# Reliability or Inventory? Contracting Strategies for After-Sales Product Support

Sang-Hyun Kim, Morris A. Cohen, and Serguei Netessine*
The Wharton School, University of Pennsylvania
Philadelphia, PA 19104

March 22, 2007

**Abstract**

Traditional sourcing arrangements for after-sales support of capital-intensive products such as airplanes, weapons systems, and manufacturing equipment have centered around physical assets. Typically, customers would pay the supplier of maintenance and repair services (usually the product manufacturer) based on the resources, such as spare parts, that were needed to maintain the products. In recent years, we have witnessed the emergence of a new service contracting strategy called Performance Based Logistics (PBL). Under PBL, the basis of supplier compensation is actual uptime (or availability) of the product. PBL implementations are now mandated by the Department of Defense for all new system acquisition programs.

The goal of this paper is to compare the inefficiencies arising under the traditional (pre-PBL or *material* contract) and the PBL contract. We use a principal-agent framework with multitasking and let the supplier set the base-stock inventory level of spare parts as well as invest in increasing product reliability. The customer (principal) sets the contract terms. We find that, in a majority of situations, both contracts result in suboptimal supply chain performance manifested by low product reliability and high inventory of spares (relative to the first-best solution). However, the PBL contract provides stronger incentives to invest in reliability improvements than does the material contract. Moreover, the efficiency of the PBL contract improves (i.e., inventory decreases and reliability increases) if the supplier owns a larger portion of the spare assets. In particular, under a PBL contract when the supplier owns all spare assets, the supply chain becomes coordinated. Our analysis calls for transforming suppliers into total service providers of support services who, under the PBL arrangement, assume complete control of service functions, including asset ownership.

*Email: shkim@wharton.upenn.edu, cohen@wharton.upenn.edu, netessine@wharton.upenn.edu

# 1 Introduction

The importance of after-sales product support is growing in such capital-intensive industries as aerospace, defense, and industrial equipment. For example, in the defense industry only about 28% of a weapon system's total ownership cost is attributed to development and procurement, whereas the costs to operate, maintain, and dispose of the system account for the remaining 72% (GAO report 2003). Given that the U.S. Department of Defense's (DoD) annual budget for operations and maintenance is projected to be approximately $70B in 2007 (a 25% increase over 2006), it is not surprising that the manufacturers of military aircraft, engines, and avionic equipment (e.g., Boeing, GE, Honeywell, Lockheed Martin, Pratt & Whitney, and Rolls-Royce) have been developing competitive strategies for the provision of spare parts and repair/maintenance services. Such after-sales services are a high-margin business (Cohen et al. 2006) that currently accounts for 40-50% of profits and 25% of total revenue across all manufacturing companies (Bijesse et al. 2002).

Traditionally, many after-sales contractual relationships in the aerospace and defense industry were governed by simple *material* contracts that were based on the number of spares needed to maintain the product. However, increasing pressure on the Department of Defense to reduce spending as well as growing dissatisfaction with the level of after-sales support from key suppliers have led to re-evaluation of these arrangements. In recent years, a novel strategy for aligning interests in the after-sales service supply chains has emerged: Performance Based Logistics (PBL). Its premise is simple: instead of paying suppliers for their parts, labor, and other services, the compensation is based on the actual availability of the product to the customer. The key idea is to align the incentives of all parties by tying suppliers' compensation to the same service value (availability) that the customer cares about. After several pilot studies, the DoD mandated the implementation of PBL contracts for all new system acquisition programs beginning in 2003 (DoD Directive 5000.1). Initial reports support the view that PBL contracts improve product availability: the U.S. Navy's implementation of PBL for its fleet of F/A-18 E/F fighter jets, for example, has resulted in an availability increase from 67% to 85%, while a similar effort has seen the availability of Aegis guided missile cruisers rise from 62% to 94% (Geary 2006).

The ultimate goal of the PBL contract – providing incentives to suppliers to attain high availability at a lower cost – can be achieved through a variety of actions. In this paper, our focus is on the tradeoff and interaction between two such actions: spare asset management and investment into product reliability. This focus is motivated by our close interactions with industry practitioners and government agencies regarding current PBL contract implementation practices, as well as by the void regarding this tradeoff in the operations management literature. In particular, the Under Secretary of

Defense for Acquisition, Technology and Logistics has issued a guidance (see Wynne 2004) regarding key performance criteria for PBL arrangements which include availability, reliability, cost of ownership, response time, and logistic footprint (which includes inventory, personnel, facilities, etc.). Clearly, spare asset management and product reliability critically affect every one of these performance criteria, a thesis that has been independently confirmed by many industry practitioners.[1] Moreover, a recent DoD Guidebook (2005) designates reliability as one of the three essential elements (along with asset availability and product maintainability) that enable mission capability. Given these considerations, we aim to address the following research questions: How does a PBL contract differ from a material contract in motivating suppliers to improve reliability and to manage the inventory of spares? What kind of inefficiencies arise under these two contracts? Does the ownership structure of the spare assets (by the customer or by the supplier) affect the answers to these questions?

In this paper, we develop a stylized economic model that draws upon two distinct bodies of literature. We employ the classical service parts inventory management model to represent repair and maintenance processes. This model is further enriched by a novel feature: endogenous product reliability improvement effort. The relationship between the customer and the supplier is modeled using the principal-agent paradigm in which the principal sets the terms of the contract in order to minimize her total cost subject to a product availability constraint. The supplier's goal is to maximize his profit given these contract terms. We allow for an arbitrary allocation of spare inventory ownership between the customer and the supplier and consider two types of contracts. Under the material contract the supplier is compensated for each spare unit that he stocks, and under the performance contract the compensation is based on product availability. We find that, in a majority of cases, both contracts result in inefficiencies such that the supplier invests less in reliability and more in inventory of spares than the integrated firm would. Compared to the material contract, we show that the PBL contract incentivizes the supplier to invest more into reliability improvement and at the same time to stock less inventory. Furthermore, we find that the full benefits of the PBL contract are realized only when spare assets are fully owned by the supplier; in fact, in this case the PBL contract achieves the first-best solution and coordinates the channel. Interestingly, our interactions with practitioners indicate that the prevailing industry practice is for the customer to own spare assets while the supplier decides on target stocking levels of spares and recommends to the customer a budget of spares acquisitions to achieve these levels. We suspect that this ownership/decision structure is largely a relic of pre-PBL practice and of the fact that customers, especially government agencies such as the armed forces, are historically reluctant to cede control of their assets to third parties due to the fear of mismanagement

[1] We are grateful to the many participants of the Wharton Service Supply Chain Thought Leaders' Forum for bringing this issue to our attention. See http://opim.wharton.upenn.edu/fd/forum/.

and the potentially catastrophic costs of product downtime. While this is understandable, our findings indicate that such resistance may actually be an impediment to the successful implementation of the PBL strategy. Thus, our analysis suggests that there are significant benefits for transforming military suppliers into total service providers who assume complete control of service functions, including asset ownership.

The rest of the paper is organized as follows. After a brief survey of the related literature, we provide our modeling assumptions and formulation in Section 3. In Section 4 we present analysis of material and performance contracts and a comparison between them. This is followed by Section 5, in which we consider the consequences of relaxing some of the basic assumptions we make in our analysis. Section 6 concludes our investigation with a summary of major findings and areas of future research.

## 2 Literature Review

Our model represents an application of the principal-agent framework to a service parts inventory management problem. Sherbrooke (1968) introduced the classical METRIC model for service parts (repairables) in the 1960's which led to numerous multi-echelon, multi-indentured inventory model extensions. In METRIC, the repair process for each part is represented by an $M/G/\infty$ queueing system, and the decision is to optimize the number of spares in stock given an exogenous part failure rate. Over the years the METRIC model and related models inspired by non-military applications (e.g., Cohen et al. 1989) have become the basis for a number of decision support systems that are currently used in both commercial and military settings (for example, see Cohen et al. 1990). Two recent books, Sherbrooke (2004) and Muckstadt (2005), present the details of the METRIC model and summarize advances within this stream of research. Recent papers in this stream include Deshpande et al. (2003) and Deshpande et al. (2006). Despite the large volume of literature in this field, the issues of contracting and outsourcing have remained largely unaddressed. We use a simplified version of the repairable model in order to minimize complexity arising from the game-theoretic aspects of the model.

We employ the principal-agent framework to represent the contractual relationship between the customer and the supplier. Bolton and Dewatripont (2005) provide a comprehensive survey of this area of research. Our model is most closely related to the multitasking literature (Holmström and Milgrom 1991, Gibbons 2005), in which the agent controls more than one action (inventory and reliability in our case). Interestingly, the original paper by Holmström and Milgrom (1991) contains the following motivating example whose idea is quite similar to our problem: “As a simple example, production workers may be responsible for producing a high volume of good quality output, or they may be required both to produce output and to care for the machines they use... if volume output is easy to measure but the quality is not, then a system of piece rates for output may lead agents to increase the volume of output at the expense of quality." Despite the similarity, our model is unique in that it ventures beyond this abstract example and attaches significance to the terms "volume" and "quality" motivated by our specific problem context of system support. We also consider other operational elements such as the target service level, which economic models ignore but which are important in practice. Thus, our model offers an interpretation of the multitask principal-agent model in terms of operational variables, namely investment in reliability and inventory, in the concrete setting of the after-sales product support supply chain for mission-critical products such as military weapon systems.

The main focus of this paper is on the tradeoff between reliability improvement and inventory level decisions made by the supplier. Thus, we endogenize the product failure rate which, to the best of our knowledge, has never been done in the service parts inventory management literature. In this respect our model has a connection to the controlled queue literature, which dates back to Naor (1969). Stidham (1985) provides an early survey of this research stream. Several recent papers study principal-agent models with controlled queues in settings other than spare parts inventory management. Ren and Zhou (2006) and Hasija et al. (2006) consider call center outsourcing, while Plambeck and Zenios (2000), Lu et al. (2006) and Baiman et al. (2007) study production systems.

Early papers that discuss incentives and contracting in the defense industry include Cummins (1977) and Rogerson (1994). Kang et al. (2005) propose a decision-support model that can help support PBL relationships by trading off reliability and maintenance tasks. While the motivation of this paper is very similar to ours, it does not explicitly model decentralized decision-making arising in support arrangements. Finally, in our earlier work (Kim et al. 2007), we considered the tradeoff between incentive and insurance against risks under a general contracting arrangement that includes PBL, and analyzed the implications to performance and cost of ownership. In the present paper we do not incorporate risk considerations, instead focusing on reliability improvement and its interaction with inventory management decisions.

In summary, the analytical contributions of our paper are two-fold. First, we endogenize reliability improvement decisions in a classical repairable inventory management model and, for the first time, study the tradeoff between inventory and reliability. Second, we study and compare two frequently used contractual arrangements (material and PBL), evaluate their inefficiencies and identify the arrangement (PBL contract with supplier ownership of spare assets) that achieves the first-best solution. From a managerial perspective, our paper sheds light on how performance-based incentives can lead to reliability improvement and on the role of supply chain re-structuring in achieving an efficient solution.

## 3 The Model

A risk-neutral customer owns and operates a fleet of $N$ identical products, whose continued usage is disrupted by random product failures occurring at a rate $\lambda$. We denote the Mean Time Between Failures (MTBF) by $\tau = 1/\lambda$, which is assumed to be bounded from below and above by $\underline{\tau}$ and $\overline{\tau}$, respectively. The supplier performs three kinds of activities to support the customer's fleet of products: (1) repairs defective units, (2) manages spare product inventory, and (3) manages product reliability. For simplicity, we only consider a single indenture level for the product, i.e., spares inventory is managed at a product level (in practice, inventory to support maintenance and repair operations primarily consists of parts at different indenture and echelon levels; see Section 5 for further discussion). We assume that a repair facility with a given capacity is already in place, and that the expected repair lead time $l$ is fixed (which is reasonable if repairs are always performed at full speed) and is normalized to one.

At the beginning of the product support phase the customer designs and offers a contract which influences the supplier's decisions on product reliability $\tau$ and the inventory of spares $s$. By assuming that $s$ is the supplier's choice, we only consider the case of Vendor Managed Inventory (VMI), which is the prevailing practice in after-sales product support relationships. We model different asset ownership structures using the parameter $0 \leq \delta \leq 1$, which represents the proportion of spare assets owned by the supplier. Thus, when $\delta = 0$, the customer owns the inventory (and hence incurs holding costs), whereas when $\delta = 1$, the ownership is with the supplier. Thus, we allow for both procurement and consignment inventory management settings.

### 3.1 Repair Process and Performance Measurement

Following the classical service parts inventory management literature, we model the repair facility as an M/G/$\infty$ queue. Product failures occur according to a Poisson process, and the failed product is replaced by a working unit from the spares inventory, if available. Otherwise, a backorder occurs. A one-for-one base stock inventory policy, characterized by a target stocking level $s$, is used: each failed product immediately undergoes a repair that takes a random amount of time with a general distribution function. Note that the Poisson failure process is not an exact representation since, in general, the failure rate in the closed-loop repair cycle (i.e., repaired units are restored back to the system) depends on the number of deployed units that are in working condition. However, this model is a good approximation as long as $\lambda = 1/\tau \ll N$ is satisfied, which is true in most environments where products fail relatively infrequently. This is indeed a standard assumption in the service parts management literature, including the paper by Sherbrooke (1968) that first introduced the METRIC model.

The Poisson failure assumption allows the application of Palm's Theorem, which postulates that the inventory on-order $O(\tau)$ (the number of units that are being repaired at a random point in time) is Poisson-distributed with mean $\lambda l = \lambda = 1/\tau$ (see Muckstadt 2005). The random variables of interest, on-hand inventory $I$ and backorder $B$, are related to $O(\tau)$ and $s$ by $I \,|\, \tau, s = (s - O(\tau))^+$ and $B \,|\, \tau, s = (O(\tau) - s)^+$, respectively. Backorders are also related to expected availability $A$ as follows: $A = 1 - E[B \,|\, \tau, s]/N$. We assume that the customer faces the service requirement $A \geq A^\circ$ (e.g., expected availability should be 95% or more), which is translated into the backorder constraint $E[B \,|\, \tau, s] \leq B^\circ$.

### 3.2 Contracting

At the beginning of the product support phase the customer offers to the supplier a contractual payment $T$, whose form will be discussed shortly. In response, the supplier sets $\tau$ and $s$ optimally in order to maximize its expected profit[2] given by

$$\pi(\tau, s) = E[T \,|\, \tau, s] - \psi(\tau, s) = E[T \,|\, \tau, s] - cs - K(\tau) - \delta h E[I \,|\, \tau, s]. \tag{1}$$

We use $\psi(\tau, s)$ to denote the supplier's internal cost, which includes three elements: the linear production/acquisition cost $cs$, the reliability improvement cost $K(\tau)$, and the expected cost of holding inventory $\delta h E[I \,|\, \tau, s]$. $c$ is the unit cost of manufacturing/purchasing each spare product, and $h$ is the net present value of the holding cost throughout the contract duration.[3] We assume that all of these costs are public knowledge (a discussion of other possible costs are discussed later in Section 5). $K(\tau)$ represents the dollar amount of investment in research and development or engineering changes required to improve reliability to $\tau$. We assume that $K(\tau)$ is increasing and convex, i.e., $K'(\tau) > 0$, $K''(\tau) > 0$. Implicit in this assumption is the simplification that the supplier has many technological choices whose combined effect maps to each value of $\tau$. Convexity is a reasonable assumption since in general the most efficient improvement opportunities will be exploited first. Furthermore, we assume that $K'''(\tau) > 0$, $K(\underline{\tau}) = 0$, and $\lim_{\tau \to \overline{\tau}} K(\tau) = \lim_{\tau \to \overline{\tau}} K'(\tau) = \infty$. Hence, $\underline{\tau}$ can be interpreted as the baseline reliability that the supplier can provide without incurring the extra cost $K(\tau)$ (i.e., the

[2]If the customer and/or the supplier are risk-averse, expected profit maximization is not an approriate criterion. See Kim et al. (2007) for an alternative model capturing risk aversion.

[3]In general, the variable costs $c$ and $h$ may depend on $\tau$. We believe that such dependence is a second-order effect because most costs associated with reliability improvement are fixed and hence we ignore this interaction. In defense logistics, most reliability improvements come from engineering changes in which a part is re-designed and the cost is absorbed in a fixed investment into research and development. Allocation of that investment to the variable cost of the product could happen, but the impact of improved reliability on the replacement rate of the product and the investment in inventory are our key interests here.

original design specifications for reliability when the product was adopted). Similarly, $\overline{\tau}$ represents the theoretical upper bound on MTBF, i.e., it becomes infinitely expensive to achieve $\overline{\tau}$. Throughout the paper, we also make the following technical assumption:

$$\underline{\tau}^3 K''(\underline{\tau}) \geq 2h. \tag{2}$$

That is, $K(\tau)$ is sufficiently convex on its domain (recall that $K'''(\tau) > 0$). For suitable choices of $\underline{\tau}, \overline{\tau}$, and $K(\tau)$, (2) is a reasonable assumption. This condition greatly simplifies our exposition and allows us to focus on the most important features of our model. We discuss the implications of relaxing (2) in Appendix B.

Anticipating the supplier's response, the customer determines contract terms that would minimize her total cost subject to the availability constraint while making sure that the supplier participates in the trade. We restrict contract forms to the following linear function:

$$T = w + ps - vB, \tag{3}$$

where $w$ is a lump-sum payment, $p$ is the unit price paid to the supplier for each unit of spare product the supplier provides, and $v$ is the penalty rate for each realized backorder.[4] We therefore assume that the customer cannot contract directly on either the inventory level $s$ or the product reliability $\tau$: instead, consistent with current industry practice, contract terms only include a combination of $w$, $p$, and $v$.

This rather general contract form allows us to model a material contract $T = w + ps$ (with $v = 0$) and a "pure" PBL contract $T = w - vB$ (with $p = 0$) as special cases of (3). The material contract represents the traditional parts pricing scheme that was ubiquitous before the introduction of the PBL. The parameter $p$ can be interpreted as a *reservation price* that secures a spare quantity $s$ for support of the customer. Note that $p$ also has a straightforward meaning as the purchase price in the case of spares procurement ($\delta = 0$, i.e., ownership is transferred to the customer). A similar meaning is not relevant when the supplier retains ownership ($\delta = 1$). With the performance contract, the

[4]There is an ongoing debate in the defense industry on whether the appropriate measure of performance is the steady-state availability random variable, which maps to $B$, or its time average, which maps to $\widetilde{B}(t) = \frac{1}{t}\int_0^t B(u)du$, where $t$ is the performance evaluation period length. The DoD Guide for Achieving Reliability, Availability, and Maintainability (2005) defines availability as "a measure of the degree to which an item is in an operable state and can be committed at the start of a mission when the mission is called for at an unknown (random) point in time", clearly advocating the usage of $B$. However, the time average is a more readily measurable metric and is widely used in practice. Although this issue merits a separate in-depth study, the distinction has a minimal impact on our model as the two are identical in expectation (i.e., $E[B \mid \tau, s] = E[\widetilde{B} \mid \tau, s]$), which is the only quantity representing the backorder in our model. Note that Kim et al. (2007) considers an arrangement in which the distinction becomes important since the assumption of risk aversion requires the computation of variance of each measures.

payment is based on the realization of the random variable $B$. In practice, payments associated with performance are made throughout the duration of the customer-supplier relationship and backorder levels are evaluated periodically. For simplicity, however, we do not model the details of cash flow dynamics and we sidestep related intertemporal incentive issues. Thus, $v$ represents the net present value of the cumulative penalty rate for the backorder $B$ over the service contract duration, measured in dollars. The magnitude of $v$ will be determined by the loss of "value" incurred by the customer due to the unavailability of one unit of the product.

It should be noted that we do not consider payment terms based on observed reliability, i.e., time between product failures. This is consistent with current practice, in which the majority of PBL contracts are usually based on availability and rarely on reliability. The main reason is that failures occur infrequently in most service environments we are concerned with and therefore it is impractical to base compensation on such rarely observed events. Although we consider only one aggregated product type in our stylized model, in reality each product consists of thousands of parts so that failure characteristics and required inventory levels are determined at the parts level. Thus, PBL contracts can be offered to multiple companies that provide different product modules or parts, and it is very common to see one or at most a few failure events occurring for each module within a performance evaluation interval, which is typically a few weeks or a month.[5] In addition, dynamic product deployment plans and scheduled/unscheduled maintenance events make it difficult for the customer to accurately measure true failure rates since the time during which the parts were not utilized should not be counted toward the time between failures. In contrast, availability does not suffer from this difficulty because it can be measured quite unambiguously at any point in time by both the customer and the supplier.

To summarize, the customer's problem can be written as

$$
\begin{aligned}
(SB) \qquad \min_{\mathcal{P}} \qquad & C^{SB} \equiv E[T \,|\, \tau^*, s^*] + (1-\delta)hE[I \,|\, \tau^*, s^*] \\
\text{s.t.} \qquad & E[B \,|\, \tau^*, s^*] \leq B^{\circ} \ , \\
& \pi(\tau^*, s^*) \geq 0, && \text{(IR)} \\
& (\tau^*, s^*) \in \arg\max \pi(\tau, s) \ , && \text{(IC)}
\end{aligned}
$$

where $\mathcal{P}$ denotes the contract parameter space, consisting of $\{w, p, v\}$. Note that $(\tau^*, s^*)$ are functions of $p$ and/or $v$. As is standard in principal-agent models, two constraints are included in the customer's problem: the individual rationality (IR) constraint ensures that the supplier participates in the trade, while the incentive compatibility (IC) constraint describes that the supplier makes decisions in a self-

[5] However, the annual number of failures is often tracked and used as a basis for renewing the service contract.

interested manner (i.e., maximizes his own profit). Without loss of generality, we assume that the supplier's reservation profit (the right-hand side of the (IR) constraint) is set to zero. We call this a second-best ($SB$) problem in view of the inefficiency that is inherent in the decentralized decision-making setup. The dimensionality of this problem is reduced after recognizing that the (IR) constraint is always binding at the optimum by adjusting $w$ accordingly. Hence and the problem becomes

$$(\widetilde{SB}) \qquad \min_{\mathcal{P}\backslash\{w\}} \widetilde{C}^{SB} \equiv cs^* + K(\tau^*) + hE[I \mid \tau^*, s^*] \text{ subject to } E[B \mid \tau^*, s^*] \leq B^\circ \text{ and (IC).}$$

Finally, we impose the following technical assumptions whose purpose is to eliminate solutions that rarely occur in practice:

$$\underline{\tau} < 1/B^\circ, \tag{4}$$

$$c < (1/B^\circ)^2 K'(1/B^\circ). \tag{5}$$

Without (4), a corner solution $(\tau^*, s^*) = (\underline{\tau}, 0)$ is possible. The condition (5) avoids a situation whereby the cost to manufacture one spare unit is so high that the supplier meets the availability constraint using reliability improvement alone, i.e., $\tau^* > \underline{\tau}$ and $s^* = 0$. Note that both conditions are satisfied if the availability target is sufficiently high.

# 4 Analysis

In this section we determine the optimal contract terms and the equilibrium solutions for $\tau$ and $s$. After introducing the mathematical conventions that are used throughout the paper, we begin with the benchmark first-best case in which the customer and the supplier are assumed to be one entity and thus contracts are unnecessary. We then analyze the supplier's response to arbitrary contract terms, and finally compare the merits of different contract types.

## 4.1 Mathematical Preliminaries

The Poisson distribution of $O(\tau)$ limits our ability to obtain insights into the game-theoretic problem because of the necessity to operate with discrete variables, which handicaps our ability to obtain analytically tractable expressions. To circumvent this difficulty, we conduct an *asymptotic* analysis by treating $O(\tau)$ and $s$ as continuous variables and restricting attention to situations in which $N$ is sufficiently large and $\tau$ is sufficiently small so that

$$1/N \ll \underline{\tau} < \overline{\tau} \lesssim 0.1 \tag{6}$$

is satisfied. This condition holds, for instance, if the fleet size is fairly large. For example, $N = 200$ and $\tau = 0.1$ imply that an average of 10 products out of 200 are being repaired at any point in time and condition (6) is satisfied. In the range of $\tau$ defined by (6), we can apply the Normal approximation of $O(\tau)$ (with $E[O(\tau)] = \mathrm{Var}[O(\tau)] = 1/\tau$), which yields very accurate evaluations of $E[B \mid \tau, s] = E[\max\{O(\tau) - s, 0\}]$ and $E[I \mid \tau, s] = E[\max\{s - O(\tau), 0\}]$, the quantities of managerial interest (see Zipkin 2000, pp. 205-209 for extensive discussion of the Normal approximation in this setting and for examples). For values of $\tau$ above 0.1, the Normal approximation becomes suspect as $\Pr(O(\tau) < 0) > 0$ can no longer be ignored. While (6) is somewhat restrictive, insights generated using this assumption will not change as long as $\tau \gg 1/N$ (including $\tau$ above 0.1) even if we switch back to the Poisson distribution.

To this end, let $\phi$ and $\Phi$ be the pdf and the cdf of the standard Normal distribution. In addition, let $\overline{\Phi}(\cdot) \equiv 1 - \Phi(\cdot)$ and the loss function $L(x) \equiv \phi(x) - x\overline{\Phi}(x)$. Define

$$z \equiv (s - 1/\tau)\,/\sqrt{1/\tau} = \sqrt{\tau}s - 1/\sqrt{\tau} \tag{7}$$

and its inverse

$$s(\tau, z) \equiv 1/\tau + z/\sqrt{\tau}. \tag{8}$$

Care is needed when applying the Normal approximation because the lower bound on $z$ (and hence on $s$) is $-\infty$, not 0. To circumvent the conceptual difficulty of having negative $s$, we will regard 0 as the lower bound on $s$ and, as a consequence, we *define* the lower bound on $z$ as $\underline{z} \equiv (0 - 1/\tau)/\sqrt{1/\tau} = -1/\sqrt{\tau}$. This definition does not cause a problem because all quantities of our interest on the domain $(-\infty, -1/\sqrt{\tau})$ are insignificant in the range of $\tau$ defined in (6). Thus,

$$\sqrt{\tau}\phi\left(\underline{z}\right) \simeq 0, \qquad \Phi\left(\underline{z}\right) \simeq 0. \tag{9}$$

These approximations require us to use the following conventions

$$L(\underline{z}) = \phi(\underline{z}) - \underline{z}\overline{\Phi}(\underline{z}) = \left[\sqrt{\tau}\phi\left(\underline{z}\right) + \overline{\Phi}\left(\underline{z}\right)\right]/\sqrt{\tau} \;\simeq\; 1/\sqrt{\tau} = -\underline{z}, \tag{10}$$

$$\Phi^{-1}(0) \;\simeq\; \underline{z} = -1/\sqrt{\tau}, \tag{11}$$

in order to be consistent with (9). The expected backorder and the expected inventory on-hand are, respectively,

$$E[B \,|\, \tau, s] = L(z)/\sqrt{\tau}, \qquad\qquad E[I \,|\, \tau, s] \simeq [z + L(z)]\,/\sqrt{\tau}.$$

Note that the expression for $E[I \,|\, \tau, s]$ contains the negative domain of $s$, but again its effect is inconsequential. The following Lemma states several intuitive results with regards to these two metrics which will prove useful shortly.

**Lemma 1** *For $\tau$ satisfying (6),*

*(i)* $\partial E[B \,|\, \tau, s]/\partial s = -\overline{\Phi}(z) \leq 0, \qquad \partial E[B \,|\, \tau, s]/\partial \tau = -\phi(z)/\left(2\tau^{3/2}\right) - \overline{\Phi}(z)/\tau^2 \leq 0,$

*(ii)* $\partial E[I \,|\, \tau, s]/\partial s = \Phi(z) \geq 0, \qquad \partial E[I \,|\, \tau, s]/\partial \tau = -\phi(z)/\left(2\tau^{3/2}\right) + \Phi(z)/\tau^2 \geq 0.$

**Proof.** Omitted proofs are found in the Appendix A. ■

As we will demonstrate shortly, the backorder constraint $E[B \,|\, \tau, s] \leq B^{\circ}$ in $(\widetilde{SB})$ will be binding in most cases, and therefore the following result characterizing interdependence of the two variables $(\tau, s)$ with respect to this constraint will be useful:

**Lemma 2** *Suppose $s(\tau; B^{\circ})$ satisfying $E[B \,|\, \tau, s(\tau; B^{\circ})] = B^{\circ}$ exists for any $\tau \in [\underline{\tau}, \overline{\tau})$. Then*

$$\textit{(i)}\ \partial s(\tau; B^{\circ})/\partial \tau < 0, \qquad \textit{(ii)}\ \partial^2 s(\tau; B^{\circ})/\partial \tau^2 > 0, \qquad \textit{(iii)}\ \partial s(\tau; B^{\circ})/\partial B^{\circ} < 0.$$

Result (i) implies that $\tau$ and $s$ are *substitutes* with respect to the binding backorder constraint, i.e., for any two pairs $(\tau_1, s_1)$ and $(\tau_2, s_2)$ such that $\tau_1 < \tau_2$ that make the backorder constraint binding, we have $s_1 > s_2$. In other words, a more reliable product needs less inventory to meet the availability requirement. Furthermore, result (ii) states that this substitution effect exhibits a diminishing rate of return, e.g., an increase of $x$ hours in MTBF has a greater impact on inventory reduction when MTBF is smaller. The last result (iii) shows that, for a fixed level of reliability, less inventory is needed if the constraint is less stringent.

## 4.2 Integrated Firm: The First-Best Solution

To establish the benchmark, we first analyze the case in which the customer and the supplier are one integrated firm minimizing its total cost subject to the availability requirement (first-best or $FB$):

$$(FB) \qquad \min_{\tau \geq \underline{\tau}, s \geq 0} \; C^{FB}(\tau, s) \equiv cs + K(\tau) + hE[I \,|\, \tau, s] \text{ subject to } E[B \,|\, \tau, s] \leq B^{\circ}.$$

The solution is stated in the following Proposition. Note that $\theta$ that appears below is the Lagrangian multiplier of the availability constraint. It plays a significant role in the solution because optimal choices of $\tau$ and $s$ are interdependent via the constraint $E[B \mid \tau, s] \le B^\circ$, which will be shown to bind at optimality.

**Proposition 1** *(Integrated firm solution) Let*

$$z^{FB}(\theta) \equiv \Phi^{-1}\left(\frac{\theta - c}{\theta + h}\right), \quad \gamma^{FB}(\tau;\theta) \equiv \frac{c}{\tau^2} + \frac{\theta + h}{2\tau^{3/2}}\phi\left(z^{FB}(\theta)\right), \quad \beta^{FB}(\theta;\tau) \equiv \frac{L(z^{FB}(\theta))}{\sqrt{\tau}}.$$

*There exists a unique solution $\overline{\theta} > c$ to the equation $\beta^{FB}(\theta;\underline{\tau}) = B^\circ$. Furthermore, if*

1. *$\gamma^{FB}(\underline{\tau};\overline{\theta}) \le K'(\underline{\tau})$, the firm chooses $\tau^{FB} = \underline{\tau}$ and $s^{FB} = s(\underline{\tau}, z^{FB}(\overline{\theta})) > 0$.*
2. *$\gamma^{FB}(\underline{\tau};\overline{\theta}) > K'(\underline{\tau})$, the firm chooses $\tau^{FB} = \tau^\dagger > \underline{\tau}$ and $s^{FB} = s(\tau^\dagger, z^{FB}(\theta^\dagger)) > 0$ where $\tau^\dagger \in (\underline{\tau}, 1/B^\circ)$ and $\theta^\dagger \in (c, \overline{\theta})$ are unique solutions to the simultaneous equations*

$$\gamma^{FB}(\tau;\theta) = K'(\tau), \qquad \beta^{FB}(\theta;\tau) = B^\circ.$$

**Proof.** The Lagrangian is

$$\mathcal{L}^{FB} = -\theta B^\circ + cs + K(\tau) + hE[I \mid \tau, s] + \theta E[B \mid \tau, s]$$

where $\theta$ is the Lagrangian multiplier. Differentiating $\mathcal{L}^{FB}$ with respect to $s$, we obtain

$$\partial\mathcal{L}^{FB}/\partial s = -(\theta - c) + (\theta + h)\Phi(z), \qquad \partial^2\mathcal{L}^{FB}/\partial s^2 = (\theta + h)\sqrt{\tau}\phi(z) \ge 0.$$

It can be shown that the condition (5) excludes the case $\theta \le c$, which makes $s = 0$ optimal. Suppose $\theta > c$. By setting $\partial\mathcal{L}^{FB}/\partial s = 0$, we obtain the optimality condition for $s$ expressed in terms of $z$, i.e., $z^{FB}(\theta) = \Phi^{-1}\left((\theta - c)\,/\,(\theta + h)\right)$. Notice that this expression is independent of $\tau$. Hence, if we consider the $(\tau, z)$ space instead of the $(\tau, s)$ space after applying the $z$-transform (7), we will find that for each $\tau \ge \underline{\tau}$ the optimal $z$ is found on the horizontal line $z = z^{FB}(\theta)$. Therefore, the original two-dimensional optimization problem is reduced to a one-dimensional problem of finding optimal $\tau$ along the line $z = z^{FB}(\theta)$. Let $\widetilde{\mathcal{L}}^{FB}(\tau;\theta)$ be the reduced Lagrangian with $z = z^{FB}(\theta)$. Then

$$\widetilde{\mathcal{L}}^{FB}(\tau;\theta) = -\theta B^\circ + c/\tau + K(\tau) + (\theta + h)\,\phi(z^{FB}(\theta))/\sqrt{\tau}.$$

Note that $\lim_{\tau \to \overline{\tau}} \widetilde{\mathcal{L}}^{FB}(\tau;\theta) = \infty$. Differentiating, we obtain

$$\begin{aligned}
\partial \widetilde{\mathcal{L}}^{FB}(\tau;\theta)/\partial\tau &= K'(\tau) - c/\tau^2 - (\theta+h)\,\phi(z^{FB}(\theta))/(2\tau^{3/2}) = K'(\tau) - \gamma^{FB}(\tau;\theta), \\
\partial^2 \widetilde{\mathcal{L}}^{FB}(\tau;\theta)/\partial\tau^2 &= K''(\tau) + 2c/\tau^3 + 3(\theta+h)\phi\left(z^{FB}(\theta)\right)/(4\tau^{5/2}) > 0.
\end{aligned}$$

Suppose $\gamma^{FB}(\underline{\tau};\theta) \leq K'(\underline{\tau})$ for some fixed $\theta$. Then $\widetilde{\mathcal{L}}^{FB}(\tau;\theta)$ is increasing on $[\underline{\tau}, \overline{\tau})$ due to convexity, so it is optimal to choose $\tau(\theta) = \underline{\tau}$. In order to verify that the optimal $\theta$ indeed satisfies the assumption $\theta > c$ for this case, notice that

$$\begin{aligned}
\lim_{\theta \to c} E[B \,|\, \underline{\tau}, s(\theta)] &= (1/\sqrt{\underline{\tau}}) \lim_{\theta \to c} L(z^{FB}(\theta)) = L(\underline{z})/\sqrt{\underline{\tau}} \simeq 1/\underline{\tau} > B^\circ, \\
\lim_{\theta \to \infty} E[B \,|\, \underline{\tau}, s(\theta)] &= (1/\sqrt{\underline{\tau}}) \lim_{\theta \to \infty} L(z^{FB}(\theta)) = 0,
\end{aligned}$$

where we have used (10) for the first limit and the assumption (4) for the inequality. Since $\beta^{FB}(\theta;\underline{\tau}) = E[B \,|\, \underline{\tau}, s(\theta)]$ decreases in $\theta$, it crosses $B^\circ$ exactly once as $\theta$ moves from $c$ to infinity, and the unique solution $\overline{\theta}$ of the equation $\beta^{FB}(\theta;\underline{\tau}) = B^\circ$ satisfies $\overline{\theta} > c$. This is the upper bound on $\theta$. To see this, note that the constraint is binding as long as $\theta > 0$, i.e., $E[B \,|\, \tau, s(\theta)] = L(z^{FB}(\theta))/\sqrt{\tau} = B^\circ$. For any $\tau$ above $\underline{\tau}$, the denominator of the left-hand side is greater and hence the numerator should be greater as well since the right-hand side is a constant. Since

$$\frac{dL(z^{FB}(\theta))}{d\theta} = -\overline{\Phi}(z^{FB}(\theta))\frac{dz^{FB}(\theta)}{d\theta} = -\overline{\Phi}(z^{FB}(\theta))\frac{c+h}{(\theta+h)^2}\frac{1}{\phi\left(z^{FB}(\theta)\right)} \leq 0, \tag{12}$$

we conclude that the optimal $\theta$ for $\tau > \underline{\tau}$ should be less than $\overline{\theta}$.

Now consider the case in which $\gamma^{FB}(\underline{\tau};\theta) > K'(\underline{\tau})$. Since $\widetilde{\mathcal{L}}^{FB}(\tau;\theta)$ initially decreases on $[\underline{\tau}, \overline{\tau})$ and is convex in $\tau$, there is a unique global minimum $\tau(\theta) > \underline{\tau}$ that solves $\gamma^{FB}(\tau;\theta) = K'(\tau)$. Since optimal $\tau$ and $z$ for fixed $\theta$ are unique, optimal $s$ is also unique via (8). Let the solutions be $(\tau(\theta), s(\theta))$. It remains to find the optimal value for $\theta$, which will be shown to satisfy the assumption $\theta > c$ under conditions (4) and (5) and is determined from the binding constraint $E[B \,|\, \tau(\theta), s(\theta)] = B^\circ$. Differentiating both sides of the first-order condition $\gamma^{FB}(\tau(\theta);\theta) = K'(\tau(\theta))$ with respect to $\theta$, we obtain

$$-\frac{2c}{\tau^3}\frac{d\tau(\theta)}{d\theta} - \frac{3(\theta+h)}{4\tau^{5/2}}\phi\left(z^{FB}(\theta)\right)\frac{d\tau(\theta)}{d\theta} + \frac{1}{2\tau^{3/2}}\phi\left(z^{FB}(\theta)\right) - \frac{\theta+h}{2\tau^{3/2}}z^{FB}(\theta)\phi\left(z^{FB}(\theta)\right)\frac{dz^{FB}(\theta)}{d\theta} = K''(\tau)\frac{d\tau(\theta)}{d\theta},$$

or after collecting terms

$$\left[K''(\tau) + \frac{2c}{\tau^3} + \frac{3(\theta+h)}{4\tau^{5/2}}\phi\left(z^{FB}(\theta)\right)\right]\frac{d\tau(\theta)}{d\theta} = \frac{L(z^{FB}(\theta))}{2\tau^{3/2}}.$$

It is clear that $d\tau(\theta)/d\theta \geq 0$. Combined with (12), we find that $E[B \mid \tau(\theta), s(\theta)] = L(z^{FB}(\theta))/\sqrt{\tau(\theta)}$ is decreasing in $\theta$. Since

$$\lim_{\theta \to c} E[B \mid \tau(\theta), s(\theta)] = L(\underline{z})/\sqrt{\tau(c)} \simeq 1/\tau(c), \qquad \lim_{\theta \to \infty} E[B \mid \tau(\theta), s(\theta)] = 0,$$

we see that the necessary condition for having a unique $\theta > c$ that satisfies the binding backorder constraint $E[B \mid \tau(\theta), s(\theta)] = B^\circ$ is $1/\tau(c) > B^\circ$. Suppose, on the contrary, $\tau(c) \geq 1/B^\circ$. Letting $\theta = c$ in the first-order condition $\gamma^{FB}(\tau(\theta); \theta) = K'(\tau(\theta))$, we obtain $c \simeq \tau(c)K'(\tau(c)) \geq (1/B^\circ)K'(1/B^\circ)$, which contradicts the assumption $c < (1/B^\circ)K'(1/B^\circ)$ from (5). Therefore, the optimal solution is characterized by two equations, $\gamma^{FB}(\tau(\theta); \theta) = K'(\tau(\theta))$ and $E[B \mid \tau(\theta), s(\theta)] = B^\circ$. ■

Intuitively, if the marginal cost of improving reliability outweighs the savings in manufacturing and holding costs associated with procuring a unit of product and the benefits of the reduced expected backorders resulting from such activities ($\gamma^{FB}(\underline{\tau}; \overline{\theta}) \leq K'(\underline{\tau})$), then the integrated firm does not need to invest in reliability improvement and instead meets the availability target by acquiring an inventory of spares alone (Case 1). Otherwise reliability improvement is combined with a positive target inventory level (Case 2).

### 4.3 Supplier's Response to Contract Terms

In this subsection we start solving the problem backwards by studying the supplier's response to the general contractual form $T = w + ps - vB$, which includes both the material and the performance contracts as special cases. The supplier's optimization problem is

$$\max_{\tau \geq \underline{\tau}, s \geq 0} \pi(\tau, s) = E[T \mid \tau, s] - \psi(\tau, s) = w + (p - c)s - K(\tau) - vE[B \mid \tau, s] - \delta h E[I \mid \tau, s].$$

We use the superscript $*$ to denote the supplier's optimal response.

**Proposition 2** *(Supplier's response to contract terms) Let*

$$z_\delta(v, p) \equiv \Phi^{-1}\left(\frac{v + p - c}{v + \delta h}\right), \qquad \gamma_\delta(\tau; v, p) \equiv \frac{c - p}{\tau^2} + \frac{v + \delta h}{2\tau^{3/2}} \phi\left(z_\delta(v, p)\right), \qquad \beta_\delta(v, p; \tau) \equiv \frac{L(z_\delta(v, p))}{\sqrt{\tau}}.$$

*Define $\widehat{z} \equiv z_\delta(v, p)$. Then if*

1. *$c - p \geq v$: $\pi(\tau, s)$ is concave. If $v \leq \underline{\tau}^2 K'(\underline{\tau})$ the supplier chooses $\tau^* = \underline{\tau}$ and $s^* = 0$. Otherwise, he chooses $\tau^* = \widehat{\widehat{\tau}} > \underline{\tau}$ and $s^* = 0$ where $\widehat{\widehat{\tau}}$ is the unique solution to the equation $\tau^2 K'(\tau) = v$.*

2. $-v < p - c < \delta h$: *$\pi(\tau, s)$ is concave if $p \leq c$ and quasiconcave if $p > c$. If $\gamma_\delta(\underline{\tau}; v, p) \leq K'(\underline{\tau})$, the supplier chooses $\tau^* = \underline{\tau}$ and $s^* = s(\underline{\tau}, \widehat{z}) > 0$. Otherwise, he chooses $\tau^* = \widehat{\tau} > \underline{\tau}$ and $s^* = s(\widehat{\tau}, \widehat{z})$ where $\widehat{\tau}$ is the unique solution to the equation $\gamma_\delta(\tau; v, p) = K'(\tau)$.*

3. $p - c \geq \delta h$: *the supplier chooses $\tau^* = \underline{\tau}$ and $s^* \to \infty$.*

In Case 1, the unit price paid to the supplier for the product is below its unit cost ($p \leq c$, since $v \geq 0$) while the backorder penalty $v$ is small ($v \leq c - p$). Since the supplier earns negative revenue for each unit he acquires for the customer, he chooses not to stock any spares. The supplier can be persuaded to act otherwise if there is a strong incentive to reduce backorders (large $v$) – but it is not strong enough in this case, hence $s^* = 0$. Moreover, the supplier may be incentivized to improve reliability, i.e., $\tau^* > \underline{\tau}$, if $v$ is sufficiently large, specifically if $\underline{\tau}^2 K'(\underline{\tau}) < v \leq c - p$.

In Case 2, $v$ is large enough so that the supplier can be incentivized to increase both $\tau$ and $s$ beyond their respective minimal values ($\underline{\tau}$ and 0) if condition $\gamma_\delta(\underline{\tau}; v, p) > K'(\underline{\tau})$ is satisfied. Note that this case includes scenarios with both negative and positive margins. When $p > c$, the supplier makes positive profit for each spare product he manufactures, but does not stock too many spares since he cannot recoup his inventory ownership cost (represented by $\delta h$) with a relatively small profit margin $p - c < \delta h$.

Case 3 differs from Case 2 in that the profit margin is greater than $\delta h$. In this case the supplier tries to procure as many spare products as possible since unit-revenue outweighs inventory ownership cost. At the same time, he is not incentivized to invest in reliability improvement because more frequent failures lead to demand for more spare products, each contributing to higher profitability. Since in this case the customer cannot effectively control the supplier's inventory decision ($s^* \to \infty$), it seems unlikely that a contract with $p - c \geq \delta h$ would be offered by the customer in equilibrium (we will demonstrate this shortly).

To gain further insights into the supplier's behavior, we study the supplier's response (i.e., changes in $\tau^*$, $s^*$, $E[B \,|\, \tau^*, s^*]$, and $E[I \,|\, \tau^*, s^*]$) to changes in contract terms $p$ and $v$ as well as to changes in the ownership parameter $\delta$. First, we show an intermediate result that is used in the proof of Proposition 3. It is introduced here for its general applicability beyond the current model.[6] Note that this result is proved for an arbitrary distribution that has an increasing failure rate (IFR) property, which is shared by the Normal distribution that we use throughout the paper.

**Lemma 3** *Let $X$ be a random variable with an IFR property whose pdf $f$ is differentiable and vanishes*

[6]The quantity $\omega(x)$ defined in Lemma 3 arises frequently in game-theoretic supply chain models (for example, see Cachon 2004 and Perakis and Roels 2006).

*at both extremes of its support* $[\underline{y}, \overline{y}]$, *where* $\underline{y} = -\infty$ *and* $\overline{y} = \infty$ *are permitted. Let* $F$ *be the cdf of* $X$ *and* $\overline{F}(\cdot) \equiv 1 - F(\cdot)$. *Then* $\omega(y) \equiv f(y)E[(X-y)^+]/[\overline{F}(y)]^2 \leq 1$.

**Proof.** Note that IFR means $\frac{d}{dy}\left(f(y)/\overline{F}(y)\right) = \left(f'(y)\overline{F}(y) + [f(y)]^2\right)/[\overline{F}(y)]^2 \geq 0$, which in turn implies

$$-f'(y)/f(y) \leq f(y)/\overline{F}(y). \tag{13}$$

It can easily be shown that $\omega'(y) \geq 0$ (a similar result for a distribution exhibiting increasing *generalized* failure rate was shown in Cachon 2004). To derive the upper bound, we only need to show $m \equiv \lim_{y\to\overline{y}} \omega(y) \leq 1$. Since $m$ is of 0/0 form, we apply L'Hopital's rule to obtain:

$$\begin{aligned} m &= \lim_{y\to\overline{y}} \frac{f(y)E[(X-y)^+]}{\left[\overline{F}(y)\right]^2} = \lim_{y\to\overline{y}} \frac{f'(y)E[(X-y)^+] - f(y)\overline{F}(y)}{-2f(y)\overline{F}(y)} = \frac{1}{2} - \frac{1}{2}\lim_{y\to\overline{y}} \frac{f'(y)E[(X-y)^+]}{f(y)\overline{F}(y)} \\ &\leq \frac{1}{2} + \frac{1}{2}\lim_{y\to\overline{y}} \frac{f(y)E[(X-y)^+]}{\left[\overline{F}(y)\right]^2} = \frac{1}{2} + \frac{m}{2}, \end{aligned}$$

where we have used (13) to prove the inequality. $m \leq 1$ follows by rearranging both sides. ■

In the following Proposition, we limit consideration to Case 2 of Proposition 2 because this is the most interesting scenario, and because this is the only case which will emerge in equilibrium. We show results only for the interior solution $\tau^* = \widehat{\tau} > \underline{\tau}$ and $s^* = s(\widehat{\tau}, \widehat{z})$ but similar results for the case $\tau^* = \underline{\tau}$ and $s^* = s(\underline{\tau}, \widehat{z})$ are straightforward to obtain.

**Proposition 3** *Suppose* $-v \leq p - c < \delta h$, $\gamma_\delta(\underline{\tau}; v, p) > K'(\underline{\tau})$, *and let* $\widehat{s} \equiv s(\widehat{\tau}, \widehat{z})$.

1. $\partial\widehat{\tau}/\partial p < 0$, $\partial\widehat{s}/\partial p > 0$, $\partial E[B \mid \widehat{\tau}, \widehat{s}]/\partial p < 0$, *and* $\partial E[I \mid \widehat{\tau}, \widehat{s}]/\partial p > 0$.

2. $\partial\widehat{\tau}/\partial v > 0$, $\partial\widehat{s}/\partial v > 0$, $\partial E[B \mid \widehat{\tau}, \widehat{s}]/\partial v < 0$, *and* $\partial E[I \mid \widehat{\tau}, \widehat{s}]/\partial v > 0$.

3. $\partial\widehat{\tau}/\partial \delta > 0$, $\partial\widehat{s}/\partial \delta < 0$, *and* $\partial E[I \mid \widehat{\tau}, \widehat{s}]/\partial \delta < 0$.

We observe that higher unit price motivates the supplier to have more spare products in stock (higher $\widehat{s}$) but at the same time the supplier is indirectly incentivized to increase his demand, i.e., design a product that fails more often (reduce $\widehat{\tau}$), so that $\widehat{s}$ can be increased even further. In contrast, the backorder penalty $v$ induces the supplier to increase both $\widehat{\tau}$ and $\widehat{s}$. Thus, the Proposition provides a crucial intuition with respect to the fundamental difference between the incentive effects of the unit price ($p$) and the backorder penalty ($v$); these two contract terms induce opposite reactions from the supplier with respect to the reliability improvement decision. The impact of spare asset ownership structure is quite intuitive; with higher ownership responsibility (higher $\delta$), the supplier cares more about reliability ($\widehat{\tau}$ goes up) while reducing the stocking quantity ($\widehat{s}$ goes down).

### 4.4 The Material Contract

The material contract represents a pre-PBL contractual relationship and, as the following Proposition shows, it is not effective in incentivizing the supplier to improve reliability beyond the default level $\underline{\tau}$.

**Proposition 4** *(Optimal material contract) With $T = w + ps$ and $\delta > 0$, the customer chooses*

$$\overline{p} = c + \delta h(\overline{\theta} - c)/(\overline{\theta} + h) \tag{14}$$

*where the unique value $\overline{\theta} > c$ satisfying the equation $\beta^{FB}(\theta; \underline{\tau}) = B^{\circ}$ was found in Proposition 1. As a response, the supplier will choose $\tau^M = \underline{\tau}$ and $s^M = s(\underline{\tau}, z_\delta(0, \overline{p})) = s(\underline{\tau}, z^{FB}(\overline{\theta}))$. On the other hand, the optimal solution does not exist if $\delta = 0$, i.e., there is no $p$ that guarantees satisfaction of the backorder constraint while incurring finite cost.*

**Proof.** We begin our proof by showing that the supplier will not invest in reliability improvements, i.e., $\tau^* = \underline{\tau}$, under the material contract. Consider each case in Proposition 2 with $v = 0$. In Case 1, $\tau^* = \underline{\tau}$ since $v \leq \underline{\tau}^2 K'(\underline{\tau})$ is satisfied. In Case 2, we have $p > c$ with $v = 0$. Observe from $\pi'(\tau) = -K'(\tau) + \gamma_\delta(\tau; 0, p)$ that

$$\begin{aligned}
\lim_{p \to c} \pi'(\tau) &= -K'(\tau) + \delta h \phi(\underline{z})/(2\tau^{3/2}) \simeq -K'(\tau) < 0, \\
\partial \pi'(\tau)/\partial p &= -1/\tau^2 - \widehat{z}/(2\tau^{3/2}) = -1/(2\tau^2) - \widehat{s}/2\tau < 0
\end{aligned}$$

for all $\tau \in [\underline{\tau}, \overline{\tau})$. Therefore $\pi'(\tau) < 0$, which implies $\tau^* = \underline{\tau}$. In Case 3, $\tau^* = \underline{\tau}$ follows automatically.

We now turn to the customer's problem. Assume $\delta > 0$. From Proposition 2 notice that $p$ will be chosen so that Case 3 will not result in at equilibrium, since in that case the contract will force the customer to make an infinite payment to the supplier who chooses $s \to \infty$ as long as $p > c$. Similarly, $v = 0$ rules out Case 1 since the supplier will choose $\tau^* = \underline{\tau}$ and $s^* = 0$ which will violate the backorder constraint due to the assumption (4) with $E[B \,|\, \underline{\tau}, 0] \simeq 1/\underline{\tau}$. Only Case 2 remains in which $0 < p - c \leq \delta h$. Recall that the supplier chooses $\tau^* = \underline{\tau}$ and that $\pi'(\underline{\tau}) = -K'(\underline{\tau}) + \gamma_\delta(\underline{\tau}; 0, p) \leq 0$, implying $s^* = s(\underline{\tau}, \widehat{z})$ where $\widehat{z} = \Phi^{-1}\left((p - c)\,/\,(\delta h)\right)$. Using these values, the Lagrangian for the customer's problem can be expressed as follows:

$$\mathcal{L}^M = -\theta B^{\circ} + K(\underline{\tau}) + \frac{c}{\underline{\tau}} + \frac{\theta + h}{\sqrt{\underline{\tau}}} \phi(\widehat{z}) + \frac{(\theta + h)\widehat{z}}{\sqrt{\underline{\tau}}} \left( \frac{p - c}{\delta h} - \frac{\theta - c}{\theta + h} \right).$$

Differentiating, we obtain

$$\frac{\partial \mathcal{L}^M}{\partial p} = \frac{\theta + h}{\delta h \sqrt{\underline{\tau}} \phi(\widehat{z})} \left( \frac{p - c}{\delta h} - \frac{\theta - c}{\theta + h} \right).$$

If $\theta \le c$, $\partial \mathcal{L}^M / \partial p > 0$ and it is best for the customer to have $p \to c$, but we have already ruled out $p \le c$ so it cannot be the optimal solution. Suppose $\theta > c$. Then the optimal $p$ is given by $p(\theta) = c + ((\theta - c) / (\theta + h)) \, \delta h$ and

$$\left. \frac{\partial^2 \mathcal{L}^M}{\partial p^2} \right|_{p = p(\theta)} = \frac{(\theta + h)\widehat{z}}{(\delta h)^2 \sqrt{\underline{\tau}} \phi(\widehat{z})} \left( \frac{p(\theta) - c}{\delta h} - \frac{\theta - c}{\theta + h} \right) + \frac{\theta + h}{(\delta h)^2 \sqrt{\underline{\tau}} \phi(\widehat{z})} = \frac{\theta + h}{(\delta h)^2 \sqrt{\underline{\tau}} \phi(\widehat{z})} > 0,$$

which implies that the solution is unique. The optimal $\theta$ is determined from the binding backorder constraint $E[B \,|\, \underline{\tau}, s(\underline{\tau}, \widehat{z})] = E[B \,|\, \underline{\tau}, s(\theta)] = B^\circ$ (it binds since $\theta > c$ by assumption). Its solution $\overline{\theta}$ was already specified in Proposition 1 and was found to satisfy $\overline{\theta} > c$, thus confirming the earlier assumption. Finally, assume $\delta = 0$. Due to the reasons noted above, $p < c$ and $p > c$ are not optimal. Suppose $p = c$. Then the supplier's expected profit (1) is independent of $s$ and thus he may choose any $s \ge 0$. Hence, the backorder constraint may or may not be satisfied. ■

With $v = 0$, the unit price $p$ is the only mechanism by which the customer can control the supplier's behavior. As is evident from the Proposition, $p$ only serves as an incentive to increase the stocking level and not the reliability.[7] Anticipating this behavior, the best the customer can do in order to meet the backorder constraint (which binds at optimality) is to choose $p$ appropriately so that it induces the supplier to pick a large enough stocking quantity of spares $s$. The resulting equilibrium solution under the optimal price $\overline{p}$ leads to lower product reliability and higher inventory than under the first-best solution. The only situation in which the first-best solution is achieved occurs when the marginal cost of reliability improvement is sufficiently high ($\gamma^{FB}(\underline{\tau}; \overline{\theta}) \le K'(\underline{\tau})$; see Proposition 1), in which case the integrated firm would also find it optimal to have minimal product reliability, $\tau^{FB} = \underline{\tau}$.

Further, we see that, at equilibrium, the optimal price is chosen such that the margin is positive, i.e., $\overline{p} - c > 0$, given that $\delta > 0$ (otherwise the supplier would not meet the availability requirement). With the positive margin, the supplier chooses a finite $s > 0$ (recall from Proposition 3 that $s^*$ increases with $p$) because of holding costs. We see from (14) that the optimal $p$ is an increasing function of $\delta$; with a larger share of the supplier's asset ownership (causing higher holding cost), a higher margin is

[7]However, the supplier might have incentives to invest into reliability improvement if we include explicit repair cost in the supplier's cost structure (see Section 5 for further discussion). Similarly, if we were to adopt a longer-term perspective beyond the life of the current contract, then a supplier would have an incentive to improve reliability since it would affect the likelihood that his product would be selected initially, or that the support contract would be renewed. Modeling such a long-term relationship is beyond the scope of this paper but will be considered in our future work.

needed to persuade the supplier to select the desired level of inventory. This argument, of course, is invalid if $\delta = 0$ (full customer ownership), since the supplier does not incur holding costs. It turns out that there is no equilibrium solution in this case: if $p > c$, the supplier chooses $s \to \infty$, resulting in an infinite compensation amount under the contract; if $p = c$, the supplier is indifferent towards inventory $s$ and his behavior becomes unpredictable. Therefore, our model suggests that the material contract should not be implemented under full customer asset ownership. Interestingly, the prevalent practice before the advent of PBL was precisely to let the customer own spare assets. We conjecture that the inherent shortcoming that we have identified using our model contributed to often adversarial relationships between the DoD and the contractors under material contracts and may have led to the eventual shift towards the PBL relationship.

We also find that the optimal material contract always leads to the same supplier's decisions, $\tau^M = \underline{\tau}$ and $s^M = s(\underline{\tau}, z^{FB}(\overline{\theta}))$, which are independent of $\delta$ as long as $\delta > 0$. Thus, efficiency loss due to inability to motivate reliability improvement is uniform across the spectrum of ownership structures and the customer is indifferent to the ownership structure under the material contract, insofar as $\delta \neq 0$.

To summarize, the material contract does not incentivize the supplier to improve reliability and therefore leads to excess inventory in the system. The intuition behind this observation is as follows. Since the supplier generates revenue by selling spare parts and repair services, his business improves when more frequent needs for them arise. Unfortunately, these needs are triggered by product failures, which have a negative impact on product availability and consequently on the customer's ability to generate value through product use. Hence, the supplier's interest in making profit in the after-sales market is in direct conflict with the customer's goal of achieving product readiness.

### 4.5 The Performance Contract

We now turn to the PBL contract. We first note that the implicit nature of the solution to the supplier's sub-problem makes analysis of second-order properties intractable; thus we are unable to show that the customer's problem is well-behaved (quasi-convex). Fortunately, the monotonicity results in Proposition 3 allow us to circumvent this difficulty because the customer's objective $\widetilde{C}^{SB} = c\widehat{s} + K(\widehat{\tau}) + hE[I \mid \widehat{\tau}, \widehat{s}]$ is monotone increasing in $v$, which, combined with the property that $\partial E[B \mid \widehat{\tau}, \widehat{s}]/\partial v < 0$, implies that the optimal $v$ is determined from the binding constraint $E[B \mid \widehat{\tau}, \widehat{s}] = B^\circ$ ($\widetilde{C}^{SB}$ is minimized at the boundary).

**Proposition 5** *(Optimal performance contract) Let* $z^P_\delta(v) \equiv z_\delta(v, 0)$, $\gamma^P_\delta(\tau; v) \equiv \gamma_\delta(\tau; v, 0)$, $\beta^P_\delta(v; \tau) \equiv$

$\beta_\delta(v, 0; \tau)$, *i.e.*,

$$z_\delta^P(v) = \Phi^{-1}\left(\frac{v-c}{v+\delta h}\right), \qquad \gamma_\delta^P(\tau; v) = \frac{c}{\tau^2} + \frac{(v+\delta h)}{2\tau^{3/2}}\phi\left(z_\delta^P(v)\right), \qquad \beta_\delta^P(v;\tau) = \frac{L(z_\delta^P(v))}{\sqrt{\tau}}.$$

*Define*

$$\overline{v} \equiv \left(\frac{c+\delta h}{c+h}\right)\overline{\theta} + \frac{(1-\delta)ch}{c+h}, \tag{15}$$

*where the unique value $\overline{\theta} > c$ that solves the equation $\beta^{FB}(\theta; \underline{\tau}) = B^\circ$ was found in Proposition 1. Then:*

1. *If $\gamma_\delta^P(\underline{\tau}; \overline{v}) \leq K'(\underline{\tau})$, the customer chooses $v^P = \overline{v}$. As a response, the supplier will choose $\tau^P = \underline{\tau}$ and $s^P = s(\underline{\tau}, z_\delta^P(\overline{v})) = s(\underline{\tau}, z^{FB}(\overline{\theta}))$.*

2. *If $\gamma_\delta^P(\underline{\tau}; \overline{v}) > K'(\underline{\tau})$, there exists a unique solution $\tau_\delta^P(v) > \underline{\tau}$ to the equation $\gamma_\delta^P(\tau; v) = K'(\tau)$ for any $v \in (c, \overline{v})$. It represents the supplier's optimal choice of $\tau$ in response to $v$. The customer chooses $v^P = v^\ddagger \in (c, \overline{v})$, which is the unique solution to equation $\beta_\delta^P(v; \tau_\delta^P(v)) = B^\circ$. As a response, the supplier will choose $\tau^P \equiv \tau_\delta^P(v^\ddagger) > \underline{\tau}$ and $s^P \equiv s(\tau_\delta^P(v^\ddagger), z_\delta^P(v^\ddagger))$.*

A reader will notice that the solution structure resembles that of the first-best solution (Proposition 1) but with some notable departures. First, the condition for having an interior solution ($\tau^P > \underline{\tau}$) is more restrictive. To see this, it suffices to compare $\gamma_\delta^P(\underline{\tau}; \overline{v}) > K'(\underline{\tau})$ and $\gamma^{FB}(\underline{\tau}; \overline{\theta}) > K'(\underline{\tau})$ and it is apparent that

$$\gamma_\delta^P(\underline{\tau}; \overline{v}) = \frac{c}{\underline{\tau}^2} + \frac{\overline{v}+\delta h}{2\underline{\tau}^{3/2}}\phi(z_\delta^P(\overline{v})) \leq \frac{c}{\underline{\tau}^2} + \frac{\overline{\theta}+\delta h}{2\underline{\tau}^{3/2}}\phi(z_\delta^P(\overline{\theta})) \leq \frac{c}{\underline{\tau}^2} + \frac{\overline{\theta}+h}{2\underline{\tau}^{3/2}}\phi(z_1^P(\overline{\theta})) = \gamma^{FB}(\underline{\tau}; \overline{\theta}),$$

where the first inequality follows from $\overline{v} \leq \overline{\theta}$ for $\delta \in [0, 1]$, as can be verified from (15), and the second inequality follows from Lemma 4 found in the Appendix. Hence, if $\tau^P$ is the interior solution under the performance contract, then $\tau^{FB}$ is the interior solution under the first-best contract. In other words, it is possible to have $\tau^{FB} > \underline{\tau}$ and $\tau^P = \underline{\tau}$, but not vice versa. Clearly, the difference between $\tau^{FB}$ and $\tau^P$ disappears when $\delta = 1$ since $\overline{v} = \overline{\theta}$ in such a case, but it becomes larger as $\delta$ decreases. Intuitively, with a higher degree of customer asset ownership, it becomes more difficult to induce the supplier to improve reliability under the performance contract. Notice also that the equilibrium interior solutions $\tau^P$ and $s^P$ are, in general, different from their first-best counterparts, $\tau^\dagger$ and $s(\tau^\dagger, z^{FB}(\theta^\dagger))$. The following result summarizes the comparison of these two solutions.

**Proposition 6** *At the equilibrium,*

1. $\tau^P = \tau^{FB}$ and $s^P = s^{FB}$ if $\gamma^{FB}(\underline{\tau}; \overline{\theta}) \leq K'(\underline{\tau})$.

2. $\tau^P \leq \tau^{FB}$ and $s^P \geq s^{FB}$ if $\gamma^{FB}(\underline{\tau}; \overline{\theta}) > K'(\underline{\tau})$. The equalities hold if and only if $\delta = 1$. Furthermore, $\tau^P(\delta_1) < \tau^P(\delta_2)$ and $s^P(\delta_1) > s^P(\delta_2)$ for $0 \leq \delta_1 < \delta_2 \leq 1$.

The last Proposition demonstrates how the first-best solution can be attained using the performance contract. In particular, the only ownership structure such that $\tau^P = \tau^{FB}$ and $s^P = s^{FB}$ regardless of relative costs (i.e., whether $\gamma^{FB}(\underline{\tau}; \overline{\theta}) \leq K'(\underline{\tau})$ or $\gamma^{FB}(\underline{\tau}; \overline{\theta}) > K'(\underline{\tau})$) involves complete supplier ownership of spare assets, $\delta = 1$. When endowed with full ownership cost responsibility, the supplier absorbs the entire cost of the integrated supply chain and becomes a full residual claimant of the stochastic financial outcome of the trade, i.e., he assumes all financial risks associated with realizations of $B$ and $I$ (provided that the backorder penalty is set equal to the first-best Lagrangian multiplier, $v = \theta^{\dagger}$). Since the supplier is risk-neutral, he accepts the risk associated with performance uncertainty as long as the ex-ante participation constraint (IR) is satisfied (see Laffont and Martimort 2002, p. 147).[8] When $\delta < 1$, on the other hand, the first-best solution is only achieved in the special Case 1, in which the condition $\gamma^{FB}(\underline{\tau}; \overline{\theta}) \leq K'(\underline{\tau})$ is satisfied. In this case, high marginal cost associated with reliability improvement leads the supplier to choose $\tau^* = \underline{\tau}$, so $v$ only serves as an incentive to increase $s$. The first-best solution is achieved in this case because the impact of $v$ is limited to one decision variable and there is no efficiency loss arising from a tradeoff between $\tau$ and $s$.

On the other hand, in Case 2, reliability improvement is a viable option for the supplier since its cost is relatively low. However, the degree to which the supplier improves reliability depends on his share of holding cost. With smaller responsibility (smaller $\delta$), it is less expensive for the supplier to hold spares, and therefore he is more willing to accept an increase in holding cost that he trades off against the cost of reliability improvement. In other words, the mix of $\tau^*$ and $s^*$ changes with decreasing $\delta$ in such a way that the supplier invests less in reliability and more in inventory. Anticipating this, the customer chooses $v$ appropriately to balance reliability and inventory in order to meet the availability requirement. It turns out that the equilibrium reliability in this case is never greater than $\tau^{FB}$, and, as a result, there is more inventory in the supply chain compared to the first-best solution. The efficiency loss occurs because (1) the supplier is only a partial residual claimant and (2) the customer cannot fully control the supplier's opportunistic behavior with a single incentive term $v$ when a tradeoff between $\tau$ and $s$ exists. The degree of the efficiency loss depends on $\delta$: the smaller the $\delta$, the further the solution is from first-best.

[8] If the supplier were risk averse, inefficiency would arise because he would demand the "risk premium". The customer would then have to set $v < \theta^{\dagger}$ in order to insure the supplier's participation in the trade (see Kim et al. 2007).

| Cases | $\tau$ | $s$ |
|---|---|---|
| $0 \leq \delta < 1$ | | |
| $\gamma^{FB}(\underline{\tau}; \overline{\theta}) \leq K'(\underline{\tau})$ | $\tau^M = \tau^P = \tau^{FB}$ | $s^M = s^P = s^{FB}$ |
| $\gamma_\delta^P(\underline{\tau}; \overline{v}) \leq K'(\underline{\tau}) < \gamma^{FB}(\underline{\tau}; \overline{\theta})$ | $\tau^M = \tau^P < \tau^{FB}$ | $s^M = s^P > s^{FB}$ |
| $\gamma_\delta^P(\underline{\tau}; \overline{v}) > K'(\underline{\tau})$ | $\tau^M < \tau^P < \tau^{FB}$ | $s^M > s^P > s^{FB}$ |
| $\delta = 1$ | | |
| $\gamma^{FB}(\underline{\tau}; \overline{\theta}) \leq K'(\underline{\tau})$ | $\tau^M = \tau^P = \tau^{FB}$ | $s^M = s^P = s^{FB}$ |
| $\gamma^{FB}(\underline{\tau}; \overline{\theta}) > K'(\underline{\tau})$ | $\tau^M < \tau^P = \tau^{FB}$ | $s^M > s^P = s^{FB}$ |

Table 1: Comparison of equilibrium $\tau$ and $s$.

### 4.6 Comparisons of the Material and Performance Contracts

For convenience, we summarize the comparisons of equilibrium decisions by the supplier ($\tau$ and $s$) in Table 1. As is evident from the Table, the most important distinction between material and performance contracts is that the latter incentivizes the supplier's voluntary reliability improvement effort while the former does not. Clearly, the difference between the two contracts will vanish if the cost of reliability improvement is high enough so it is not economically feasible to increase $\tau$ beyond the default level $\underline{\tau}$. Thus, if the technology underlying the product is such that improving its reliability is too costly or impractical, the material contract performs just as well as does the performance contract. However, if product modification is a viable option for the supplier in the sense that improved reliability is worth the cost, then a performance contract is the preferred option.

The distortions of equilibrium solutions for $\tau$ and $s$ are due to misalignment of incentives borne by the limited nature of the contracts. It is convenient to represent the distortions as follows.[9] We can construct a vector $\mathbf{y}^j = (\tau^j - \underline{\tau})\mathbf{e}_\tau + s^j \mathbf{e}_s$ where $\mathbf{e}_\tau$ and $\mathbf{e}_s$ are orthogonal unit vectors in the direction of $\tau$- and $s$-axes respectively, with $j \in \{FB, M, P\}$. One way to measure the distortion is by an angle between $\mathbf{y}^{FB}$ and either $\mathbf{y}^M$ or $\mathbf{y}^P$. Denote this angle by $\varphi_i$, $i \in \{M, P\}$. From

$$(\tau^i - \underline{\tau})(\tau^{FB} - \underline{\tau}) + (s^i)(s^{FB}) = \mathbf{y}^i \cdot \mathbf{y}^{FB} = \left|\mathbf{y}^i\right| \left|\mathbf{y}^{FB}\right| \cos \varphi_j,$$

we obtain

$$\cos \varphi_i = \frac{(\tau^i - \underline{\tau})(\tau^{FB} - \underline{\tau}) + (s^i)(s^{FB})}{\sqrt{(\tau^i - \underline{\tau})^2 + (s^i)^2}\sqrt{(\tau^{FB} - \underline{\tau})^2 + (s^{FB})^2}}. \tag{16}$$

As discussed above, the distortion is greater with the material contract than with the performance contract, i.e., $\varphi_M > \varphi_P$. This is illustrated in Figure 1.

[9] This representation is inspired by a geometric interpretation of multitask principal-agent models in Baker (2002).

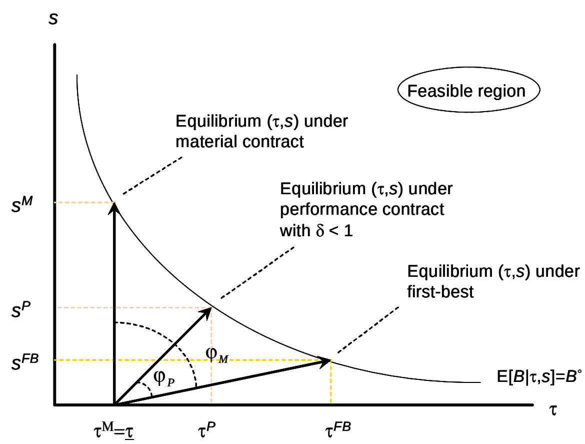


Figure 1: Equilibrium solution points in the (τ,*s*) space.

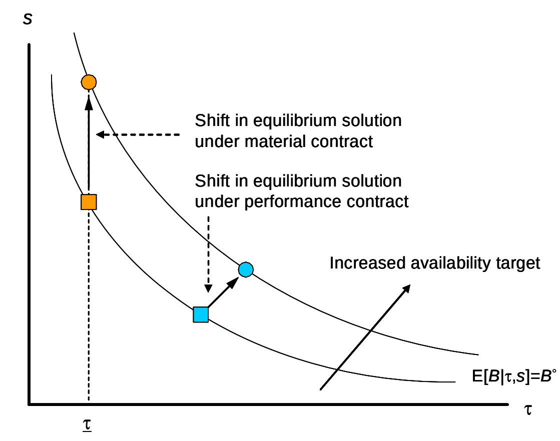


Figure 2: Equilibrium solutions and the availability target.

Finally, we analyze changes in contract terms and equilibrium decisions by the supplier in response to changes in the availability target.

**Proposition 7** *Let $p^M$ and $v^P$ be the optimal unit price and penalty rate under the material and performance contracts, respectively.*

*(i) Under the material contract, $\partial p^M/\partial B^\circ < 0$. Furthermore, $\partial \tau^M/\partial B^\circ = 0$ and $\partial s^M/\partial B^\circ < 0$.*

*(ii) Under the performance contract, $\partial v^P/\partial B^\circ < 0$. Furthermore, $\partial \tau^P/\partial B^\circ \leq 0$ and $\partial s^P/\partial B^\circ < 0$, where the equality holds only if $\tau^P = \underline{\tau}$.*

Intuitively, incentive terms ($p$ and $v$) decrease as the availability target decreases ($B^\circ$ increases). Since the material contract does not induce the supplier to improve reliability, lower target availability also results in less inventory. Under the performance contract, on the other hand, both reliability and inventory decrease to meet the new (lower) availability target. These effects are illustrated in Figure 2.

# 5 Discussion of Modeling Assumptions

In order to highlight the main issues of interest, we have made several simplifying assumptions throughout the paper. In this section, we outline possible consequences of relaxing some of these assumptions. First, in the paper, we treat each spare product as an integrated "kit" instead of an assembled product consisting of many different parts. In reality, contracts are often enforced at the subsystem or part level (e.g., a PBL contract is often awarded for an engine or an avionics subsystem). An explicit model of subsystems raises the issue of how to break down the availability requirement for the final product into the requirements for each component. The solution is well-known in the literature and is computational in nature ("the greedy algorithm", see Sherbrooke 2004). Understandably, a game-theoretic analysis of the setup in which there are multiple suppliers of an assembly system is quite difficult and insights are limited (see Kim et al. 2007). The complexity is exacerbated further when product reliability is set endogenously. In this paper we ignored the multi-indentured structure of the supply chain because of the apparent complexity it would introduce into our model, but it may prove to be a promising direction for future research.

Furthermore, we have chosen to omit the variable repair cost from the supplier's cost function $\psi(\tau, s)$. In practice, product repairs may incur both fixed and variable costs, and the latter is typically proportional to the failure rate $\lambda$. We believe that in practice the fixed cost associated with setting up the repair facility, buying equipment, and paying salaried workers is orders of magnitude higher than any variable costs associated with repairs, making our simplification quite reasonable. In cases when the variable cost of repairs cannot be ignored, most of our insights remain unchanged except that under the material contract the supplier may choose to improve reliability, i.e., $\tau^* > \underline{\tau}$. This is because the variable repair cost motivates the supplier to reduce the number of costly product failures. While this is an interesting observation, introduction of the variable repair cost poses analytical difficulty described above: the customer's objective $\widetilde{C}^{SB}$ may not be monotone in $p$ (as opposed to $v$ above) and the availability constraint may not bind. Again, we believe that such situations rarely, if ever, arise in practice and hence our modeling choice is reasonable.

Finally, after analyzing the material and performance contracts, a legitimate question to ask is whether the hybrid contract $T = w + ps - vB$ is preferred over them. It turns out that the answer to this question is negative as the following Proposition shows: mixing the material and the performance contract terms does not add value to a simpler performance contract.[10] In fact, the material contract term $p$ tends to deteriorate supply chain performance because the profit margin for each unit motivates the supplier to increase inventory and reduce reliability.

**Proposition 8** *Among the set of contracts of the form $T = w + ps - vB$, the performance contract with $p = 0$ is optimal for all $0 \le \delta \le 1$.*

**Proof.** Since Cases 1 and 3 in Proposition 2 never lead to equilibrium solutions, the customer will choose $p$ and $v$ such that $-v \le p - c < \delta h$. For expositional clarity, we only consider the ranges of $p$ and $v$ that induce the supplier to choose interior solutions $\tau^* = \widehat{\tau} > \underline{\tau}$ (incorporating $\tau^* = \underline{\tau}$ is straightforward). Then Proposition 3 applies, and, as a result, $\widetilde{C}^{SB} = c\widehat{s} + K(\widehat{\tau}) + hE[I \,|\, \widehat{\tau}, \widehat{s}]$ is

[10] In contrast, Kim et al. (2007) shows that the optimal contract in the presence of risk aversion has a cost-sharing contract term in addition to the performance term. Cost-sharing does not play a role in the current model since the supply chain members are assumed to be risk neutral. Moreover, reliability improvement is not considered in Kim et al. (2007).

increasing in $v$ and $\partial E[B \mid \widehat{\tau}, \widehat{s}]/\partial v < 0$. Thus one can always find $v$ on $[c-p, \infty)$ for any fixed $p$ that makes the backorder constraint binding. Denote $\widehat{\tau}(p,v)$ and $\widehat{s}(p,v)$ as the supplier's optimal choices as functions of $p$ and $v$, and $E[B \mid p,v] \equiv E[B \mid \widehat{\tau}(p,v), \widehat{s}(p,v)]$. From Proposition 6 we know that $\tau^P = \widehat{\tau}(0,v^P) \leq \tau^{FB}$ and $s^P = \widehat{s}(0,v^P) \geq s^{FB}$, where $v^P > c$ satisfies $E[B \mid 0, v^P] = B^\circ$. Consider increasing $p$ by $\Delta p \in [0, c+\delta h)$. Since $\partial E[B \mid p,v]/\partial p < 0$, we end up with $E[B \mid \Delta p, v^P] < B^\circ$. To make the constraint binding, the customer should decrease $v$ by $\Delta v$ such that $E[B \mid \Delta p, v^P - \Delta v] = B^\circ$, since $\partial E[B \mid p,v]/\partial v < 0$. But notice that we have $\widehat{\tau}(\Delta p, v^P - \Delta v) < \widehat{\tau}(0, v^P) = \tau^P \leq \tau^{FB}$, since $\partial\widehat{\tau}/\partial p < 0$ and $\partial\widehat{\tau}/\partial v > 0$. By Lemma 2 we also have $\widehat{s}(\Delta p, v^P - \Delta v) > \widehat{s}(0, v^P) = s^P \geq s^{FB}$. In other words, we move further away from the first-best equilibrium solution as we increase $p$ beyond 0.
■

## 6 Conclusion

In this paper we propose a stylized economic model to evaluate the trade-off between investing in reliability improvements and stocking spares under two contracts that are commonly observed in after-sales support for complex equipment. The motivation for our research comes from the new contracting strategy, Performance Based Logistics, which is gaining wide acceptance in the aerospace and defense industries today. Performance contracts are designed to replace more conventional material contracts in an attempt to better align the incentives of customers and suppliers. However, even several years after the PBL strategy has been announced, significant confusion surrounds the implementation of PBL contracts. Our conversations with many suppliers to the Department of Defense indicate that they face difficulties estimating the costs and benefits of PBL contracts, whereas this was relatively straightforward under material contracts, when suppliers were paid per spare part. At the same time, early implementations of PBL relationships seem to indicate their superiority to material contracts.

Our model suggests that material contracts are not as effective as PBL contracts in incentivizing suppliers to invest into reliability improvements. Instead, under a material contract, suppliers tend to meet the availability target by increasing the inventory of spares. Under a PBL contract, on the other hand, the supplier achieves the availability target by both improving reliability and stocking inventory. When improving reliability is impractical, both contracts can be used to achieve the first-best solution. However, for a majority of reasonable situations, both contracts result in inefficiencies manifested in less reliable products and more inventory than the first-best solution prescribes. The inefficiency of the material contract is typically larger, i.e., it results in even less reliability and more inventory than the PBL contract.

Moreover, we found that the efficiency of the PBL contract depends heavily on the asset ownership structure, while the efficiency of the material contract is largely independent of it. The predominant PBL practice we observe today is for the customer to own spare assets. However, our model predicts that this arrangement is suboptimal because the supplier does not fully internalize the operating (holding) costs. Instead, our analysis advocates giving suppliers full ownership responsibility and thereby transforming them into total service providers. When this is done, the PBL contract does achieve the first-best solution, thus coordinating the supply chain. Naturally, practical implementation of our policy recommendation will not be straightforward since many military customers believe that asset ownership protects them from mismanagement by the supplier and endows them with more control over fleet availability.

While focusing on the tradeoff between investment in reliability and spare parts management, we ignored several important aspects of the contractual relationships prevalent in the defense and aerospace industries. Perhaps the most important aspect is the long-term nature of most such relationships, which is partially driven by the fact that there is a single monopolistic customer and very few potential system suppliers. We found that, in practice, in addition to explicit contractual terms (such as availability), the customer often evaluates suppliers based on a variety of other metrics which are used to award contract renewals. A natural modeling framework for such practice is a repeated game, which introduces additional methodological challenges. We are currently pursuing this avenue of research in a separate paper. Furthermore, while we believe that reliability improvement is an important consideration for suppliers, an argument can be made that suppliers can also affect repair lead times. We hope that future work will explore this avenue of research. Last but not least, practitioners we communicated with expressed interest in formalizing insights from stylized economic models into a decision-support tool that can aid the negotiation process between customers and suppliers. Clearly, this is an important and difficult problem that requires an explicit model of the multi-echelon, multi-indentured structure of the military supply chain, a direction we wish to pursue in the future.

## References

[1] Baiman, S., S. Netessine and R. Saouma. 2007. The Interaction of Management Control and Job Design: Incentive Compensation and The Choice of Inventory Buffer. Working paper, University of Pennsylvania.

[2] Baker, G. 2002. Distortion and Risk in Optimal Incentive Contracts. *Journal of Human Resources*, 37, 728-51.

[3] Bijesse, J., M. McCluskey, and L. Sodano. 2002. Service Lifecycle Management (Part 1): The Approaches and Technologies to Build Sustainable Competitive Advantage for Service. *AMR Research Report*.

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

[5] Cachon, G. P. 2004. The Allocation of Inventory Risk in a Supply Chain: Push, Pull, and Advance-Purchase Discount Contracts. *Management Science*, 50(2), 222-38.

[6] Cohen, M. A., P. R. Kleindorfer, and H. L. Lee. 1989. Near-Optimal Service Constrained Stocking Policies for Spare Parts. *Operations Research*, 37(1), 104-117.

[7] Cohen, M. A., N. Agrawal, and V. Agrawal. 2006. Winning in the Aftermarket. *Harvard Business Review*, 84(5), 129-38.

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

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

[10] Deshpande, V., M. A. Cohen and K. Donohue. 2003. An Empirical Study of Service Differentiation for Weapon System Service Parts. *Operations Research*, 51(4), 518-530.

[11] Deshpande, V., A. V. Iyer and R. Cho. 2006. Efficient Supply Chain Management at the U.S. Coast Guard Using Part-Age Dependent Supply Replenishment Policies. *Operations Research*, 54(6), 1028-1040.

[12] Department of Defense. 2003. Department of Defense Directive 5000.1. http://www.dtic.mil/ whs/ directives/ corres/ html/ 50001.htm.

[13] Department of Defense. 2005. DoD Guide for Achieving Reliability, Availability, and Maintainability. http://www.dote.osd.mil/ reports/ RAMGuide.pdf.

[14] Geary, S. 2006. Ready for Combat. *DC Velocity*, 4(7), 75-80.

[15] Gibbons, R. 2005. Incentives Between Firms (and Within). *Management Science*, 51(1), 2-17.

[16] Government Accountability Office. 2003. Best Practices: Setting Requirements Differently Could Reduce Weapon Systems' Total Ownership Costs. GAO-03-57.

[17] Hasija, S., E. Pinker and R. Shumsky. 2006. Call Center Outsourcing Contracts Under Information Asymmetry. Working Paper, University of Rochester.

[18] 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.

[19] Kang, K., K. H. Doerr, U. Apte and M. Boudreau. 2005. Decisions Support Models for Valuing Improvements in Component Reliability and Maintenance. Working paper, Naval Postgraduate School.

[20] Kim, S.-H., M. A. Cohen, and S. Netessine. 2007. Performance Contracting in After-Sales Service Supply Chains. Forthcoming, *Management Science*.

[21] Laffont, J.-J. and D. Martimort. 2002. *The Theory of Incentives: The Principal-Agent Model.* Princeton University Press.

[22] Lu, X. L., J. A. Van Mieghem, and R. C. Savaskan. 2006. Incentives for Quality through Endogenous Routing. Working paper, Northwestern University.

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

[24] Naor, P. 1969. The Regulation of Queue Size by Levying Tolls. *Econometrica*, 37(1), 15-24.

[25] Perakis, G. and G. Roels. 2006. The Price of Anarchy in Supply Chains: Quantifying the Efficiency of Price-Only Contracts. Forthcoming, *Management Science*.

[26] 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.

[27] Ren, Z. J. and Y.-P. Zhou. 2006. Call Center Outsourcing: Coordinating Staffing Level and Service Quality. Forthcoming, *Management Science*.

[28] Rogerson, W. P. 1994. Economic Incentives and the Defense Procurement Process. *Journal of Economic Perspectives*, 8(4), 65-90.

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

[30] Sherbrooke, C. C. 2004. *Optimal Inventory Modeling of Systems: Multi-Echelon Techniques.* Springer.

[31] Stidham, S. 1985. Optimal Control of Admission to a Queueing System. *IEEE Trans. Auto. Control*, 30(8), 705-13.

[32] Wynne, M. W. 2004. Performance Based Logistics: Purchasing Using Performance Based Criteria. The Under Secretary of Defense, Memorandum for Secretaries of the Military Departments, August 16.

[33] Zipkin, P. H. 2000. *Foundations of Inventory Management.* McGraw-Hill.

## Appendix A: Proofs

**Proof of Lemma 1.** Only $\partial E[I \,|\, \tau, s]/\partial\tau \geq 0$ needs to be shown. Fix $\tau$ and let $\iota(s) \equiv \partial E[I \,|\, \tau, s]/\partial\tau$. It can be shown that $\iota(0) \geq 0$ for $\tau \lesssim 0.1$ since $\iota'(s) = z\phi(z)/\left(2\tau\right) + \phi(z)/\tau^{3/2} \geq 0$ for all $s \geq 0$ when (6) is satisfied. ■

**Proof of Lemma 2.** We suppress arguments in $s(\tau; B^\circ)$ for notational convenience. Differentiating both sides of the binding constraint $E[B \,|\, \tau, s] = L(z)/\sqrt{\tau} = B^\circ$ with respect to $\tau$, we obtain

$$\begin{aligned}\partial z/\partial\tau &= -B^\circ/\left(2\sqrt{\tau}\overline{\Phi}(z)\right) < 0,\\ \partial^2 z/\partial\tau^2 &= B^\circ[\overline{\Phi}(z)/\left(2\sqrt{\tau}\right) - \sqrt{\tau}\phi\left(z\right)\left(\partial z/\partial\tau\right)]/(2\tau\left[\overline{\Phi}(z)\right]^2) > 0.\end{aligned}$$

Furthermore, from $s = 1/\tau + z/\sqrt{\tau}$ and $z \geq \underline{z} = -1/\sqrt{\tau}$ (see (11)), we obtain

$$\begin{aligned}\frac{\partial s}{\partial\tau} &= -\frac{1}{\tau^2} - \frac{z}{2\tau^{3/2}} + \frac{1}{\sqrt{\tau}}\frac{\partial z}{\partial\tau} \leq -\frac{1}{2\tau^2} + \frac{1}{\sqrt{\tau}}\frac{\partial z}{\partial\tau} < 0,\\ \frac{\partial^2 s}{\partial\tau^2} &= \frac{2}{\tau^3} + \frac{3z}{4\tau^{5/2}} - \frac{1}{\tau^{3/2}}\frac{\partial z}{\partial\tau} + \frac{1}{\sqrt{\tau}}\frac{\partial^2 z}{\partial\tau^2} \geq \frac{5}{4\tau^3} - \frac{1}{\tau^{3/2}}\frac{\partial z}{\partial\tau} + \frac{1}{\sqrt{\tau}}\frac{\partial^2 z}{\partial\tau^2} > 0.\end{aligned}$$

Similarly, differentiating $L(z)/\sqrt{\tau} = B^\circ$ with respect to $B^\circ$ yields $\partial z/\partial B^\circ = -\sqrt{\tau}/\overline{\Phi}(z)$ and

$$\partial s/\partial B^\circ = \left(1/\sqrt{\tau}\right)\left(\partial z/\partial B^\circ\right) = -1/\overline{\Phi}(z) < 0.$$

■

**Proof of Proposition 2.** After differentiating $\pi(\tau, s)$ with respect to $s$, we obtain

$$\partial\pi(\tau, s)/\partial s = p - c + v - (v + \delta h)\Phi(z), \qquad \partial^2\pi(\tau, s)/\partial s^2 = -(v + \delta h)\sqrt{\tau}\phi(z) \leq 0.$$

(Case 1) If $p - c + v \leq 0$, $\partial\pi(\tau, s)/\partial s \leq 0$ regardless of $\tau$ and it is optimal to set $s = 0$. Then the simplified profit expression is $\pi(\tau) \equiv \pi(\tau, 0) = w - K(\tau) - vE[B \mid \tau, 0] - \delta hE[I \mid \tau, 0] \simeq w - K(\tau) - v/\tau$ and

$$d\pi(\tau)/d\tau = -K'(\tau) + v/\tau^2, \qquad d^2\pi(\tau)/d\tau^2 = -K''(\tau) - 2v/\tau^3 \leq 0.$$

and Case 1 follows.

(Case 2) $\widehat{z} = \Phi^{-1}\left((v + p - c) / (v + \delta h)\right)$ is obtained at optimality and we see that it is independent of $\tau$. As described in the proof of Proposition 1, the original problem is reduced to a one-dimensional optimization

$$\max_\tau \; \pi(\tau) = w - (c - p)/\tau - K(\tau) - \phi(\widehat{z})\,(v + \delta h)/\sqrt{\tau}.$$

Observe that $\pi(\tau) \to -\infty$ as $\tau \to \overline{\tau}$. Differentiating,

$$\begin{aligned}
\pi'(\tau) &= -K'(\tau) + \frac{c-p}{\tau^2} + \frac{v+\delta h}{2\tau^{3/2}}\phi(\widehat{z}) = -K'(\tau) + \gamma_\delta(\tau; v, p), \\
\pi''(\tau) &= -K''(\tau) - \frac{2(c-p)}{\tau^3} - \frac{3(v+\delta h)}{4\tau^{5/2}}\phi(\widehat{z}).
\end{aligned}$$

It is straightforward to see that $\pi(\tau)$ is concave if $p \leq c$. To see that it is quasiconcave if $p > c$, we evaluate $\pi''(\tau)$ at its critical point $\widehat{\tau}$ (by multiplying the first-order condition $\pi'(\tau) = 0$ by $3/(2\tau)$ and substituting it in $\pi''(\tau)$) to obtain

$$\begin{aligned}
\pi''(\widehat{\tau}) &= -K''(\widehat{\tau}) - 3K'(\widehat{\tau})/\left(2\widehat{\tau}\right) + (p-c)/\left(2\widehat{\tau}^3\right) < -K''(\widehat{\tau}) - 3K'(\widehat{\tau})2\widehat{\tau}/ + h/\left(2\widehat{\tau}^3\right) \\
&\leq -3\left(\widehat{\tau}^2 K'(\widehat{\tau}) + h\right)/\left(2\widehat{\tau}^3\right) < 0,
\end{aligned}$$

where the first inequality comes from the assumption that $p - c < \delta h \leq h$ and the second inequality comes from (2), because $\underline{\tau} \leq \widehat{\tau}$ and $K'''(\tau) > 0$ imply $\widehat{\tau}^3 K''(\widehat{\tau}) \geq \underline{\tau}^3 K''(\underline{\tau}) \geq 2h$ or $-K''(\widehat{\tau}) \leq -2h/\widehat{\tau}^3$. Finally, $\pi''(\widehat{\tau}) < 0$ implies that a critical point, if it exists, is never a local minimum, and quasiconvexity follows. The solution is then characterized by the first-order condition as stated in the Proposition.

(Case 3) For a finite $z$, $\Phi(z) < 1$ and thus $\partial\pi(\tau, s)/\partial s \geq p - c + v - (v + \delta h) = p - c - \delta h > 0$, so it is optimal for the supplier to choose $s = \infty$. The same result is true for $z = \infty$. Observe that

$$\lim_{s\to\infty} \frac{\partial \pi(\tau, s)}{\partial \tau} = \lim_{s\to\infty} \left\{ -K'(\tau) + \frac{v\phi(z)}{2\tau^{3/2}} + \frac{v\overline{\Phi}(z)}{\tau^2} + \frac{\delta h\phi(z)}{2\tau^{3/2}} - \frac{\delta h \Phi(z)}{\tau^2} \right\} = -K'(\tau) - \frac{\delta h}{\tau^2} < 0$$

for any $\tau \geq \underline{\tau}$. Therefore, the supplier chooses $\tau = \underline{\tau}$. ■

**Proof of Proposition 3.** Before we prove this result, we need the following auxiliary Lemma, which is used in the proof along with the result in Lemma 3.

**Lemma 4** *Assume $p - c < \delta h$ and let $\eta_\delta(v,p) \equiv (v+\delta h)\phi\left(z_\delta(v,p)\right)$. For $v$ defined on $(c-p,\infty)$, $\frac{\partial \eta_\delta(v,p)}{\partial v} = L(z_\delta(v,p)) > 0$ and $\frac{\partial \eta_\delta(v,p)}{\partial \delta} = h\left(\phi\left(z_\delta(v,p)\right) + \frac{z_\delta(v,p)}{v+\delta h}\right) > 0$.*

**Proof.** $\lim_{v\to c-p} \eta_\delta(v,p) = 0$ by (9) and by L'Hopital's rule,

$$\lim_{v\to\infty} \eta_\delta(v,p) = \lim_{v\to\infty} \frac{\phi\left(z_\delta(v,p)\right)}{1/(v+\delta h)} = \lim_{v\to\infty} \frac{-\frac{\delta h - p + c}{(v+\delta h)^2} z_\delta(v,p)}{-1/(v+\delta h)^2} = \lim_{v\to\infty} (\delta h - p + c) z_\delta(v,p) = \infty.$$

Also,

$$\frac{\partial \eta_\delta(v,p)}{\partial v} = \phi\left(z_\delta(v,p)\right) - \left(\frac{\delta h - p + c}{v + \delta h}\right) z_\delta(v,p) = L(z_\delta(v,p)) > 0$$

for $v \in (c-p,\infty)$. These results imply that $\lim_{v\to c-p}\left[\eta_\delta(v,p) + z_\delta(v,p)\right] = 0$, $\lim_{v\to\infty}\left[\eta_\delta(v,p) + z_\delta(v,p)\right] = \infty$ , and

$$\frac{\partial}{\partial v}\left[\eta_\delta(v,p) + z_\delta(v,p)\right] = L(z_\delta(v,p)) + \frac{\delta h - p + c}{(v+\delta h)^2 \phi\left(z_\delta(v,p)\right)} > 0.$$

Thus $\eta_\delta(v,p) + z_\delta(v,p) > 0$ for $v \in (c-p,\infty)$ and

$$\frac{\partial \eta_\delta(v,p)}{\partial \delta} = h\phi\left(z_\delta(v,p)\right) + \frac{h z_\delta(v,p)}{v+\delta h} = \frac{h\left[\eta_\delta(v,p) + z_\delta(v,p)\right]}{v+\delta h} > 0.$$

■

We are now ready to prove the main results. Note that

$$\frac{\partial \widehat{z}}{\partial p} = \frac{1}{v+\delta h}\frac{1}{\phi\left(\widehat{z}\right)} > 0, \qquad \frac{\partial \widehat{z}}{\partial v} = \frac{\delta h - p + c}{(v+\delta h)^2}\frac{1}{\phi\left(\widehat{z}\right)} > 0, \qquad \frac{\partial \widehat{z}}{\partial \delta} = -\frac{h}{(v+\delta h)^2}\frac{1}{\phi\left(\widehat{z}\right)} < 0. \tag{17}$$

Case 1. By differentiating both sides of the first-order condition $\gamma_\delta(\widehat{\tau}; v, p) = K'(\widehat{\tau})$ with respect to $p$, we obtain

$$K''(\widehat{\tau})\frac{\partial \widehat{\tau}}{\partial p} = -\frac{2(c-p)}{\widehat{\tau}^3}\frac{\partial \widehat{\tau}}{\partial p} - \frac{3(v+\delta h)}{4\widehat{\tau}^{5/2}}\phi(\widehat{z})\frac{\partial \widehat{\tau}}{\partial p} - \frac{1}{\widehat{\tau}^2} - \frac{\widehat{z}}{2\widehat{\tau}^{3/2}}.$$

After collecting terms, we have

$$\frac{\partial \widehat{\tau}}{\partial p} = \frac{1}{\pi''(\widehat{\tau})}\left(\frac{1}{\widehat{\tau}^2} + \frac{\widehat{z}}{2\widehat{\tau}^{3/2}}\right) = \frac{1}{\pi''(\widehat{\tau})}\left(\frac{1}{2\widehat{\tau}^2} + \frac{\widehat{s}}{2\widehat{\tau}}\right) < 0, \tag{18}$$

where $\pi''(\widehat{\tau}) = -K''(\widehat{\tau}) - 2(c-p)/\widehat{\tau}^3 - 3\phi(\widehat{z})(v+\delta h)/\left(4\widehat{\tau}^{5/2}\right) < 0$ since $\widehat{\tau}$ is the local maximizer.

Moreover,

$$\frac{\partial \widehat{s}}{\partial p} = \frac{\partial}{\partial p}\left(\frac{1}{\widehat{\tau}} + \frac{\widehat{z}}{\sqrt{\widehat{\tau}}}\right) = -\left(\frac{1}{\widehat{\tau}^2} + \frac{\widehat{z}}{2\widehat{\tau}^{3/2}}\right)\frac{\partial \widehat{\tau}}{\partial p} + \frac{1}{\sqrt{\widehat{\tau}}}\frac{\partial \widehat{z}}{\partial p} = -\left(\frac{1}{2\widehat{\tau}^2} + \frac{\widehat{s}}{2\widehat{\tau}}\right)\frac{\partial \widehat{\tau}}{\partial p} + \frac{1}{\sqrt{\widehat{\tau}}}\frac{\partial \widehat{z}}{\partial p} > 0,$$

since $\partial \widehat{z}/\partial p > 0$ and $\partial \widehat{\tau}/\partial p < 0$, and similarly (using Lemma 1)

$$\begin{aligned}
\frac{\partial E[I \,|\, \widehat{\tau}, \widehat{s}]}{\partial p} &= \frac{\partial E[I \,|\, \widehat{\tau}, \widehat{s}]}{\partial \tau}\frac{\partial \widehat{\tau}}{\partial p} + \frac{\partial E[I \,|\, \widehat{\tau}, \widehat{s}]}{\partial s}\frac{\partial \widehat{s}}{\partial p} \\
&= \left[-\frac{\phi(z)}{2\widehat{\tau}^{3/2}} + \frac{1}{\widehat{\tau}^2}\left(\frac{v+p-c}{v+\delta h}\right)\right]\frac{\partial \widehat{\tau}}{\partial p} + \left(\frac{v+p-c}{v+\delta h}\right)\left[-\left(\frac{1}{\widehat{\tau}^2} + \frac{\widehat{z}}{2\widehat{\tau}^{3/2}}\right)\frac{\partial \widehat{\tau}}{\partial p} + \frac{1}{\sqrt{\widehat{\tau}}}\frac{\partial \widehat{z}}{\partial p}\right] \\
&= -\frac{E[I \,|\, \widehat{\tau}, \widehat{s}]}{2\widehat{\tau}}\frac{\partial \widehat{\tau}}{\partial p} + \left(\frac{v+p-c}{v+\delta h}\right)\frac{1}{\sqrt{\widehat{\tau}}}\frac{\partial \widehat{z}}{\partial p} > 0.
\end{aligned}$$

$\partial E[B \,|\, \widehat{\tau}, \widehat{s}]/\partial p < 0$ will be shown below.

Case 2. We first differentiate both sides of the first-order condition $\gamma_\delta(\widehat{\tau}; v, p) = K'(\widehat{\tau})$ with respect to $s$ to obtain

$$\frac{\partial \widehat{\tau}}{\partial v} = -\frac{L(\widehat{z})}{2\widehat{\tau}^{3/2}\pi''(\widehat{\tau})} > 0. \tag{19}$$

Next,

$$\frac{\partial E[B \,|\, \widehat{\tau}, \widehat{s}]}{\partial v} = \frac{\partial}{\partial v}\left(\frac{L(\widehat{z})}{\sqrt{\widehat{\tau}}}\right) = -\frac{L(\widehat{z})}{2\widehat{\tau}^{3/2}}\frac{\partial \widehat{\tau}}{\partial v} - \frac{\overline{\Phi}(\widehat{z})}{\sqrt{\widehat{\tau}}}\frac{\partial \widehat{z}}{\partial v} < 0.$$

To show $\partial \widehat{s}/\partial v > 0$, notice that (2) together with the assumption $p - c < \delta h$ imply that $\widehat{\tau}^3 K''(\widehat{\tau}) \geq \underline{\tau}^3 K''(\underline{\tau}) \geq 2h \geq 2\delta h > \delta h + p - c$. Furthermore

$$\begin{aligned}
-\pi''(\widehat{\tau}) &= K''(\widehat{\tau}) + \frac{2(c-p)}{\widehat{\tau}^3} + \frac{3(v+\delta h)}{4\widehat{\tau}^{5/2}}\phi(\widehat{z}) > \frac{\delta h + p - c}{\widehat{\tau}^3} + \frac{2(c-p)}{\widehat{\tau}^3} + \frac{3(\delta h - p + c)\widehat{z}}{4\widehat{\tau}^{5/2}} \\
&= \frac{\delta h - p + c}{\widehat{\tau}^3} + \frac{3(\delta h - p + c)}{4\widehat{\tau}^{5/2}}\left(\sqrt{\widehat{\tau}}\widehat{s} - \frac{1}{\sqrt{\widehat{\tau}}}\right) = \frac{\delta h - p + c}{4\widehat{\tau}^3} + \frac{3(\delta h - p + c)\widehat{s}}{4\widehat{\tau}^2} \\
&> \frac{\delta h - p + c}{2\widehat{\tau}}\left(\frac{1}{2\widehat{\tau}^2} + \frac{\widehat{s}}{2\widehat{\tau}}\right),
\end{aligned} \tag{20}$$

where we used $\phi(\widehat{z}) \geq \widehat{z}\overline{\Phi}(\widehat{z}) = \widehat{z}\,(\delta h - p + c)\,/\,(v + \delta h)$ in the first inequality. From Lemma 3,

$$\frac{\phi(\widehat{z})L(\widehat{z})}{[\overline{\Phi}(\widehat{z})]^2} = \left(\frac{v+\delta h}{\delta h - p + c}\right)^2 \phi(\widehat{z})L(\widehat{z}) \leq 1,$$

or similarly,

$$-L(\widehat{z}) + \left(\frac{\delta h - p + c}{v + \delta h}\right)^2 \frac{1}{\phi(\widehat{z})} \geq 0. \tag{21}$$

After combining (17), (19), (20), and (21), we obtain

$$
\begin{aligned}
\frac{\partial \widehat{s}}{\partial v} &= -\left(\frac{1}{2\widehat{\tau}^2}+\frac{\widehat{s}}{2\widehat{\tau}}\right)\frac{\partial \widehat{\tau}}{\partial v}+\frac{1}{\sqrt{\widehat{\tau}}}\frac{\partial \widehat{z}}{\partial v} \\
&= -\left(\frac{1}{2\widehat{\tau}^2}+\frac{\widehat{s}}{2\widehat{\tau}}\right)\frac{L(\widehat{z})}{2\widehat{\tau}^{3/2}\left(-\pi''(\widehat{\tau})\right)}+\frac{\delta h-p+c}{(v+\delta h)^2}\frac{1}{\sqrt{\widehat{\tau}}\phi\left(\widehat{z}\right)} \\
&> -\left(\frac{1}{2\widehat{\tau}^2}+\frac{\widehat{s}}{2\widehat{\tau}}\right)\frac{L(\widehat{z})}{\sqrt{\widehat{\tau}}(\delta h-p+c)\left(\frac{1}{2\widehat{\tau}^2}+\frac{\widehat{s}}{2\widehat{\tau}}\right)}+\frac{\delta h-p+c}{(v+\delta h)^2}\frac{1}{\sqrt{\widehat{\tau}}\phi\left(\widehat{z}\right)} \\
&= \frac{1}{\sqrt{\widehat{\tau}}(\delta h-p+c)}\left[-L(\widehat{z})+\left(\frac{\delta h-p+c}{v+\delta h}\right)^2\frac{1}{\phi\left(\widehat{z}\right)}\right]\geq 0.
\end{aligned}
$$

Note that

$$
\frac{\partial \widehat{s}}{\partial v}=-\left(\frac{1}{\widehat{\tau}^2}+\frac{\widehat{z}}{2\widehat{\tau}^{3/2}}\right)\frac{\partial \widehat{\tau}}{\partial v}+\frac{1}{\sqrt{\widehat{\tau}}}\frac{\partial \widehat{z}}{\partial v}=\frac{L(\widehat{z})}{2\widehat{\tau}^{3/2}}\frac{\partial \widehat{\tau}}{\partial p}+\frac{\overline{\Phi}(\widehat{z})}{\sqrt{\widehat{\tau}}}\frac{\partial \widehat{z}}{\partial p}=-\frac{\partial E[B\,|\,\widehat{\tau},\widehat{s}]}{\partial p}
$$

where the second equality follows from (17), (18), (19), and Lemma 1. Thus $\partial E[B\,|\,\widehat{\tau},\widehat{s}]/\partial p<0$ in Case 1. In addition,

$$
\frac{\partial E[I\,|\,\widehat{\tau},\widehat{s}]}{\partial v}=\frac{\partial E[I\,|\,\widehat{\tau},\widehat{s}]}{\partial \tau}\frac{\partial \widehat{\tau}}{\partial v}+\frac{\partial E[I\,|\,\widehat{\tau},\widehat{s}]}{\partial s}\frac{\partial \widehat{s}}{\partial v}>0.
$$

Case 3. We differentiate both sides of the first-order condition $\gamma_\delta(\widehat{\tau};v,p)=K'(\widehat{\tau})$ with respect to $\delta$, obtaining

$$
\frac{\partial \widehat{\tau}}{\partial \delta}=-\frac{h}{2\widehat{\tau}^{3/2}\pi''(\widehat{\tau})}\left(\phi(\widehat{z})+\frac{\widehat{z}}{v+\delta h}\right)>0,
$$

where the inequality follows from the second result of Lemma 4. With (17), this implies

$$
\frac{\partial \widehat{s}}{\partial \delta}=-\left(\frac{1}{2\widehat{\tau}^2}+\frac{\widehat{s}}{2\widehat{\tau}}\right)\frac{\partial \widehat{\tau}}{\partial \delta}+\frac{1}{\sqrt{\widehat{\tau}}}\frac{\partial \widehat{z}}{\partial \delta}<0.
$$

Also,

$$
\frac{\partial E[I\,|\,\widehat{\tau},\widehat{s}]}{\partial \delta}=-\frac{E[I\,|\,\widehat{\tau},\widehat{s}]}{2\widehat{\tau}}\frac{\partial \widehat{\tau}}{\partial \delta}+\left(\frac{v+p-c}{v+\delta h}\right)\frac{1}{\sqrt{\widehat{\tau}}}\frac{\partial \widehat{z}}{\partial \delta}<0
$$

where the equality is derived similarly to $\partial E[I\,|\,\widehat{\tau},\widehat{s}]/\partial p$ above. ■

**Proof of Proposition 5.** With $p=0$, only Cases 1 and 2 in Proposition 2 are possible. It can be shown that the condition (5) excludes the case $v\leq c$, which lead to $s=0$. Thus only Case 2 remains. Suppose $v>c$. According to Proposition 2, the supplier chooses $\tau^*=\underline{\tau}$ and $s^*=s(\underline{\tau},z_\delta^P(v))$ if

$\gamma_\delta^P(\underline{\tau};v) \le K'(\underline{\tau})$ and $\tau^* = \widehat{\tau}$ and $s^* = s(\tau_\delta^P(v), z_\delta^P(v))$ otherwise, where $\tau_\delta^P(v)$ solves $\gamma_\delta^P(\tau;v) = K'(\tau)$. Consider $\gamma_\delta^P(\underline{\tau};v) \le K'(\underline{\tau})$ first. In this case, the Lagrangian for the customer's problem is

$$\mathcal{L}^P = -\theta B^\circ + K(\underline{\tau}) + \frac{c}{\underline{\tau}} + \frac{\theta + h}{\sqrt{\underline{\tau}}}\phi(z_\delta^P(v)) + \frac{(\theta+h)z_\delta^P(v)}{\sqrt{\underline{\tau}}}\left(\frac{v-c}{v+\delta h} - \frac{\theta - c}{\theta + h}\right).$$

Differentiating it with respect to $v$, we obtain

$$\frac{\partial \mathcal{L}^P}{\partial v} = \frac{(\theta+h)(c+\delta h)}{(v+\delta h)^2\sqrt{\underline{\tau}}\phi(z_\delta^P(v))}\left(\frac{v-c}{v+\delta h} - \frac{\theta-c}{\theta+h}\right).$$

If $\theta \le c$, $\partial \mathcal{L}^P > \partial v$, and hence $v = c$ is optimal, violating the assumption. Thus $\theta > c$ and the optimal $v$ is

$$v(\theta) = \theta\,(c+\delta h)\,/\,(c+h) + (1-\delta)ch/\,(c+h)\,.$$

In view of this relationship, finding optimal $v$ is equivalent to finding optimal $\theta$. Since $z_\delta^P(v(\theta)) = \Phi^{-1}\left((v(\theta)-c)\,/\,(v(\theta)+\delta h)\right) = \Phi^{-1}\left((\theta - c)\,/\,(\theta+h)\right)$, the optimal $\theta$ is the solution to the binding constraint which is identical to that of the first-best case: $\beta^P(v(\theta);\underline{\tau}) = \beta^{FB}(\theta;\underline{\tau}) = B^\circ$. The solution $\overline{\theta}$ was shown to be unique in Proposition 1, hence uniqueness of $\overline{v} = v(\overline{\theta})$ follows. Next, suppose $\gamma^P(\underline{\tau};\overline{v}) > K'(\underline{\tau})$. Existence and uniqueness of $v^\ddagger$ can be shown similarly to those of $\theta^\dagger$ in the proof of Proposition 1, so we omit details. ■

**Proof of Proposition 6.** Case 1 is clear after comparing Propositions 1 and 4. For Case 2, there are two cases to consider: $\gamma^P(\underline{\tau};\overline{v}) \le K'(\underline{\tau}) < \gamma^{FB}(\underline{\tau};\overline{\theta})$ and $\gamma^P(\underline{\tau};\overline{v}) > K'(\underline{\tau})$. For the former, $\tau^P = \underline{\tau} < \tau^\dagger = \tau^{FB}$ and $s^P = s(\underline{\tau}, z^{FB}(\overline{\theta})) > s(\tau^\dagger, z^{FB}(\theta^\dagger)) = s^{FB}$ since $\overline{\theta} \ge \theta^\dagger$ implies $z^{FB}(\overline{\theta}) \ge z^{FB}(\theta^\dagger)$. Since $\overline{v} \le \overline{\theta}$ and $\theta^\dagger \le \overline{\theta}$, the sign of $\overline{v} - \theta^\dagger$ is not clear. For the latter, let $v^\dagger$ be the penalty rate that induces the supplier to choose first-best $\tau$, $\tau^\dagger$. We will now demonstrate that, for $\delta \in [0,1)$, the customer has to choose $v > \theta^\dagger$ in order to induce $\tau^* = \tau^\dagger$, the first-best interior solution for $\tau$. Suppose that $v^\dagger$ induces the supplier to choose $\tau^* = \tau^\dagger$ at equilibrium. From the optimality conditions in Propositions 1 and 2, we have

$$\frac{c}{(\tau^\dagger)^2} + \frac{v^\dagger + \delta h}{2(\tau^\dagger)^{3/2}}\phi(z_\delta^P(v^\dagger)) = K'(\tau^\dagger) = \frac{c}{(\tau^\dagger)^2} + \frac{\theta^\dagger + h}{2(\tau^\dagger)^{3/2}}\phi(z^{FB}(\theta^\dagger)),$$

or similarly $(v^\dagger + \delta h)\phi(z_\delta(v^\dagger, 0)) = (\theta^\dagger + h)\phi(z_1(\theta^\dagger, 0))$. Since Lemma 4 implies that $(v+\delta h)\phi(z_\delta(v,0))$ increases in $v$ and $(v+\delta h)\phi(z_\delta(v,0)) < (v+\delta h)\phi(z_1(v,0))$ for $\delta \in [0,1)$, we conclude that $v^\dagger > \theta^\dagger$.

Using this result, we obtain that

$$z^P(v^\dagger) = \Phi^{-1}((v^\dagger - c)/(v^\dagger + \delta h)) \geq \Phi^{-1}((\theta^\dagger - c)/(\theta^\dagger + \delta h)) \geq \Phi^{-1}((\theta^\dagger - c)/(\theta^\dagger + h)) = z^{FB}(\theta^\dagger)$$

where the first inequality comes from the fact that $(x - c)/(x + \delta h)$ is increasing in $x$. Since $L(z)$ is a decreasing function, we have

$$E[B \,|\, \tau^\dagger, s(\tau^\dagger, z^P(v^\dagger))] = L(z^P(v^\dagger))/\sqrt{\tau^\dagger} \leq L(z^{FB}(\theta^\dagger))/\sqrt{\tau^\dagger} = E[B \,|\, \tau^\dagger, s(\tau^\dagger, z^{FB}(\theta^\dagger))].$$

Combined with the fact that $E[B \,|\, \tau^P(v), s(\tau^P(v), z^P(v))]$ is decreasing in $v$ (see Proposition 3), this inequality implies that $v^\ddagger$, the optimal $v$ under performance contract, should be such that $v^\ddagger \leq v^\dagger$ in order to have the backorder constraint binding. Hence $\tau^P(v^\ddagger) \leq \tau^P(v^\dagger) = \tau^\dagger$, since $\tau^P(v)$ is increasing. Then $s(\tau^P(v^\ddagger), z^P(v^\ddagger)) \geq s(\tau^\dagger, z^{FB}(\theta^\dagger))$ by Lemma 2. The last statement in the Proposition follows from similar logic. ■

**Proof of Proposition 7.** Case 1. Since $\tau^M = \underline{\tau}$ is fixed as long as $v = 0$, it does not change with $B^\circ$. $\partial s^M/\partial B^\circ < 0$ follows from Lemma 2. $\partial p^M/\partial B^\circ < 0$ is inferred from (14) and $\partial\overline{\theta}/\partial B^\circ < 0$, which is evident from the relation $\beta^{FB}(\theta; \underline{\tau}) = L(z^{FB}(\theta))/\sqrt{\underline{\tau}} = B^\circ$. Case 2. If $\gamma^{FB}(\underline{\tau}; \overline{\theta}) < K'(\underline{\tau})$, then $\tau^P = \underline{\tau}$ and the results identical to Case 1 are obtained. The statements in (ii) are true with $\partial\tau^P/\partial B^\circ = 0$ and $\left|\partial s^P/\partial B^\circ\right| = \left|\partial s^M/\partial B^\circ\right|$. Suppose $\gamma^{FB}(\underline{\tau}; \overline{\theta}) \geq K'(\underline{\tau})$, in which case $\tau^P > \underline{\tau}$ is optimal. The condition (2) ensures that the backorder constraint is binding, i.e., $\beta^P(v; \tau^P_\delta(v)) = B^\circ$. Since $\beta^P(v; \tau^P_\delta(v))$ is a decreasing function of $v$ (see Proposition 3), we find $\partial v^P/\partial B^\circ < 0$. This in turn implies $\partial\tau^P/\partial B^\circ = \left(\partial\tau^P/\partial v\right)\left(\partial v^P/\partial B^\circ\right) < 0$ and $\partial s^P \partial B^\circ/ = \left(\partial s^P/\partial v\right)\left(\partial v^P/\partial B^\circ\right) < 0$ via results in Proposition 3. ■

# Appendix B: Discussion of Technical Assumption

We comment on what happens if condition (2) is not imposed, i.e., if $K(\tau)$ is not sufficiently convex. First, in this case the supplier's profit function may not be well-behaved. In particular, it can be shown that $\pi(\tau, s)$ is not quasiconcave in general and can have a corner solution even in the presence of an interior maximum. As a result, the supplier's optimal choices of $\tau$ and $s$ may exhibit jumps with respect to changes in contract parameters, so it becomes very difficult to analyze the supplier's response. Second, three results in Proposition 3 ($\partial\widehat{s}/\partial v > 0$, $\partial E[B \,|\, \widehat{\tau}, \widehat{s}]/\partial p < 0$, and $\partial E[I \,|\, \widehat{\tau}, \widehat{s}]/\partial v > 0$) do not necessarily hold without assumption (2). To illustrate, recall that

$$\widehat{s} = 1/\widehat{\tau} + \Phi^{-1}\left((v+p-c)\,/\,(v+\delta h)\right)\Big/\sqrt{\widehat{\tau}}\;.$$

Thus, when $v$ increases, there are two opposing forces at work. On the one hand, the supplier reduces the probability of being backordered by increasing the $z$-value (or the normalized fractile), $\Phi^{-1}\left((v+p-c)\,/\,(v+\delta h)\right)$. This decision results in a higher $\widehat{s}$. On the other hand, the supplier reduces the number of backorders by investing into product reliability which leads to a smaller $\widehat{s}$. The combined effect is such that $\widehat{s}$ may not be monotone in $v$. Consequently, the customer's objective function $\widetilde{C}^{SB}$ may not be monotone in $v$ either so it becomes possible that the availability constraint does not bind at the equilibrium, i.e., $E[B\,|\,\widehat{\tau},\widehat{s}] < B^{\circ}$. The implication is that the system may end up with expected availability beyond the specified target that comes at a cost beyond what the customer is willing to pay. Although this observation is of some interest, we do not pursue it further because it obscures the central theme of our paper, the tradeoff between reliability and inventory. Moreover, it causes several related technical challenges: for example, it becomes difficult to compare equilibrium solutions under various scenarios (first-best, material contract, performance contract) and some of the main results of our paper (e.g., Table 1) become hard to show. We believe that little is lost by ignoring this rather impractical scenario.