, v_i, a_i, s_i\}} \sum_{i=1}^n \left(c_i s_i - a_i + k_i a_i^2/2 + \left(r_0\alpha_i^2 + r_i(1-\alpha_i)^2\right) Var[\varepsilon_i]/2 + (r_0 + r_i)\, v_i^2 Var[B_i\,|\,s_i]/2\right),$$
$$\text{s.t. } \sum_{i=1}^n E\left[B_i\,|\,s_i\right] \le \widehat{B}_0.$$

Notice that $(\mathcal{A}'_{FB})$ reduces to the classic resource allocation problem faced by a single decision maker (Muckstadt [18]) in the special case $a_i = r_0 = r_i = 0$. Clearly, the objective function is minimized when $v_i = 0$ for all i. With this observation, the Lagrangian with the associated multiplier θ becomes

$$
\begin{aligned}
\mathcal{L}(a, s, \theta) &= \sum_{i=1}^n \left(c_i s_i - a_i + k_i a_i^2/2 + \left(r_0\alpha_i^2 + r_i(1-\alpha_i)^2\right) Var[\varepsilon_i]\right)/2 + \theta\left(\sum_{i=1}^n E\left[B_i\,|\,s_i\right] - \widehat{B}_0\right) \\
&= -\theta\widehat{B}_0 + \sum_{i=1}^n \left(c_i s_i + \theta E\left[B_i\,|\,s_i\right] - a_i + k_i a_i^2/2 + \left(r_0\alpha_i^2 + r_i(1-\alpha_i)^2\right) Var[\varepsilon_i]/2\right). \quad (14)
\end{aligned}
$$

It is apparent that the minimization can be done separately for each supplier. As the objective is a decreasing function of $\{s_i\}$ the optimal values are always at the corner and the (AR) constraint is binding, implying that $\theta > 0$. Note $\partial^2 \mathcal{L}_i/\partial a_i^2 = k_i > 0$, $\partial^2 \mathcal{L}_i/\partial s_i^2 = \theta f(s_i) > 0$, and $\partial^2 \mathcal{L}_i/\partial \alpha_i^2 = r_0 + r \ge 0$. In the absence of cross partial terms $\partial^2 \mathcal{L}_i/\partial a_i s_i = \partial^2 \mathcal{L}_i/\partial s_i \alpha_i = \partial^2 \mathcal{L}_i/\partial \alpha_i a_i = 0$, so the Hessian for supplier i is positive definite and hence the problem is convex, establishing the uniqueness of the equilibrium solution. (6), (7), and (9) are obtained from the first-order condition of supplier i. Clearly, the optimal s_i is a function of θ, which is determined from the (AR) constraint, as in (8). The supplier's profit and the prime's expenditures follow immediately. ∎

Proof of Proposition 2. The following identities are needed to prove solution uniqueness:

$$\partial Var[B_i\,|\,s_i]/\partial s_i = -2F_i(s_i)E[B_i\,|\,s_i] \le 0, \quad (15)$$
$$\partial^2 Var[B_i\,|\,s_i]/\partial s_i^2 = -2f_i(s_i)E[B_i\,|\,s_i] + 2F_i(s_i)[1 - F_i(s_i)]. \quad (16)$$

Let us drop the subscript i for notational convenience. Differentiating the supplier's expected utility function (4) with respect to s, we find that

$$\partial U/\partial s = -(1-\alpha)c + v[1 - F(s)] + rv^2 F(s)E[B\,|\,s], \quad (17)$$

which is greater than zero for all s if $\alpha = 1$ and $v > 0$, because U increases without bound in this case. If $\alpha < 1$ and $v[1 - F(0)] \geq (1 - \alpha)c$, then $\partial U/\partial s \geq 0$ at $s = 0$, so U is nondecreasing initially. Notice also that $\lim_{s \to \infty} \partial U/\partial s = -(1 - \alpha)c < 0$, so there exists at least one critical point on $[0, \infty)$. Setting $\partial U/\partial s = 0$, we obtain $E[B \mid s^*] = (1 - \alpha)c - v[1 - F(s^*)]/\left(rv^2 F(s^*)\right)$. Substituting this result into the second derivative

$$\partial^2 U/\partial s^2 = -vf(s) + rv^2 f(s)E[B \mid s] - rv^2 F(s)[1 - F(s)] \tag{18}$$

(note (15) and (16) are used), we obtain

$$
\begin{aligned}
\partial U/\partial s|_{s=s^*} &= -vf(s^*) - [v - (1 - \alpha)c]f(s^*)/F(s^*) + vf(s^*) - rv^2 F(s^*)[1 - F(s^*)] \\
&= -[v - (1 - \alpha)c]f(s^*)/F(s^*) - rv^2 F(s^*)[1 - F(s^*)] < 0,
\end{aligned}
$$

where the inequality follows from the condition $v \geq v[1 - F(0)] \geq (1 - \alpha)c$. Since the second derivative is negative at every critical point, s^* cannot be a minimizer. Combining this result with $\partial U/\partial s|_{s=0} > 0$ and $\lim_{s \to \infty} \partial U/\partial s < 0$, we conclude that U has a unique maximizer. Optimal solutions follow from the first-order conditions. ∎

Proof of Corollary 1. We drop the subscript i for notational convenience. After differentiating the first-order condition (17) implicitly with respect to r (optimal s is a function of r, i.e., $s^* = s(r)$) and collecting the terms we obtain

$$\frac{\partial s^*}{\partial r} = \frac{v^2 F(s^*)E[B \mid s^*]}{vf(s^*) - rv^2 f(s^*)E[B \mid s^*] + rv^2 F(s^*)[1 - F(s^*)]}.$$

Notice that the denominator has the sign opposite of that in (18). Hence $\partial s^*/\partial r > 0$. Similarly,

$$
\begin{aligned}
\partial a^*/\partial \alpha &= -1/k < 0, \\
\frac{\partial s^*}{\partial \alpha} &= \frac{c}{vf(s^*) - rv^2 f(s^*)E[B \mid s^*] + rv^2 F(s^*)[1 - F(s^*)]} > 0, \\
\partial a^*/\partial v &= 0, \\
\frac{\partial s^*}{\partial v} &= \frac{[1 - F(s^*)] + 2rvF(s^*)E[B \mid s^*]}{vf(s^*) - rv^2 f(s^*)E[B \mid s^*] + rv^2 F(s^*)[1 - F(s^*)]} > 0.
\end{aligned}
$$

∎

Proof of Proposition 3. With $r_0 = r_1 = ... = r_n = 0$, the prime's Lagrangian for supplier i becomes

$$\mathcal{L}_i(a_i, s_i, \theta) = c_i s_i + \theta E\left[B_i \mid s_i\right] - a_i + k_i a_i^2/2$$

and solutions are given by (6), (7), and (8). From the supplier's utility

$$U_i(a_i, s_i, w_i, v_i, \alpha_i) = w_i - (1 - \alpha_i)(c_i s_i - a_i) - v_i E\left[B_i \mid s_i\right] - k_i a_i^2/2$$

it is clear that setting $\alpha_i = 0$ and $v_i = \theta$ yields the same Lagrangian (with the reverse sign) as $\mathcal{L}_i$ plus a constant, reproducing the first-best solutions. The supplier's profit and the prime's expenditure follow immediately. ∎

Proof of Lemma 1. Define $\gamma \equiv 4cF(s)E\left[B \mid s\right]/[1 - F(s)]^2$ and note that (13) can be rewritten as

$$v(\alpha) = \frac{2c}{1 - F(s)} \frac{1}{r\gamma} \left(-1 + \sqrt{1 + r\gamma(1 - \alpha)}\right),$$

from which we obtain

$$v'(\alpha) = -\frac{c}{1 - F(s)} \frac{1}{\sqrt{1 + r\gamma(1 - \alpha)}}, \qquad v''(\alpha) = -\frac{1}{2} \frac{c}{1 - F(s)} \frac{r\gamma}{[1 + r\gamma(1 - \alpha)]^{3/2}},$$

and

$$\frac{\partial(v^2)}{\partial \alpha} = 2v(a)v'(a) = -\frac{4c^2}{[1 - F(s)]^2} \frac{1}{r\gamma} \left(1 - \frac{1}{\sqrt{1 + r\gamma(1 - \alpha)}}\right),$$

$$\frac{\partial^2(v^2)}{\partial \alpha^2} = 2(v'(a))^2 + 2v(a)v''(a) = \frac{2c^2}{[1 - F(s)]^2} \frac{1}{[1 + r\gamma(1 - \alpha)]^{3/2}}.$$

Differentiating the Lagrangian (12) and substituting $\partial(v^2)/\partial a$ and $\partial^2(v^2)/\partial a^2$, we find that

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \frac{\alpha}{k} + [(r_0 + r)\alpha - r]Var[\varepsilon] - \frac{2(r_0 + r)c^2}{[1 - F(s)]^2} \frac{1}{r\gamma} \left(1 - \frac{1}{\sqrt{1 + r\gamma(1 - \alpha)}}\right) Var[B \mid s],$$

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = \frac{1}{k} + (r_0 + r)Var[\varepsilon] + \frac{(r_0 + r)c^2}{[1 - F(s)]^2} \frac{1}{[1 + r\gamma(1 - \alpha)]^{3/2}} Var[B \mid s] > 0.$$

∎

Proof of Proposition 5. We use the results in the proof of Lemma 1. Define

$$\ell_{FB}(\alpha) \equiv [(r_0 + r)\alpha - r]Var[\varepsilon],$$

$$\ell_{SO}(\alpha) \equiv \frac{\alpha}{k} + [(r_0 + r)\alpha - r]Var[\varepsilon],$$

$$\ell_{AO}(\alpha) \;\equiv\; [(r_0+r)\alpha - r]Var[\varepsilon] - \frac{2(r_0+r)c^2}{[1-F(s)]^2}\frac{1}{r\gamma}\left(1 - \frac{1}{\sqrt{1+r\gamma(1-\alpha)}}\right)Var[B\,|\,s],$$

$$\ell_{SB}(\alpha) \;\equiv\; \frac{\partial \mathcal{L}}{\partial \alpha}(\alpha) = \frac{\alpha}{k} + [(r_0+r)\alpha - r]Var[\varepsilon] - \frac{2(r_0+r)c^2}{[1-F(s)]^2}\frac{1}{r\gamma}\left(1 - \frac{1}{\sqrt{1+r\gamma(1-\alpha)}}\right)Var[B\,|\,s].$$

α^{FB}, α^{SO}, α^{AO}, and α^{SB} are the solutions to $\ell_{FB}(\alpha)=0$, $\ell_{SO}(\alpha)=0$, $\ell_{AO}(\alpha)=0$, and $\ell_{SB}(\alpha)=0$, respectively. Observe $\ell_{AO}(\alpha) \le \ell_{SB}(\alpha) \le \ell_{SO}(\alpha)$ for any α. Since $\ell'_j(\alpha) > 0$ for all j, $\alpha^{SO} \le \alpha^{SB} \le \alpha^{AO}$. In contrast, both $\ell_{FB}(\alpha) \le \ell_{SB}(\alpha)$ and $\ell_{FB}(\alpha) \ge \ell_{SB}(\alpha)$ are possible. To see this, substitute $\alpha^{FB} = r/(r_0+r)$ in ℓ_{SB} to obtain

$$\ell_{SB}(\alpha^{FB}) = \frac{1}{k}\frac{r}{r_0+r} - \frac{2(r_0+r)c^2}{[1-F(s)]^2}\frac{1}{r\gamma}\left(1 - \frac{1}{\sqrt{1+r\gamma\frac{r_0}{r_0+r}}}\right)Var[B\,|\,s].$$

Let $\delta \equiv r/r_0$ and rewrite $\ell_{SB}(\alpha^{FB})$ as a function of δ:

$$\widetilde{\ell}_{SB}(\delta) \equiv \ell_{SB}(\alpha^{FB}) = \frac{1}{k}\frac{\delta}{1+\delta} - \frac{2c^2}{[1-F(s)]^2}\frac{1}{r\gamma}\left(1+\frac{1}{\delta}\right)\left(1 - \frac{1}{\sqrt{1+r\gamma\frac{1}{1+\delta}}}\right)Var[B\,|\,s].$$

Differentiating, we see that

$$\widetilde{\ell}'_{SB}(\delta) = \frac{1}{k}\frac{1}{(1+\delta)^2} + \frac{2c^2}{[1-F(s)]^2}\frac{1}{r\gamma}\left[\frac{1}{\delta^2}\left(1 - \frac{1}{\sqrt{1+r\gamma\frac{1}{1+\delta}}}\right) + \frac{(1+1/\delta)}{(1+\delta)^2}\frac{r\gamma}{2\left(1+r\gamma\frac{1}{1+\delta}\right)^{3/2}}\right]Var[B\,|\,s] > 0.$$

Hence, $\widetilde{\ell}_{SB}(\delta)$ is increasing. Notice $\lim_{\delta\to 0}\widetilde{\ell}_{SB}(\delta) = -\infty$, $\lim_{\delta\to\infty}\widetilde{\ell}_{SB}(\delta) = 1/k$. Therefore there is a unique $\delta^\dagger$ such that $\widetilde{\ell}_{SB}(\delta^\dagger) = 0$. Since $\widetilde{\ell}_{SB}(\delta)$ is increasing, $\widetilde{\ell}_{SB}(\delta) = \ell_{SB}(\alpha^{FB}) > 0$ for all $\delta > \delta^\dagger$, implying that $\alpha^{SB} < \alpha^{FB}$, since $\ell_{SB}(\alpha)$ is also increasing. Likewise, $\alpha^{SB} > \alpha^{FB}$ for all $\delta < \delta^\dagger$.

Differentiating $\ell_{SB}(\alpha^{SB}) = 0$ with respect to $Var[\varepsilon]$ and collecting terms, we see that

$$\frac{d\alpha^{SB}}{d(Var[\varepsilon])} = \frac{r - (r_0+r)\alpha^{SB}}{\frac{1}{k} + (r_0+r)Var[\varepsilon] + \frac{(r_0+r)c^2}{[1-F(s)]^2}\frac{1}{r\gamma}\frac{1}{[1+r\gamma(1-\alpha^{SB})]^{3/2}}Var[B\,|\,s]}.$$

The numerator is negative if $\alpha^{SB} > \alpha^{FB}$, zero if $\alpha^{SB} = \alpha^{FB}$, and positive if $\alpha^{SB} < \alpha^{FB}$. The sign of $dv^{SB}/d(Var[\varepsilon])$ is the reverse of that of $d\alpha^{SB}/d(Var[\varepsilon])$ via (13). ∎

	$r_0 = 0.01$			$r_0 = 0.1$			$r_0 = 1$		
r_2	0.01	0.1	1	0.01	0.1	1	0.01	0.1	1
α_2^{FB}	0.500	0.909	0.990	0.091	0.500	0.909	0.010	0.091	0.500
α_1^{SB}	0.279	0.279	0.279	0.318	0.318	0.318	0.430	0.430	0.429
α_2^{SB}	0.060	0.279	0.755	0.184	0.318	0.726	0.436	0.430	0.539
v_1^{SB}	0.994	0.993	0.987	0.952	0.944	0.934	0.823	0.801	0.758
v_2^{SB}	1.408	0.993	0.288	1.215	0.944	0.317	0.826	0.801	0.490
a_1^{SB}	3.603	3.603	3.603	3.409	3.409	3.409	2.849	2.852	2.856
a_2^{SB}	4.699	3.603	1.224	4.082	3.409	1.371	2.820	2.852	2.303
s_1^{SB}	8.729	8.725	8.678	8.779	8.725	8.650	8.905	8.725	8.330
s_2^{SB}	8.722	8.725	8.773	8.672	8.725	8.802	8.559	8.725	9.151
$-U_0$	13.342	13.923	15.153	14.000	14.435	15.676	19.104	19.153	20.141

Table 4: Effects of changing r_2. Italics indicate symmetric parameters.

	$r_0 = 0.01$ ($\alpha_2^{FB} = 0.909$)			$r_0 = 0.1$ ($\alpha_2^{FB} = 0.5$)			$r_0 = 1$ ($\alpha_2^{FB} = 0.091$)		
$Var[\varepsilon_2]$	1	10	100	1	10	100	1	10	100
α_1^{SB}	0.279	0.279	0.279	0.318	0.318	0.318	0.429	0.430	0.430
α_2^{SB}	0.172	0.279	0.650	0.268	0.318	0.442	0.654	0.430	0.156
v_1^{SB}	0.993	0.993	0.992	0.944	0.944	0.944	0.794	0.801	0.800
v_2^{SB}	1.127	0.993	0.507	1.007	0.944	0.785	0.505	0.801	1.149
a_1^{SB}	3.603	3.603	3.603	3.409	3.409	3.409	2.853	2.852	2.852
a_2^{SB}	4.138	3.603	1.752	3.658	3.409	2.790	1.731	2.852	4.221
s_1^{SB}	8.724	8.725	8.718	8.725	8.725	8.723	8.670	8.725	8.715
s_2^{SB}	8.727	8.725	8.733	8.726	8.725	8.728	8.782	8.725	8.735
$-U_0$	13.653	13.923	15.139	14.172	14.435	16.780	17.794	19.153	24.341

Table 5: Effects of changing $Var[\varepsilon_2]$ when $r_1 = r_2 = 0.1$. Italics indicate symmetric parameters.

$\widehat{B}_0$	1	2	3	4	5	6	7
α_1^{SB}	0.421	0.404	0.398	0.396	0.396	0.397	0.399
α_2^{SB}	0.592	0.604	0.614	0.624	0.634	0.644	0.653
v_1^{SB}	1.678	1.147	0.935	0.820	0.748	0.699	0.666
v_2^{SB}	0.792	0.557	0.463	0.412	0.379	0.357	0.341
a_1^{SB}	2.896	2.980	3.011	3.021	3.021	3.015	3.006
a_2^{SB}	2.039	1.981	1.929	1.879	1.830	1.782	1.737
s_1^{SB}	11.852	10.384	9.355	8.511	7.765	7.078	6.425
s_2^{SB}	12.205	10.752	9.751	8.948	8.259	7.644	7.083
$(s_2^{SB} - s_1^{SB})/s_1^{FB}$	2.94%	3.48%	4.15%	5.00%	6.17%	7.69%	9.75%
IIR_1	11.852	10.384	9.355	8.511	7.765	7.078	6.425
IIR_2	12.205	10.752	9.751	8.948	8.259	7.644	7.083
NCR_1	2.057	2.092	2.104	2.108	2.108	2.106	2.102
NCR_2	1.623	1.589	1.557	1.526	1.495	1.465	1.435
CRP_1	0.610	0.586	0.577	0.574	0.574	0.576	0.578
CRP_2	1.708	1.696	1.687	1.680	1.675	1.671	1.668
PRP_1	1.252	1.160	1.116	1.090	1.074	1.062	1.055
PRP_2	0.574	0.587	0.600	0.611	0.621	0.628	0.632
$-U_0$	24.521	21.484	19.425	17.781	16.364	15.088	13.905

Table 6: Effects of changing $\hat{B}_0$, with $r_0 = 0.5$, $r_1 = 0.1$ and $r_2 = 1$. IIR stands for investment in resources and is equal to $c_i s_i^{SB}$. NCR is $-a_i^{SB} + \frac{1}{2}k_i(a_i^{SB})^2$, the net cost reduction. CRP is the residual cost risk premium, $\frac{1}{2}(r_0(\alpha_i^{SB})^2 + r_i(1 - \alpha_i^{SB})^2)Var[\varepsilon_i]$, and PRP is the residual performance risk premium, $\frac{1}{2}(r_0 + r_i)(v_i^{SB})^2 Var[B_i|s_i^{SB}]$.