Dynamic Business Share Allocation in a Supply Chain with Competing Suppliers

Hongmin Li[1], Hao Zhang[2], and Charles H. Fine[3]

This paper studies a repeated game between a manufacturer and two competing suppliers with imperfect monitoring. We present a principal-agent model for managing long-term supplier relationships using a unique form of measurement and incentive scheme. We measure a supplier's overall performance with a rating equivalent to its continuation utility (the expected total discounted utility of its future payoffs), and incentivize supplier effort with larger allocations of future business. We obtain the vector of the two suppliers' ratings as the state of a Markov decision process, and solve an infinite horizon contracting problem in which the manufacturer allocates business volume between the two suppliers and updates their ratings dynamically based on their current ratings and the current performance outcome.

Our contributions are both theoretical and managerial: We propose a repeated principal-agent model with a novel incentive scheme to tackle a common, but challenging incentive problem in a multi-period supply chain setting. Assuming binary effort choices and performance outcomes by the suppliers, we characterize the structure of the optimal contract through a novel fixed-point analysis. Our results provide a theoretical foundation for the emergence of "business-as-usual" (low effort) trapping states and tournament competition (high effort) recurrent states as the long-run incentive drivers for motivating critical suppliers.

Keywords: Asymmetric Information, Performance-Based Contract, Volume Incentive, Repeated Moral Hazard, Principal-Agent Model, Supply Chain Contracting

[1] W.P. Carey School of Business, Arizona State University, Tempe, AZ 85287. Email: hongmin.li@asu.edu.

[2] Sauder School of Business, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada. Email: hao.zhang@sauder.ubc.ca.

[3] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139. Email: charley@mit.edu.

1 Introduction

We model and analyze the use of business share (or volume) to motivate performance improvements from critical suppliers. Throughout the last few decades, many companies have reduced the number of suppliers they use, and focused on improving the quality of the relationships they have with those remaining (Giunipero, 1990). In consumer electronics, information technology, and other industries, dual-sourcing (or multi-sourcing from a few suppliers) has become a common practice. For example, Apple Inc. often sources critical parts from two suppliers: Solid State Drives for MacBook Air from Toshiba and Samsung (O'Grady, 2011); DRAM for iPhone 4S from Samsung and Elpida (Shimpi and Klug, 2011); and assembly manufacturing for iPhone 4S from Foxconn and Pegatron (Whitney, 2012). Supply base reduction allows a firm to focus on long-term ties with suppliers but may potentially reduce the power of the buying firm. How can the manufacturer prevent a supplier from getting "too comfortable" to improve? Krause et al. (2000) surveyed 527 purchasing executives and found that supplier assessment and supplier incentives are the two most important enablers of supplier development efforts. The incentives identified in their research are (1) promise of higher order volume for current business, and (2) promise of preferred status for future business. That is, performance-based business share allocation is used to drive competition among suppliers and keep the suppliers on their toes. In a Japanese vertical, Keiretsu-style supply chain, a lead firm often multi-sources to a few suppliers and uses business share incentives to drive supplier efficiency improvements (Tezuka, 1997). A supplier that fails to meet the competitive standard over some extended period of time will lose business share and its preferred status.

In this paper, we focus on the incentive issues arised when a manufacturer cannot directly observe or verify its suppliers' effort decisions that affect the delivered value to the manufacturer. For example, in each contract period, a supplier may boost its quality-control effort to reduce the defect rate, optimize the equipment maintenance schedule to decrease machine down time, or assign the most effective account manager to manage the production and delivery for this manufacturer. The level of these efforts is not easily verifiable by the manufacturer, but can affect the supplier's performance and thus the delivered value to the manufacturer greatly. We explore via a principal-agent model how a manufacturer can induce the desired supplier behavior through business share allocation based on supplier performance. We examine this in the context of a cost-plus contract in which the transfer price between each supplier and the manufacturer is the unit cost of the component plus a margin.

We make both technical and managerial contributions to the supply chain management and contract design literature. On the technical side, we propose a novel principal-agent model for performance-based supplier incentive schemes in a dual-sourced supply chain. Our model is an (infinitely) repeated moral hazard model with imperfect monitoring, which is known for its theoretical challenge: "Generally speaking, the design of an optimal compensation scheme in the dynamic principal-agent context is considered an intractable problem. In fact, even in the simpler repeated principal-agent setting, the analysis of optimal schemes is formidable and involves complex and subtle economic reasoning" (Plambeck and Zenios, 2000). We are among the very few to tackle a *two-agent* repeated moral hazard model. We characterize the optimal contract through a novel fixed-point analysis. Extending the dynamic programming approach of Spear and Srivastava (1987) for a single-agent model, we formulate the two-agent problem in a recursive fashion and construct the fixed point (function) directly, which allows us to obtain interesting structural results.

Managerially, our study provides theoretical explanations to popular business practices. In the study by Giunipero (1990), 46% of the firms studied use formal quantitative rating systems to monitor and motivate suppliers. Empirical research has documented many instances of rating/scoring systems for suppliers. For example, Nike regularly rates its subcontractors for environmental and labor performance (Sabel et al., 2000). High scorers often garner more lucrative orders and low scorers risk losing contracts. Intel tracks a supplier's cost, availability, service, support responsiveness and quality, and rewards suppliers who have the best ratings with more business (Datta, 2004). Despite the apparent prevalence in practice, there are no published theoretical results addressing these widely used supplier management practices. Our results fill this gap and explain the relationship between a quantitative supplier assessment system and the manufacturer's decisions on suppliers' business shares.

A central managerial finding in this paper relates to the longitudinal behaviors of the supply chain under the optimal contract. In our model, the state of the system is given by the vector of the two suppliers' ratings (quantified as their sustainable continuation values, or values-to-go). Under the optimal contract, three types of states emerge. (i) A set of "trapping" states in which the suppliers choose low effort *forever*. Each trapping state represents a "business-as-usual" scenario with a state-dependent but fixed volume allocation for *all* future periods, which is reached after both or at least one supplier over-perform for some extended time. Since each supplier prefers a trapping state that yields a higher volume for itself, this creates incentive for suppliers to continually exert high effort in order to influence the direction of the state transition. (ii) A "recurrent" class of states, in which suppliers engage in a tournament-like competition and both choose high effort *forever* in an effort to win a preferential status for future business. This represents an ideal situation for the manufacturer but a punishing situation for the suppliers, and is usually reached after both suppliers repeatedly under-perform. (iii) "Transient" states, from which the system eventually evolves into either a business-as-usual scenario or a tournament competition situation. Therefore, cases (i) and (ii) form the long-run incentive drivers, as the "carrot" or "stick", for the suppliers to work hard.

The rest of the paper is organized as follows: Section 2 reviews the relevant literature in economics and operations. Section 3 provides the problem description and assumptions. We present the solution of the history-dependent dynamic contract problem in Section 4. In Section 5, we further explore properties of the optimal contract numerically. Section 6 discusses extensions of the basic model and Section 7 concludes. The proofs of the results are given in Appendix A, and more details of the extensions are given in Appendix B.

2 Related Literature

Moral Hazard (Hidden Action). In this paper, we consider an incentive problem with moral hazard where a manufacturer (the buyer) does not directly observe its suppliers' effort decisions and needs to design incentive mechanisms to induce desired supplier behavior. Single-period moral hazard problems have been extensively studied in economics; see Laffont and Martimort (2002), Bolton and Dewatripont (2005), and references therein. Moral hazard problems have frequently emerged in operations management (supply chain management in particular) in recent years, involving various operational and managerial decisions across the supply chain, such as managers' manufacturing and marketing efforts, suppliers' capacity investment and cost reduction decisions, manufacturers' quality improvement efforts, and buyer's processing and testing efforts; we refer the reader to Porteus and Whang (1991), Baiman et al. (2001), Corbett et al. (2005), Kaya and Özer (2009), Kim et al. (2007), and Kim et al. (2011). In contrast to these papers, which focus on single period settings or steady state analysis that reduces to a static setting, we solve a moral hazard problem with repeated interactions.

Because a multi-period contract can use both immediate compensation and future promises as incentives to induce desired behaviors, it is potentially more powerful than a static contract. The main obstacle to finding an optimal multi-period contract is history dependency. In theory, the optimal contract could compensate a supplier based on its entire performance history and that of competing suppliers (if any). As more performance data becomes available, the information set expands and the computational complexity grows exponentially.

Using formal contracts, Plambeck and Zenios (2000) solve a dynamic moral hazard problem in operations management. They assume that the agent has an exponential utility function and can borrow and lend freely from a bank, which leads to a memoryless optimal contract. Building upon this paper, also assuming the agent's exponential utility and free access to banking, Fuloria and Zenios (2001) study dynamic outcome-adjusted reimbursement for a health-care provider who privately chooses the intensity of treatment in every period, and Plambeck and Zenios (2003) study a make-to-stock queueing system in which the production rate of the server is privately controlled by the agent. Our paper however, presents an incentive structure based on suppliers' full performance history, *without* the aforementioned assumptions. Abreu et al. (1986, 1990) introduce a recursive representation of the dynamic contract using the agent's expected future utility as the state variable, which is then extended by Spear and Srivastava (1987) to the Principal-Agent framework. We use a similar approach to solve a *two-agent* problem with common business/resource constraint, whereas the above papers all solve a single-agent problem.

The literature on relational contracts examines informal contractual agreements between players. In the presence of moral hazard, a relational contract can induce desired actions from the players by the threat of termination of the business relationship or the worst payoffs thereafter if a deviation is caught (the so called "trigger strategies;" see Friedman 1971). Levin (2003) shows that under certain assumptions (risk neutral players, sufficiently high discount factor, etc.), there exists a history-independent, stationary optimal contract, which can be solved as a one-period problem. This result has since been extended to supply chain management by Plambeck and Taylor (2006) and Taylor and Plambeck (2007a,b). In contrast to this approach, we solve a repeated moral hazard problem with risk averse agents and formal contracts, without any restriction on the discount factor.

Relational contracts can also be history dependent, when the players adopt "review strategies" (Radner 1985). Ren et al. (2010) examine a supply chain in which a supplier reviews a demand forecast from a buyer in every period before investing in capacity. If the buyer does not pass the truth-telling test, a limited-time punishment phase follows. They show that truthful information sharing is induced under large discount factors. In comparison, we consider hidden efforts, utilize the entire performance history, and allow any level of discount factor.

Lastly, we note some additional work in supply chain management on multi-period games with hidden information. Zhang et al. (2010) investigate the optimal wholesale contract for a supplier in face of a retailer who carries inventory privately. Oh and Özer (2012) study a supplier's choice between making its own demand forecasts and screening the information from a downstream manufacturer before a capacity investment. The work by Li and Debo (2009a,b) examines the option value of future supplier-switching or second-sourcing of a manufacturer facing uncertain demand when suppliers have private cost information.

Volume Allocation. Many papers on reverse auction or dual sourcing address volume allocations, which is an important aspect of the problem we are studying. Anton and Yao (1989) compare the split-award auction with a winner-take-all auction in a single-stage Nash equilibrium. Klotz and Chatterjee (1995) consider a two-period dual-sourcing model where the buyer reserves a fixed volume share for each supplier and leaves the rest to a competitive bidding in which the lower-cost provider takes all. Seshadri (1995) studies a dual-sourcing model with a cost-plus contest that awards each supplier its actual audited cost plus a fraction of the fixed incentive money. Benjaafar et al. (2007) consider a performance-based proportional allocation mechanism in a single-period model. Cachon and Zhang (2007) compare several performance-based allocation policies that assign incoming jobs to two servers who control their own service rate. They analyze open-loop strategies in steady state and effectively solve a static problem. We extend this research stream by considering *dynamic* volume allocation in an infinite-horizon problem.

A few recent papers examine volume allocation in dynamic environments. Lu and Lariviere (2011) consider a dynamic stochastic game in which a car manufacturer allocates its scarce capacity to its retailers through a fixed (equal) or "turn-and-earn" allocation scheme (which allocates a higher volume to the retailer with more sales). In contrast, we do not assume a particular mathematical form of the allocation policy. Belavina and Girotra (2012) model sourcing decisions with an intermediary and consider business allocations between two suppliers in an infinitely repeated game. They examine cooperative behavior of the suppliers under relational governance whereas we study formal contracts for inducing efforts from *competing* suppliers.

3 Problem Description and Model Formulation

In this section, we formulate the volume allocation problem for a manufacturer facing two substitutable suppliers.

3.1 Problem Description and Assumptions

We consider a single manufacturer sourcing a critical component from two chosen suppliers: Supplier 1 and Supplier 2. Both suppliers are able to meet the minimum cost and quality requirement for the manufacturer. However, the total cost of ownership to the manufacturer could differ between the two suppliers on a number of key measures such as the defect rate, technology innovation, percentage of on-time delivery, etc. The manufacturer constantly evaluates each supplier using these measures and generates an overall rating for the supplier, which serves as a basis for determining business allocations in future time periods. Each supplier, in order to earn more business, has an incentive to expend additional resources to improve the performance outcome (or measure). Such an action can be costly, and does not always work – it only increases the performance outcome probabilistically. From the manufacturer's perspective, additional supplier effort is desirable and ideally the manufacturer would like its suppliers to engage in continuous improvement over the long run. However, the manufacturer needs to provide enough incentive so that a supplier would voluntarily engage in such activities. These incentives could come at a cost to the manufacturer. Therefore, it is not necessarily optimal or feasible to always induce high effort from the suppliers. In this paper, we strive to find the optimal contract that generates the maximal long-run payoffs for the manufacturer.

We make the following assumptions regarding the manufacturer and its suppliers.

(1) The manufacturer is risk neutral and the suppliers are risk averse, which approximates a typical situation with a large buyer and relatively small suppliers.

(2) The transfer price between each supplier and the manufacturer is determined through a cost-plus model. That is, the manufacturer promises to pay each supplier the cost of the component plus a margin r for each unit of the component for an agreed quantity $q_i, i = 1, 2$. In this paper, we focus on the case where the manufacturer uses volume allocation as an incentive lever and thus we treat r as a constant and for simplicity, assume that the two suppliers receive the same margin r. We later relax this assumption and show how the optimal contract may change if the margins are asymmetric (Section 6.1) and how the problem of allocating a total volume is similar to the problem of allocating a total payment (Section 6.2).

(3) In the base model, the total volume to be allocated between the two suppliers is fixed, as the order quantity of a critical part is typically determined by the production plan for the final product. In Section 6.3, we will allow the total volume to deviate from a target level and show that the main insights from the optimal contract stay true with this generalization.

(4) The suppliers are identical with regard to their effort choice options, utility functions, and cost functions, which allows us to focus on the performance differences caused solely by suppliers' efforts. A supplier's utility from the one-period margin rq_i is $\phi(rq_i)$, which is an increasing and concave function and, without loss of generality, satisfies $\phi(0) = 0$. In addition, the supplier's utility is additively separable across time, as is standard in the dynamic contract literature.

(5) The suppliers have two effort choices, "high" and "low," from the set $\mathcal{A} = \{H, L\}$, and their disutility of effort choice $a \in \mathcal{A}$ is $\psi(a)$ (or ψ_a), with $\Delta_\psi = \psi_H - \psi_L > 0$. Treating the disutility of effort a separately from the utility of margin rq_i is standard in the literature, because the cost-of-

effort might not easily translate to a monetary cost. For the performance-enhancing efforts that the suppliers engage in, activities are often process based and therefore only incur fixed costs.[4] We will relax the assumption of binary effort choices in Section 6.4 and demonstrate that the main results remain true.

(6) The suppliers' production functions are independent and the set of possible performance outcomes is $\mathcal{X} = \{0, 1\}$, representing "poor" and "good" outcomes, respectively.[5] We assume that the performance outcomes are public information to the manufacturer and the two suppliers.[6] The probability for outcome $x \in \mathcal{X}$ after a supplier chooses effort $a \in \mathcal{A}$ is $p_a(x)$, which satisfies $p_H(1) > p_L(1)$, i.e., a good outcome is more likely to result from the high effort. We assume that the effort choice in each period directly affects the performance in the current period only. This is often the case with management, maintenance, or operational type of effort, and is arguably the more interesting situation for inducing supplier efforts because incentive must be provided constantly and suppliers cannot sit back and enjoy the lasting effects of their previous efforts.

(7) The value of a supplier's performance outcome $x \in \mathcal{X}$ to the manufacturer is $q \cdot \pi(x)$, where q is the quantity provided by that supplier and $\pi(1) > \pi(0)$. That is, the performance outcome is linked to a per unit dollar value $\pi(x)$.[7]

(8) The manufacturer and the suppliers have the same discount factor $\delta \in (0, 1)$.

3.2 Model Formulation

Now, we formulate the model. In each period t, the manufacturer assigns a quantity q_{it} to supplier i and the supplier privately chooses an effort level $a_{it} \in \mathcal{A}$. The supplier's performance $x_{it} \in \mathcal{X}$ depends on a_{it} through the probabilities $p_{a_{it}}(x_{it})$. Let $h^t = \{(x_{11}, x_{21}), \ldots, (x_{1t}, x_{2t})\}$ denote the suppliers' performance history up to the *end* of period t, and $H^t = (\mathcal{X} \times \mathcal{X})^t$ denote the set of possible h^t's. Supplier i's utility from the quantity q_{it} is $\phi(rq_{it})$ and disutility from the effort is

[4] A general disutility function may also include a variable element which depends on the business volume q_i allocated to a supplier. If the variable element of the disutility function has a linear form cq_i, it can be viewed as part of the variable cost and directly compensated by the manufacturer (see Swinney and Netessine 2009 for a similar argument). Assuming $\psi(a)$ independent of q_i facilitates our analysis and allows us to concentrate on the key trade-offs in motivating suppliers to make high efforts.

[5] It is known that a manufacturer can filter out common industry noise by observing the performance from multiple suppliers (see Holmstrom, 1982; Swinney and Netessine, 2009; and Chen et al., 2011). In this paper, we treat performance outcomes as the outcomes after common noise filtration.

[6] In practice, this is key for inspiring the suppliers and inducing competition. For example, Sun Microsystems Inc. gave each supplier its scorecard results, along with the highest scores of other suppliers in the same commodity area (Farlow et al., 1995); Waste Management Inc. publishes scores of all its suppliers (without disclosing names) to let suppliers see how they performed relative to other vendors (Duffy, 2005).

[7] For example, at Sun Microsystems, if a supplier receives a total score of 86 from the scorecard evaluation, the commodity manager may calculate the Total Cost of Ownership (TCO) for Sun using the formula (100-score)/100+1 and inform the supplier that every dollar Sun spends with the supplier actually costs Sun \$1.14 (Farlow et al., 1995).

$\psi(a_{it})$. Therefore, a dynamic contract can be represented by $\sigma = \{q_{it}(h^{t-1}), a_{it}(h^{t-1})\}_{i=1,2;t=1,\cdots,\infty}$, which defines the strategy profile for the manufacturer and two suppliers. Because suppliers' efforts cannot be observed by the manufacturer, $\{a_{it}(h^{t-1})\}_{t=1,\cdots,\infty}$ can be viewed as the manufacturer's suggested effort plan to supplier i. Notice that q_{it} and a_{it} depend on h^{t-1}, the performance outcomes observed before period t, because the purchase volumes from the suppliers in period t must be determined before entering period t and the suppliers must exert efforts before the outcomes are realized. By default, $h^0 = \emptyset$, representing no initial information. We denote the vectors (q_{1t}, q_{2t}), (a_{1t}, a_{2t}), and (x_{1t}, x_{2t}) by $\mathbf{q}_t$, $\mathbf{a}_t$, and $\mathbf{x}_t$, respectively.

The manufacturer maximizes its total discounted value through the following problem:

$$V = \max_{\{\mathbf{q}_t(\cdot),\mathbf{a}_t(\cdot)\}_{t=1}^{\infty}} \sum_{i=1,2} E\left\{ \sum_{t=1}^{\infty} \delta^{t-1}[\pi(x_{it})q_{it}(h^{t-1}) - rq_{it}(h^{t-1})] \middle| \{\mathbf{a}_t(\cdot)\}_{t=1}^{\infty} \right\} \tag{3.1}$$

$$s.t.\quad E\left\{ \sum_{\tau=t}^{\infty} \delta^{\tau-t}[\phi(rq_{i\tau}(h^{\tau-1})) - \psi(a_{i\tau}(h^{\tau-1}))] \middle| \{\mathbf{a}_\tau(\cdot)\}_{\tau=t}^{\infty}, h^{t-1} \right\} \geq \underline{u}_i,$$
$$h^{t-1} \in H^{t-1},\ t = 1, \ldots, \infty,\ i \in \{1, 2\}, \tag{3.2}$$

$$E\left\{ \sum_{\tau=t}^{\infty} \delta^{\tau-t}[\phi(rq_{i\tau}(h^{\tau-1})) - \psi(a_{i\tau}(h^{\tau-1}))] \middle| \{\mathbf{a}_\tau(\cdot)\}_{\tau=t}^{\infty}, h^{t-1} \right\} \geq$$
$$E\left\{ \sum_{\tau=t}^{\infty} \delta^{\tau-t}[\phi(rq_{i\tau}(\widehat{h}^{\tau-1})) - \psi(\hat{a}_{i\tau}(\widehat{h}^{\tau-1}))] \middle| \{\hat{a}_{i\tau}(\cdot) \in \mathcal{A}, a_{j\tau}(\cdot)\}_{\tau=t}^{\infty}, h^{t-1} \right\},$$
$$h^{t-1} \in H^{t-1},\ t = 1, \ldots, \infty,\ j \neq i \in \{1, 2\}, \tag{3.3}$$

$$\sum_{i=1,2} q_{it}(h^{t-1}) = Q,\ q_{1t}(h^{t-1}) \geq 0,\ q_{2t}(h^{t-1}) \geq 0,\ h^{t-1} \in H^{t-1},\ t = 1, \ldots, \infty. \tag{3.4}$$

Inequality (3.3) for $i \in \{1, 2\}$ (and $j \neq i$) is supplier i's *incentive compatibility* (IC) constraint, which implies that the supplier would voluntarily follow the manufacturer's suggested effort plan, from any period t onward and after any performance history h^{t-1}. Note that the deviated effort plan $\{\hat{a}_{i\tau}(\cdot)\}_{\tau=t}^{\infty}$ would alter the performance path stochastically, and we denote a deviated path after history h^{t-1} by $\{\widehat{h}^{\tau}\}_{\tau=t}^{\infty}$ (assuming $\widehat{h}^{t-1} = h^{t-1}$). Inequality (3.2) for $i \in \{1, 2\}$ is the *participation constraint* for supplier i, which ensures that the supplier would voluntarily participate in the contract, after any performance history h^{t-1}, given its reservation utility $\underline{u}_i$. Expression (3.4) represents a *volume constraint* which requires the total business volume to be fixed and is mathematically akin to the "budget constraint" in the literature. In this infinite-horizon problem, the information set H^{t-1} (i.e., performance history set) grows with t and eventually becomes too large to allow computation of the equilibrium strategy.

3.3 Model Transformation

Abreu et al. (1986, 1990) and Spear and Srivastava (1987) address the computational complexity issue in a repeated game between a principal and a *single* agent by a recursive formulation, which can be extended to the two-agent setting of (3.1)-(3.4). In what follows, we describe the basic idea of this extended approach. Because the future looks exactly the same from any period onward, the subgame following every public history is conceptually identical. It can be easily shown that each agent (supplier)'s expected future utility u_i following any public history can be decomposed into an immediate utility $\phi(rq_i) - \psi(a_i)$ in the current period and a continuation utility U_i from the next period onward, contingent on the random outcome of the current period:[8]

$$u_i = \phi(rq_i) - \psi(a_i) + \delta E[U_i(\mathbf{x}) |\, \mathbf{a}], \quad i = 1, 2. \tag{3.5}$$

Because of the infinite future, the set of feasible continuation utility vectors from any period onward should be identical. That is, the vectors (u_1, u_2) and $(U_1, U_2)(\mathbf{x})$ should all belong to the same continuation utility set. The vector $\mathbf{u} = (u_1, u_2)$ can be interpreted as the state of an (induced) Markov decision process, since the transition from state $\mathbf{u}$ to state $\mathbf{U}$ is determined by the current-period efforts $\mathbf{a}$ stochastically (through the current-period outcomes $\mathbf{x}$).

The recursive formulation reduces the history-dependent contract problem to a dynamic programming problem with a state variable $\mathbf{u}$. Consequently, the problem of searching for the optimal volume allocation contract $\sigma = \{\mathbf{q}_t(h^{t-1}), \mathbf{a}_t(h^{t-1})\}_{t=1,\cdots,\infty}$ is reduced to one of finding the optimal variables $\{\mathbf{a}, \mathbf{q}, \mathbf{U}(\mathbf{x})\}$ for each feasible $\mathbf{u}$.[9] The state variable $\mathbf{u}$ in this stationary representation has dual interpretations. On the one hand, it is a proxy of the suppliers' performance history as from any given initial state, the value of $\mathbf{u}$ at time t is determined by the sequence of performance outcomes $h^{t-1} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{t-1}\}$. On the other hand, u_i represents supplier i's expected future (or continuation) utility. The manufacturer may simply treat it as an equivalent of the supplier's preferential status, and update it in each period with new performance data. Thus, we shall refer to it as the supplier's "rating."

[8] Let h_t^τ denote the performance history from the beginning of period t to the end of period τ, for $\tau \geq t$, i.e., $h_t^\tau = \{\mathbf{x}_t, \ldots, \mathbf{x}_\tau\}$; by default, $h_t^\tau = \emptyset$ if $\tau < t$. Then, h^τ is equivalent to (h^{t-1}, h_t^τ), for $\tau \geq t$. Based on the formulation (3.1)-(3.4), at the beginning of period t after any performance history h^{t-1}, define $u_i(h^{t-1}) = E\left\{\sum_{\tau=t}^{\infty} \delta^{\tau-t}[\phi(rq_{i\tau}(h^{t-1}, h_t^{\tau-1})) - \psi(a_{i\tau}(h^{t-1}, h_t^{\tau-1}))] \,\big|\, \{\mathbf{a}_\tau(\cdot)\}_{\tau=t}^{\infty}, h^{t-1}\right\}$ and $U_i(h^{t-1}, \mathbf{x}_t) = E\left\{\sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1}[\phi(rq_{i\tau}(h^{t-1}, \mathbf{x}_t, h_{t+1}^{\tau-1})) - \psi(a_{i\tau}(h^{t-1}, \mathbf{x}_t, h_{t+1}^{\tau-1}))] \,\big|\, \{\mathbf{a}_\tau(\cdot)\}_{\tau=t+1}^{\infty}, h^{t-1}, \mathbf{x}_t\right\}$. In the backward induction, the past information h^{t-1} plays no explicit role and can be suppressed without loss of generality. Hence, noticing $h^t = (h^{t-1}, \mathbf{x}_t)$, we arrive at the equation $u_i = \phi(rq_{it}) - \psi(a_{it}) + \delta E[U_i(\mathbf{x}_t) |\, \mathbf{a}_t]$.

[9] The vectors $\mathbf{q}$, $\mathbf{a}$, and $\mathbf{U}(\mathbf{x})$ depend on the suppliers' current continuation utility vector $\mathbf{u}$ implicitly, but for notational simplicity, this dependence is suppressed.

Figure 1 shows the sequence of events in the recursive framework. At the beginning of period t, the suppliers' ratings are given by $\mathbf{u}$. The manufacturer announces the volume allocation for the current period and ratings $\mathbf{U}(\mathbf{x})$ for the next period, contingent on the outcomes of the current period. Then the suppliers privately choose effort levels. After delivery, the manufacturer observes the suppliers' performance outcomes and updates their ratings. The game enters the next period.

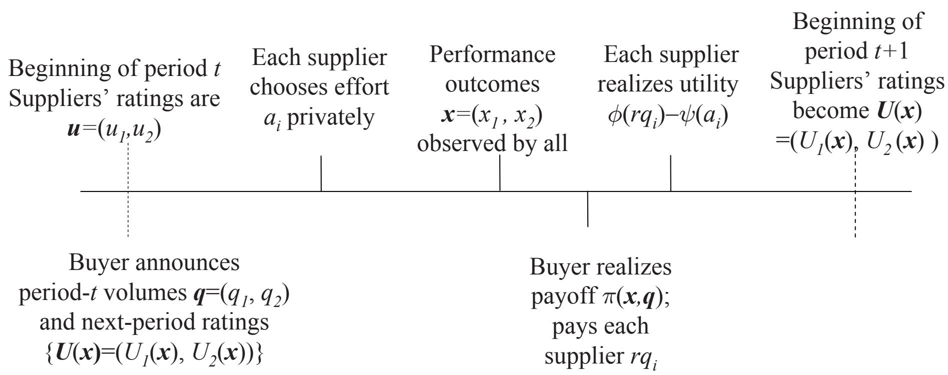


Figure 1: Sequence of Events in Period t under a Dynamic Volume Contract.

Let $V(\mathbf{u})$ be the expected future payoff for the manufacturer given the suppliers' expected future utilities $\mathbf{u} = (u_1, u_2)$. For each feasible $\mathbf{u}$, the manufacturer chooses volumes $\mathbf{q} = (q_1, q_2)$, efforts $\mathbf{a} = (a_1, a_2)$, as well as the suppliers' continuation utilities $\mathbf{U}(\mathbf{x}) = (U_1(\mathbf{x}), U_2(\mathbf{x}))$ to maximize its expected future payoff, provided that the suppliers voluntarily choose $\mathbf{a}$:

$$V(\mathbf{u}) = \max_{\mathbf{a},\mathbf{q},\{\mathbf{U}(\mathbf{x})\}} E[\pi(x_1)q_1 + \pi(x_2)q_2 + \delta V(U_1(\mathbf{x}), U_2(\mathbf{x}))|\, \mathbf{a}] - rQ \tag{3.6}$$

$$\text{s.t.} \quad \phi(rq_i) - \psi_{a_i} + \delta E[U_i(\mathbf{x})|\, \mathbf{a}] = u_i,\ i \in \{1, 2\} \tag{3.7}$$

$$\phi(rq_i) - \psi_{a_i} + \delta E[U_i(\mathbf{x})|\, \mathbf{a}] \geq \phi(rq_i) - \psi_{\widehat{a}_i} + \delta E[U_i(\mathbf{x})|\, \widehat{a}_i, a_j],\ \widehat{a}_i \neq a_i,\ j \neq i \in \{1, 2\} \tag{3.8}$$

$$q_1 + q_2 = Q,\ q_1, q_2 \geq 0. \tag{3.9}$$

Equation (3.7) is the *promise keeping* (PK) constraint, the same as (3.5). Constraints (3.8) and (3.9) are again the *incentive compatibility* (IC) constraint and the *volume constraint*, respectively. This problem is parameterized by $\mathbf{u}$. Both the parameter $\mathbf{u}$ and the decision variables $\{\mathbf{U}(\mathbf{x})\}$ are drawn from the same feasible continuation utility set, say $S \subset \mathbb{R}^2$, and the manufacturer's optimal value function $V(\cdot)$ is determined recursively through the above problem. Our goal is to characterize this function $V : S \to \mathbb{R}$. Note that the original participation constraint (3.2) is equivalent to $\mathbf{u} \geq \underline{\mathbf{u}}$, for a reservation utility vector $\underline{\mathbf{u}}$; we will later normalize $\underline{\mathbf{u}}$ to $\mathbf{0}$ (without loss of generality) and require $\mathbf{u}$ and $\mathbf{U}(\mathbf{x}) \geq \mathbf{0}$, or, $S \subset \mathbb{R}^2_+$.

In general, the optimal value function $V(\cdot)$ may not be concave. However, when randomized contracts are allowed, $V(\cdot)$ must be concave with a convex domain. To see this, suppose that the optimal solutions to the problem given any feasible $\mathbf{u}'$ and $\mathbf{u}''$ are $\{\mathbf{a}', \mathbf{q}', \mathbf{U}'(\mathbf{x})\}$ and $\{\mathbf{a}'', \mathbf{q}'', \mathbf{U}''(\mathbf{x})\}$, respectively. Then the randomized contract that executes $\{\mathbf{a}', \mathbf{q}', \mathbf{U}'(\mathbf{x})\}$ with probability λ and $\{\mathbf{a}'', \mathbf{q}'', \mathbf{U}''(\mathbf{x})\}$ with probability $1-\lambda$ would generate continuation utility vector $\lambda \mathbf{u}' + (1-\lambda)\mathbf{u}''$ for the suppliers and continuation value $\lambda V(\mathbf{u}') + (1-\lambda)V(\mathbf{u}'')$ for the manufacturer. Therefore, the suppliers' continuation utility vector $\lambda \mathbf{u}' + (1-\lambda)\mathbf{u}''$ is feasible and the manufacturer's optimal continuation value at $\lambda \mathbf{u}' + (1-\lambda)\mathbf{u}''$ is at least $\lambda V(\mathbf{u}') + (1-\lambda)V(\mathbf{u}'')$, which implies the concavity of $V(\cdot)$. Randomization is commonly assumed in the repeated game/dynamic contract literature (e.g., Fudenberg and Tirole 1991, Phelan and Stacchetti 2001, Judd et al. 2003, Doepke and Townsend 2006) and is permitted in this paper as well. In essence, the manufacturer may randomly choose among a set of deterministic contracts according to a public lottery (with probabilities dependent on the suppliers' ratings $\mathbf{u}$), which allows the manufacturer to potentially improve its value function.

4 Solving the Dynamic Volume Allocation Problem

The manufacturer's volume allocation problem couples the two suppliers together through the volume constraint (3.9). The manufacturer wishes to create incentives for the suppliers to exert high effort. However, to maintain the total volume, the manufacturer cannot penalize the suppliers simultaneously when their performance outcomes are both poor or reward them at the same time when the outcomes are both good. The manufacturer thus faces an intricate problem of providing the right incentives for the suppliers through dynamic volume allocation. In the following, we discuss step-by-step how to solve for the dynamic contract. Specifically, The problem can be facilitated by four subproblems, given the intended effort pair (H, H), (H, L), (L, H), and (L, L), respectively. We first analyze each subproblem and obtain useful properties of the solution (Section 4.1) and then derive the optimal contract from these subproblems (Section 4.2). For the ease of representation, let $\overline{\pi}_L = E(\pi(x_i)|\, a_i = L)$ and $\overline{\pi}_H = E(\pi(x_i)|\, a_i = H)$.

4.1 Inducing a Given Effort Pair

Given an effort pair (a_1, a_2) to implement, the manufacturer's problem (3.6)-(3.9) reduces to

$$(\Gamma_{a_1a_2}V)(\mathbf{u}) = \max_{\mathbf{q}\in\mathbb{R}^2,\{\mathbf{U}(\mathbf{x})\in S\}_{\mathbf{x}\in\{0,1\}^2}} E[\pi(x_1)q_1 + \pi(x_2)q_2 + \delta V(\mathbf{U}(\mathbf{x}))|\, a_1, a_2] - rQ \tag{4.1}$$

$$\text{s.t.} \quad u_1 = \delta E[U_1(\mathbf{x})|\, a_1, a_2] + \phi(rq_1) - \psi_{a_1} \tag{4.2}$$

$$u_2 = \delta E[U_2(\mathbf{x})|\, a_1, a_2] + \phi(rq_2) - \psi_{a_2} \tag{4.3}$$

$$u_1 \geq \delta E[U_1(\mathbf{x})|\, \widehat{a}_1, a_2] + \phi(rq_1) - \psi_{\widehat{a}_1},\ \widehat{a}_1 \neq a_1 \tag{4.4}$$

$$u_2 \geq \delta E[U_2(\mathbf{x})|\, a_1, \widehat{a}_2] + \phi(rq_2) - \psi_{\widehat{a}_2},\ \widehat{a}_2 \neq a_2 \tag{4.5}$$

$$q_1 + q_2 = Q,\ q_1, q_2 \geq 0. \tag{4.6}$$

This problem implicitly defines a functional operator $\Gamma_{a_1a_2}$, mapping a value function $V : S \to \mathbb{R}$ to another value function $\Gamma_{a_1a_2}V : S_{a_1a_2} \to \mathbb{R}$. Using this operator, the manufacturer's volume allocation problem (3.6)-(3.9) can be succinctly written as

$$V^*(\mathbf{u}) = \max_{(a_1,a_2)\in\{H,L\}^2} (\Gamma_{a_1a_2}V^*)(\mathbf{u}) \tag{4.7}$$

(the superscript "$*$" represents "optimum" throughout this paper). Problem (4.1)-(4.6), given (a_1, a_2), can be simplified by the following results:

Lemma 1. *Given any concave function $V(\cdot)$ and feasible continuation utility vector $\mathbf{u}$, there exists an optimal solution to problem (4.1)-(4.6) such that: (1) if $(a_1, a_2) = (L, L)$, the IC constraints (4.4) and (4.5) do not bind and $U_i(\mathbf{x}) \equiv U_i^*$ for $i \in \{1, 2\}$; (2) if $(a_1, a_2) = (H, L)$, (4.4) binds, (4.5) does not, and $U_i(x_1, 0) = U_i(x_1, 1) = U_i^*(x_1)$, for $i \in \{1, 2\}$ and $x_1 \in \{0, 1\}$; (3) if $(a_1, a_2) = (L, H)$, (4.5) binds, (4.4) does not, and $U_i(0, x_2) = U_i(1, x_2) = U_i^*(x_2)$, for $i \in \{1, 2\}$ and $x_2 \in \{0, 1\}$; (4) if $(a_1, a_2) = (H, H)$, both (4.4) and (4.5) bind.*

The lemma confirms the intuition that to induce high effort from a supplier, the supplier's future utility must be contingent on (in fact, increase with) its performance outcome x_i and its IC constraint should be active.

4.1.1 Inducing Effort Pair (L, L)

By Lemma 1, if $(a_1, a_2) = (L, L)$, problem (4.1)-(4.6) becomes

$$(\Gamma_{LL}V)(\mathbf{u}) = \delta \max_{\mathbf{q}\in\mathbb{R}^2,\mathbf{U}\in S} V(\mathbf{U}) + (\overline{\pi}_L - r)Q \tag{4.8}$$

$$\text{s.t.} \quad u_1 = \delta U_1 + \phi(rq_1) - \psi_L \tag{4.9}$$

$$u_2 = \delta U_2 + \phi(rq_2) - \psi_L \tag{4.10}$$

$$q_1 + q_2 = Q,\ q_1, q_2 \geq 0. \tag{4.11}$$

This problem is relatively straightforward and can be solved directly given any input function $V(\cdot)$.

4.1.2 Inducing Effort Pair (H, L) or (L, H)

We focus on the (H, L) problem below; the (L, H) problem is symmetric and can be analyzed similarly. For $(a_1, a_2) = (H, L)$, problem (4.1)-(4.6) becomes:

$$
\begin{aligned}
(\Gamma_{HL}V)(\mathbf{u}) = & \max_{\mathbf{q}\in\mathbb{R}^2, \{\mathbf{U}(x_1)\in S\}_{x_1\in\{0,1\}}} \{\overline{\pi}_H q_1 + \overline{\pi}_L q_2 + \delta E[V(\mathbf{U}(x_1)) |\, a_1 = H]\} - rQ && (4.12)\\
\text{s.t.}\quad & u_1 = \delta E[U_1(x_1) |\, a_1 = H] + \phi(rq_1) - \psi_H && (4.13)\\
& u_2 = \delta E[U_2(x_1) |\, a_1 = H] + \phi(rq_2) - \psi_L && (4.14)\\
& u_1 = \delta E[U_1(x_1) |\, a_1 = L] + \phi(rq_1) - \psi_L && (4.15)\\
& q_1 + q_2 = Q,\ \ q_1, q_2 \geq 0. && (4.16)
\end{aligned}
$$

Notice that the variables $\mathbf{U}(x_1)$ do not depend on a_2, as shown in Lemma 1. This problem can be decomposed as follows:

Proposition 1. *Problem (4.12)-(4.16) can be solved in two steps: At the lower level, given an **expected** continuation utility vector $\widehat{\mathbf{U}}$ and an input value function $V : S \to \mathbb{R}$, solve*

$$
\begin{aligned}
\widehat{V}_{HL}(\widehat{\mathbf{U}}) = & \max_{\{\mathbf{U}(x_1)\in S\}_{x_1\in\{0,1\}}} E[V(\mathbf{U}(x_1)) |\, a_1 = H] && (4.17)\\
s.t.\quad & U_1(0) = \widehat{U}_1 - p_H(1)\mu, && (4.18)\\
& U_1(1) = \widehat{U}_1 + p_H(0)\mu, && (4.19)\\
& p_H(0)U_2(0) + p_H(1)U_2(1) = \widehat{U}_2, && (4.20)
\end{aligned}
$$

where $\mu = \delta^{-1}\Delta_\psi/(p_H(1) - p_L(1)) > 0$. Let $\widehat{S}_{HL}$ be the feasible parameter set of this problem. At the upper level, given the promised continuation utility vector $\mathbf{u}$ and the above function $\widehat{V}_{HL} : \widehat{S}_{HL} \to \mathbb{R}$, solve

$$
\begin{aligned}
(\Gamma_{HL}V)(\mathbf{u}) = & \max_{\mathbf{q}\in\mathbb{R}^2, \widehat{\mathbf{U}}\in\widehat{S}_{HL}} \{\overline{\pi}_H q_1 + \overline{\pi}_L q_2 + \delta \widehat{V}_{HL}(\widehat{\mathbf{U}})\} - rQ && (4.21)\\
s.t.\quad & u_1 = \delta \widehat{U}_1 + \phi(rq_1) - \psi_H && (4.22)\\
& u_2 = \delta \widehat{U}_2 + \phi(rq_2) - \psi_L && (4.23)\\
& q_1 + q_2 = Q,\ \ q_1, q_2 \geq 0. && (4.24)
\end{aligned}
$$

The upper level problem focuses on the optimal choice of volume allocation $\mathbf{q}$ and the expected continuation utility vector $\widehat{\mathbf{U}}$ (from the next period onward) that render the continuation utility

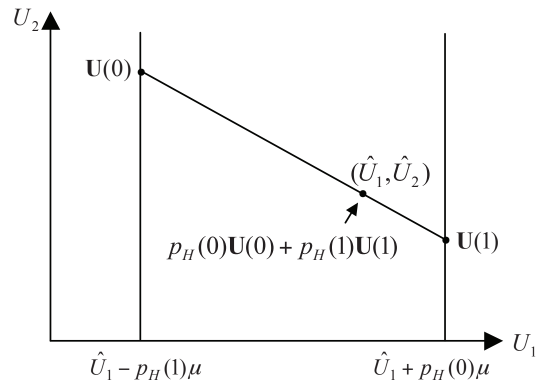


Figure 2: Positions of $\mathbf{U}(0)$ and $\mathbf{U}(1)$ given $\widehat{\mathbf{U}}$.

vector $\mathbf{u}$; while the lower level problem focuses on the optimal choice of the continuation utility vectors $\{\mathbf{U}(x_1)\}$ that yield the expected utility $\widehat{\mathbf{U}}$, subject to supplier 1's incentive compatibility with the high effort. The proposition suggests that in order to motivate supplier 1 to exert high effort, its future compensation must differ substantially based on its performance x_1, i.e., $U_1(1)-U_1(0)=\mu$. Geometrically, as shown in Figure 2, the future utility points $\mathbf{U}(0)$ and $\mathbf{U}(1)$ must lie on the vertical lines with horizontal coordinates $\widehat{U}_1 - p_H(1)\mu$ and $\widehat{U}_1 + p_H(0)\mu$, respectively, and their expectation $p_H(0)\mathbf{U}(0) + p_H(1)\mathbf{U}(1)$ is exactly $\widehat{\mathbf{U}}$.

The lower level problem for any given $\widehat{\mathbf{U}}$ has essentially one free decision variable ($U_2(0)$ or $U_2(1)$) and the upper level problem given $\mathbf{u}$ has also one free decision variable (q_1 or q_2). The challenge comes from the fact that these problems are parameterized and must be solved for all possible $\widehat{\mathbf{U}}$ and $\mathbf{u}$, for a given input function $V(\cdot)$.

4.1.3 Inducing Effort Pair (H, H)

When $(a_1, a_2) = (H, H)$, according to Lemma 1, problem (4.1)-(4.6) becomes

$$(\Gamma_{HH}V)(\mathbf{u}) = \delta \max_{\mathbf{q}\in\mathbb{R}^2, \{\mathbf{U}(\mathbf{x})\in S\}_{\mathbf{x}\in\{0,1\}^2}} E[V(\mathbf{U}(\mathbf{x}))|\, a_1 = H, a_2 = H] + (\overline{\pi}_H - r)Q \tag{4.25}$$

$$\text{s.t.} \quad u_1 = \delta E[U_1(\mathbf{x})|\, a_1 = H, a_2 = H] + \phi(rq_1) - \psi_H \tag{4.26}$$

$$u_2 = \delta E[U_2(\mathbf{x})|\, a_1 = H, a_2 = H] + \phi(rq_2) - \psi_H \tag{4.27}$$

$$u_1 = \delta E[U_1(\mathbf{x})|\, a_1 = L, a_2 = H] + \phi(rq_1) - \psi_L \tag{4.28}$$

$$u_2 = \delta E[U_2(\mathbf{x})|\, a_1 = H, a_2 = L] + \phi(rq_2) - \psi_L \tag{4.29}$$

$$q_1 + q_2 = Q, \;\; q_1, q_2 \geq 0. \tag{4.30}$$

This problem can be decomposed as follows.

Proposition 2. *Problem (4.25)-(4.30) can be solved in two steps: At the lower level, given an* ***expected*** *continuation utility vector* $\widehat{\mathbf{U}}$ *and an input value function* $V : S \to \mathbb{R}$*, solve*

$$\widehat{V}_{HH}(\widehat{\mathbf{U}}) = \max_{\{\mathbf{U}(\mathbf{x}) \in S\}_{\mathbf{x} \in \{0,1\}^2}} E[V(\mathbf{U}(\mathbf{x})) |\, a_1 = H, a_2 = H] \tag{4.31}$$

$$s.t. \quad p_H(0)U_1(0,0) + p_H(1)U_1(0,1) = \widehat{U}_1 - p_H(1)\mu, \tag{4.32}$$

$$p_H(0)U_1(1,0) + p_H(1)U_1(1,1) = \widehat{U}_1 + p_H(0)\mu, \tag{4.33}$$

$$p_H(0)U_2(0,0) + p_H(1)U_2(1,0) = \widehat{U}_2 - p_H(1)\mu, \tag{4.34}$$

$$p_H(0)U_2(0,1) + p_H(1)U_2(1,1) = \widehat{U}_2 + p_H(0)\mu, \tag{4.35}$$

where $\mu = \delta^{-1}\Delta_\psi/(p_H(1) - p_L(1)) > 0$*. Let* $\widehat{S}_{HH}$ *be the feasible parameter set of this problem. At the upper level, given the promised continuation utility vector* $\mathbf{u}$ *and the above function* $\widehat{V}_{HH} : \widehat{S}_{HH} \to \mathbb{R}$*, solve*

$$(\Gamma_{HH}V)(\mathbf{u}) = \delta \max_{\mathbf{q} \in \mathbb{R}^2, \widehat{\mathbf{U}} \in \widehat{S}_{HH}} \widehat{V}_{HH}(\widehat{\mathbf{U}}) + (\overline{\pi}_H - r)Q \tag{4.36}$$

$$s.t. \quad u_1 = \delta\widehat{U}_1 + \phi(rq_1) - \psi_H \tag{4.37}$$

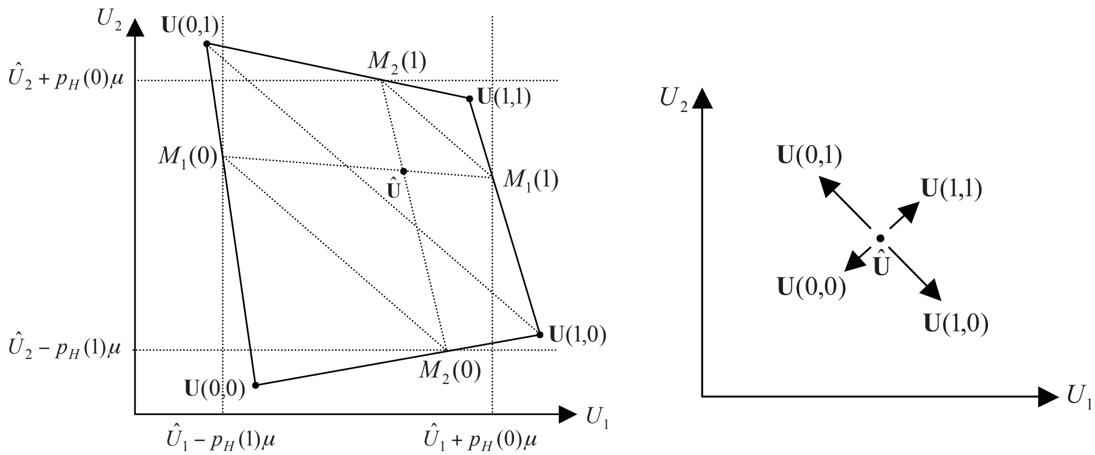
$$u_2 = \delta\widehat{U}_2 + \phi(rq_2) - \psi_H \tag{4.38}$$

$$q_1 + q_2 = Q, \;\; q_1, q_2 \geq 0. \tag{4.39}$$

With four free decision variables, the lower level problem in this case is considerably harder than its counterpart in the (H, L) or (L, H) case. Notice that $E[U_1(1, x_2)|\, a_2 = H] - E[U_1(0, x_2)|\, a_2 = H] = E[U_2(x_1, 1)|\, a_1 = H] - E[U_2(x_1, 0)|\, a_1 = H] = \mu$. Once again, to motivate the suppliers to choose high effort, their future compensation must increase with their individual performance, and the gap between the two scenarios must be sufficiently large. The resulting continuation utility points $\{\mathbf{U}(\mathbf{x})\}_{\mathbf{x} \in \{0,1\}^2}$ also possess strong geometric properties, as summarized below and illustrated in Figure 3(a). Let $l(\overline{N_1N_2})$ denote the length of a line segment $\overline{N_1N_2}$.

Proposition 3. *Given* $\widehat{\mathbf{U}}$*, (1) the points (convex combinations)* $M_1(x_1) = p_H(0)\mathbf{U}(x_1, 0) + p_H(1)\mathbf{U}(x_1, 1)$*,* $x_1 \in \{0, 1\}$*, lie on the vertical lines with horizontal coordinates* $\widehat{U}_1 - p_H(1)\mu$ *and* $\widehat{U}_1 + p_H(0)\mu$*, respectively; (2) the points* $M_2(x_2) = p_H(0)\mathbf{U}(0, x_2) + p_H(1)\mathbf{U}(1, x_2)$*,* $x_2 \in \{0, 1\}$*, lie on the horizontal lines with vertical coordinates* $\widehat{U}_2 - p_H(1)\mu$ *and* $\widehat{U}_2 + p_H(0)\mu$*, respectively; (3) the line segments* $\overline{M_1(0)M_1(1)}$ *and* $\overline{M_2(0)M_2(1)}$ *intersect at* $\widehat{\mathbf{U}}$*; and (4) the line segments* $\overline{M_1(0)M_2(0)}$ *and* $\overline{M_2(1)M_1(1)}$ *are parallel to* $\overline{\mathbf{U}(0,1)\mathbf{U}(1,0)}$*, with lengths* $l(\overline{M_1(0)M_2(0)}) = p_H(1) \cdot l(\overline{\mathbf{U}(0,1)\mathbf{U}(1,0)})$ *and* $l(\overline{M_2(1)M_1(1)}) = p_H(0) \cdot l(\overline{\mathbf{U}(0,1)\mathbf{U}(1,0)})$*.*[10]

[10] Proposition 3 suggests a geometric method to determine points $\{\mathbf{U}(\mathbf{x})\}$ from $\widehat{\mathbf{U}}$: first, freely choose $\mathbf{U}(0,1)$ and $\mathbf{U}(1,0)$; then the points $M_1(0)$, $M_2(0)$, $M_1(1)$, and $M_2(1)$ are uniquely determined according to part (4); finally, $\mathbf{U}(0,0)$ and $\mathbf{U}(1,1)$ are uniquely determined by the expressions of $\{M_i(x_i)\}$ in part (1) or (2).



(a) Geometric structure (b) A common pattern

Figure 3: Positions of $\mathbf{U}(0,0)$, $\mathbf{U}(0,1)$, $\mathbf{U}(1,0)$ and $\mathbf{U}(1,1)$ given $\widehat{\mathbf{U}}$.

The geometric properties reveal a common pattern of the suppliers' continuation utilities, as illustrated in Figure 3(b), and are useful for retrieving structural properties of the optimal contract later.

4.2 Finding Optimal Contract

Now we return to the volume allocation problem (3.6)-(3.9), or equivalently, (4.7).

4.2.1 Suppliers' Continuation Utility Set and Randomized Volume Allocation

The domain of the manufacturer's optimal value function $V^*(\cdot)$ is a subset of $\mathbb{R}^2$. To derive this set, we introduce a set operation. The *Minkowski sum* of two sets Y and Z in an Euclidean space $\mathbb{R}^n$ is the set

$$Y \oplus Z = \{\mathbf{y} + \mathbf{z} : \mathbf{y} \in Y, \mathbf{z} \in Z\}.$$

Consider problem (4.8)-(4.11) of inducing efforts (L, L). Let $S \subset \mathbb{R}^2$ be the domain of the input function $V(\cdot)$ and $S_{LL} \subset \mathbb{R}^2$ be that of the output function $(\Gamma_{LL}V)(\cdot)$. Define the set

$$\begin{aligned} T &= \{(\phi(rq_1), \phi(rq_2)) : q_1 + q_2 = Q,\ q_1, q_2 \in [0, Q]\} \\ &= \{(t_1, t_2) : \phi^{-1}(t_1) + \phi^{-1}(t_2) = rQ,\ t_1, t_2 \in [\phi(0), \phi(rQ)]\}. \end{aligned} \tag{4.40}$$

Every vector $\mathbf{t}$ in T represents the suppliers' utilities from a certain volume allocation $\mathbf{q}$. Using the Minkowski sum operation, constraints (4.9)-(4.11) can be condensed to

$$S_{LL} = (\delta S) \oplus T - (\psi_L, \psi_L). \tag{4.41}$$

The output set S_{LL} so defined may not be convex even if the input set S is convex, because T is a curve in $\mathbb{R}^2$ and is a non-convex set for risk-averse suppliers. However, by the argument at the end of Section 3, when randomization is permitted, problem (4.8)-(4.11) can be modified so that the output domain is convex (and the output function is concave). When the input domain S is convex (and the input function $V(\cdot)$ is concave), which is true under our model, it suffices to randomize over the utility set T because the Minkowski sum of two convex sets is also convex. To that end, denote the convex hull of T by

$$conv(T) = \{\lambda \mathbf{t}' + (1-\lambda)\mathbf{t}'' : \mathbf{t}', \mathbf{t}'' \in T, \lambda \in [0,1]\}. \tag{4.42}$$

Every $\mathbf{t} \in conv(T)\backslash T$ gives the suppliers' *expected utilities* from a *randomized volume allocation* that randomizes between two deterministic allocations $\mathbf{q}'$ and $\mathbf{q}''$. After incorporating randomization, equation (4.41) becomes

$$S_{LL} = (\delta S) \oplus conv(T) - (\psi_L, \psi_L). \tag{4.43}$$

Similarly, randomized contracts are allowed in problems (4.12)-(4.16) and (4.25)-(4.30).

4.2.2 Benchmark Contract: Inducing (L, L) Forever

To always induce effort pair (L, L) is a feasible strategy for the manufacturer and provides a useful benchmark solution to the dynamic volume allocation problem although it may not be optimal. Let $V_{LL}^{\infty}(\cdot)$ be the manufacturer's value function in this solution. It is the fixed point of the operator Γ_{LL} defined in (4.8)-(4.11), i.e., satisfying $(\Gamma_{LL} V_{LL}^{\infty})(\cdot) = V_{LL}^{\infty}(\cdot)$.

This fixed point property has two implications. First, the domain of V_{LL}^{∞}, denoted by S_{LL}^{∞}, is self-generated through (4.8)-(4.11) and hence, by (4.43), satisfies

$$S_{LL}^{\infty} = (\delta S_{LL}^{\infty}) \oplus conv(T) - (\psi_L, \psi_L). \tag{4.44}$$

This equation can be solved through the properties of the Minkowski sum (Gritzmann and Sturmfels, 1993; Zhang, 2010). Second, if we can show that $V_{LL}^{\infty}(\mathbf{u}) \equiv V_{LL}^{\infty}$, it follows immediately that $V_{LL}^{\infty} = \delta V_{LL}^{\infty} + (\overline{\pi}_L - r)Q$. Along these lines, we obtain the following result:

Theorem 1. *Suppose without loss of generality that both suppliers' reservation utility is* 0. *To induce efforts (L, L) forever, the set of suppliers' continuation utility vectors is $S_{LL}^{\infty} = (1-\delta)^{-1}[conv(T) - (\psi_L, \psi_L)] \cap \mathbb{R}_+^2$, and the manufacturer's value function is $V_{LL}^{\infty}(\mathbf{u}) = (1-\delta)^{-1}(\overline{\pi}_L - r)Q$, for any $\mathbf{u} \in S_{LL}^{\infty}$. At any $\mathbf{u} \in S_{LL}^{\infty}$, an optimal choice of $\mathbf{U}$ is $\mathbf{u}$. When $\mathbf{u}$ lies on the upper boundary of S_{LL}^{∞}, denoted by $\overline{S_{LL}^{\infty}}$, this optimal $\mathbf{U}$ is unique and the optimal volume allocation $\mathbf{q}$ satisfies $\phi(rq_1)/\phi(rq_2) = u_1/u_2$.*

The set S_{LL}^{∞} is illustrated by the shaded areas in Figure 4 for $\psi_L = 0$ and $\psi_L > 0$, recalling that $\phi(0) = 0$. The theorem implies that every point $(\mathbf{u}, V_{LL}^{\infty}(\mathbf{u}))$ can be self-generated or self-sustained: On the upper boundary of S_{LL}^{∞}, i.e., for $\mathbf{u} \in \overline{S_{LL}^{\infty}}$, the manufacturer provides the suppliers with the same business volume allocation $\mathbf{q}$ in every period which satisfies $\phi(rq_1)/\phi(rq_2) = u_1/u_2$, and the suppliers' ratings are the same $\mathbf{u}$ forever; If $\mathbf{u} \notin \overline{S_{LL}^{\infty}}$, each point can still be self-generated, but through a randomized volume allocation. Therefore, every point $(\mathbf{u}, V_{LL}^{\infty}(\mathbf{u}))$, for $\mathbf{u} \in \overline{S_{LL}^{\infty}}$, is a "trapping" state and represents a "business as usual" situation: each supplier maintains its status quo (i.e., does not undertake additional effort to improve performance) and the manufacturer simply compensates them according to this status quo and maintains the same volume allocation from period to period. Although good performance can still be observed in this scenario (unless $p_L(1)$ is zero), it is not interpreted as an indication of high effort and the manufacturer does not differentiate good and bad performance observations. As we explain in the following sections, this benchmark scenario serves as an effective long-run incentive, which seems counterintuitive but can be well explained once the longitudinal behavior of the optimal contract is revealed.

4.2.3 Properties of the Optimal Solution

Let S^* denote the domain of the manufacturer's optimal value function $V^*(\cdot)$ and $S_{a_1a_2}^*$ denote the feasible domain of the subproblem of inducing efforts (a_1, a_2) given the input function $V^*(\cdot)$. After incorporating randomized contracts, the volume allocation problem (4.7) implies that

$$S^* = conv(S_{LL}^* \cup S_{HL}^* \cup S_{LH}^* \cup S_{HH}^*), \tag{4.45}$$

where

$$S_{LL}^* = (\delta S^*) \oplus conv(T) - (\psi_L, \psi_L) \tag{4.46}$$

by equation (4.43), and the other $S_{a_1a_2}^*$ can be derived from the upper and lower level problems defined in Propositions 1 and 2.

We characterize the optimal solution along the upper and lower boundaries of S^* by examining the sets $\{S_{a_1a_2}^*\}$. A representative S^* is illustrated in Figure 4, for $\psi_L = 0$ and $\psi_L > 0$. The sets S_{LL}^*, S_{HL}^*, S_{LH}^*, and S_{HH}^* are illustrated in Figure 5, for the numerical example discussed in Section 5 (see Table 1 for the parameters). We denote the upper (lower) boundary of a set S by $\overline{S}$ ($\underline{S}$).

Theorem 2. *The upper boundary of S^* coincides with the upper boundaries of S_{LL}^* and S_{LL}^{∞}, and the manufacturer's optimal value $V^*(\mathbf{u}) = (1-\delta)^{-1}(\overline{\pi}_L - r)Q$ for any $\mathbf{u} \in \overline{S^*}$. The optimal solution at any $\mathbf{u} \in \overline{S^*}$, including the volume allocation and next period ratings, is identical to that in Theorem 1.*

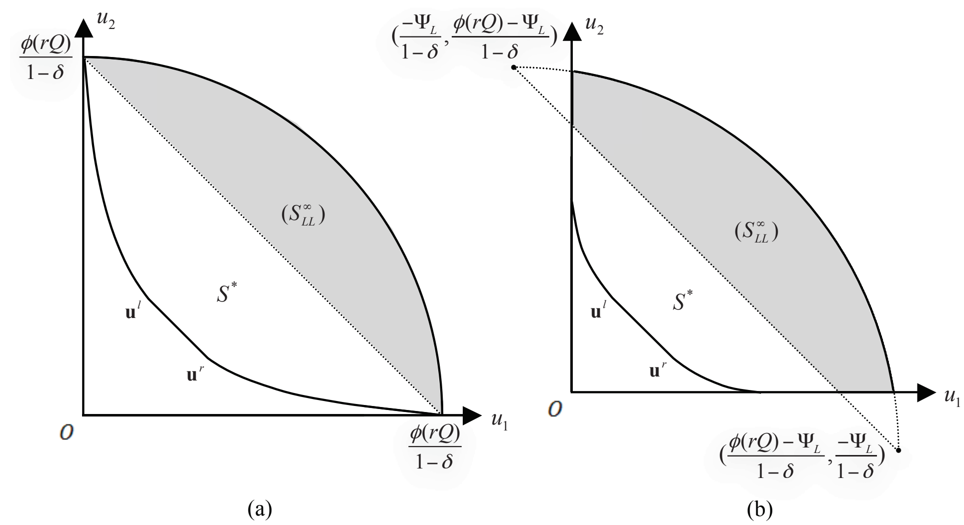


Figure 4: A typical S^* for (a) $\psi_L = 0$ and (b) $\psi_L > 0$.

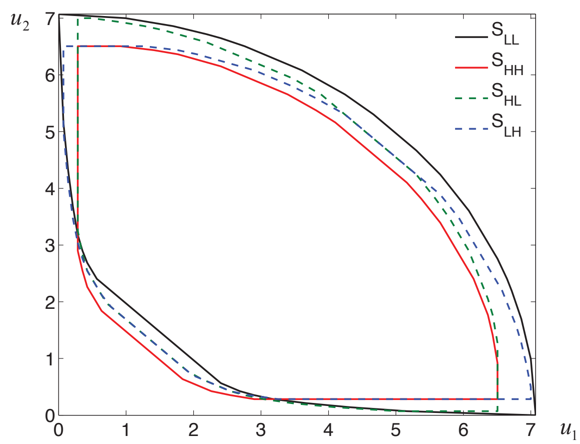


Figure 5: An example of sets S^*_{LL}, S^*_{HL}, S^*_{LH}, and S^*_{HH}.

The theorem identifies a set of self-generated points along the upper boundary of S^*. That is, if the suppliers' continuation utility vector $\mathbf{u}$ enters $\overline{S^*}$, it will be "trapped" there forever. Next, we examine the lower boundary of S^*.

Theorem 3. *(1) If $\frac{\psi_L}{\psi_H} \leq \frac{p_L(1)}{p_H(1)}$ and $\phi(rQ) \geq 2(1-\delta p_H(0))\mu$, the center of the lower boundary of S^* is a $-45°$ line segment self-generated under the (H,H) effort pair, with end points $\mathbf{u}^l = (1-\delta)^{-1}(\delta p_H(1)\mu - \psi_H, -\delta p_H(1)\mu + \phi(rQ) - \psi_H)$ and $\mathbf{u}^r = (1-\delta)^{-1}(-\delta p_H(1)\mu + \phi(rQ) - \psi_H, \delta p_H(1)\mu - \psi_H)$.*

(2) If $\frac{\psi_L}{\psi_H} > \frac{p_L(1)}{p_H(1)}$ and $\phi(rQ) \geq 2((1-\delta)\mu + \psi_H)$, the lower boundary of S^* is a $-45°$ line segment self-generated under the (H,H) effort pair, with end points $\underline{\mathbf{u}}^l = (1-\delta)^{-1}(0, \phi(rQ) - 2\psi_H)$ and $\underline{\mathbf{u}}^r = (1-\delta)^{-1}(\phi(rQ) - 2\psi_H, 0)$.

(3) In the above cases, for any $\mathbf{u} \in \overline{\mathbf{u}^l\mathbf{u}^r}$ *or* $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$, *the manufacturer's optimal value* $V^*(\mathbf{u}) = (1-\delta)^{-1}(\overline{\pi}_H - r)Q$ *and the optimal volume allocation* $\mathbf{q}$ *is randomized between* $(0, Q)$ *and* $(Q, 0)$.

Part (3) implies that the manufacturer can achieve the highest possible (first-best) expected value $(1-\delta)^{-1}(\overline{\pi}_H - r)Q$ by keeping the suppliers' ratings in the line segment $\overline{\mathbf{u}^l\mathbf{u}^r}$ (or $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$) and inducing both of them to exert high effort; the line segment $\overline{\mathbf{u}^l\mathbf{u}^r}$ is labeled in Figure 4. The conditions in parts (1) and (2) of the theorem are sufficient but not necessary. They enable sufficient variations in the suppliers' future utilities for incentive provision [11] and can be easily met when the (possible) reward is sufficiently high (e.g., high Q, r, or δ) and/or the cost of the high effort (ψ_H) is sufficiently low. [12]

Although the lower boundary of S^* is also generated from points on the lower boundary, no individual point on $\underline{S^*}$ can be a trapping point as those on the upper boundary, because, to provide incentive for efforts $(a_1, a_2) \neq (L, L)$, a utility vector $\mathbf{u} \in \underline{S^*}$ must be generated from at least two distinct points in the feasible domain to reward a good outcome and punish a bad one. However, as shown below and illustrated in Figure 6, the suppliers' continuation utilities can still be "locally trapped" on the lower boundary, i.e., confined to a closed line segment which forms a "recurrent" class of the induced Markov process.

Proposition 4. *Let* $\widetilde{\mathbf{u}}^l = \mathbf{u}^l + (\mu, -\mu)$ *and* $\widetilde{\mathbf{u}}^r = \mathbf{u}^r + (-\mu, \mu)$. *In the first case of Theorem 3, there exists an optimal solution such that (1) for any* $\mathbf{u} \in \overline{\widetilde{\mathbf{u}}^l\widetilde{\mathbf{u}}^r}$, $\mathbf{U}(0,0) = \mathbf{U}(1,1) = \mathbf{u}$, $\mathbf{U}(0,1) = \mathbf{u} + (-\mu, \mu)$, *and* $\mathbf{U}(1,0) = \mathbf{u} + (\mu, -\mu)$; *(2) for any* $\mathbf{u} \in \overline{\mathbf{u}^l\widetilde{\mathbf{u}}^l}$, $\mathbf{U}(0,0) = \mathbf{U}(0,1) = \mathbf{u}^l$, $\mathbf{U}(1,1) = \mathbf{u}^l + \frac{p_H(1) - p_H(0)}{p_H(1)}(\mu, -\mu)$, *and* $\mathbf{U}(1,0) = \mathbf{u}^l + (2\mu, -2\mu)$; *and (3) for any* $\mathbf{u} \in \overline{\widetilde{\mathbf{u}}^r\mathbf{u}^r}$, $\mathbf{U}(0,0) = \mathbf{U}(1,0) = \mathbf{u}^r$, $\mathbf{U}(1,1) = \mathbf{u}^r + \frac{p_H(1) - p_H(0)}{p_H(1)}(-\mu, \mu)$, *and* $\mathbf{U}(0,1) = \mathbf{u}^r + (-2\mu, 2\mu)$. *In the second case of Theorem 3, there exists an optimal solution similar to the above, with* $\mathbf{u}^l$, $\mathbf{u}^r$, $\widetilde{\mathbf{u}}^l$, *and* $\widetilde{\mathbf{u}}^r$ *replaced by* $\underline{\mathbf{u}}^l$, $\underline{\mathbf{u}}^r$, $\widetilde{\underline{\mathbf{u}}}^l = \underline{\mathbf{u}}^l + (\mu, -\mu)$, *and* $\widetilde{\underline{\mathbf{u}}}^r = \underline{\mathbf{u}}^r + (-\mu, \mu)$, *respectively.*

The proposition reveals an interesting and intuitive solution for the manufacturer. Once the suppliers' ratings fall into the middle section of the trapping segment $\overline{\mathbf{u}^l\mathbf{u}^r}$ on the lower boundary, the manufacturer can keep the suppliers on their toes through the following "tournament": when

[11] In case (1), the distance between the two end points $\mathbf{u}^l$ and $\mathbf{u}^r$ is given by $(1-\delta)^{-1}(\phi(rQ) - 2\delta p_H(1)\mu)$ along both axes. Thus the condition $\phi(rQ) \geq 2(1 - \delta p_H(0))\mu$ implies that these two points are at least 2μ apart along both axes. The assumption $\frac{\psi_L}{\psi_H} \leq \frac{p_L(1)}{p_H(1)}$ is equivalent to $\delta p_H(1)\mu \geq \psi_H$ and thus $u_1^l = u_2^r \geq 0$. In case (2), the line segment $\overline{\mathbf{u}^l\mathbf{u}^r}$ is truncated by the two axes to $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$. Since the distance between $\underline{\mathbf{u}}^l$ and $\underline{\mathbf{u}}^r$ is given by $(1-\delta)^{-1}(\phi(rQ) - 2\psi_H)$, the assumption $\phi(rQ) \geq 2((1-\delta)\mu + \psi_H)$ similarly ensures that $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$ is long enough for incentive provision.

[12] For example, when δ is close to 1, the main assumption in case (2), $\phi(rQ) \geq 2((1-\delta)\mu + \psi_H)$, is approximately $\phi(rQ) \geq 2\psi_H$, which is necessary to just cover the disutility of high effort for the two suppliers (under randomized volume allocation).

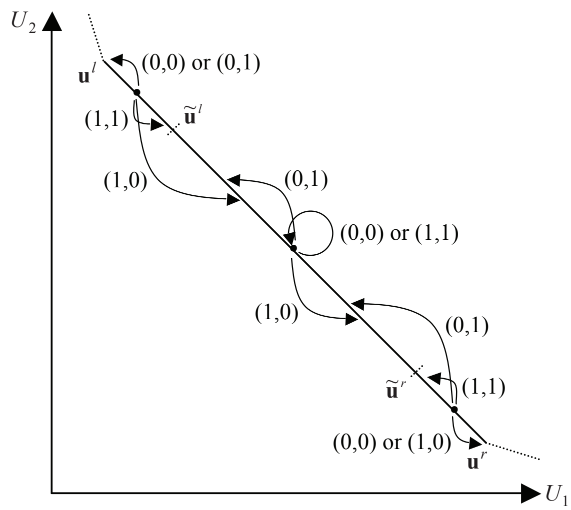


Figure 6: Local Trapping on the Lower Boundary.

one supplier performs better than the other (i.e., the outcome vector is either $(0,1)$ or $(1,0)$), promote the former supplier and demote the latter; if they perform equally well or equally poor (with outcome vector $(0,0)$ or $(1,1)$), keep their ratings unchanged. This strategy highlights the role of competition in motivating suppliers. When the suppliers' ratings move too close to one end of the trapping segment, i.e., into $\overline{\mathbf{u}^l\widetilde{\mathbf{u}}^l}$ or $\overline{\widetilde{\mathbf{u}}^r\mathbf{u}^r}$, the above tournament becomes non-sustainable and the manufacturer's strategy needs to be modified: for example, the manufacturer should punish poor performance by the lower-rated supplier even if the competing supplier performs equally poor.

4.2.4 State Evolution under the Optimal Contract

Our solution approach to the repeated moral hazard problem rests upon the idea that the suppliers' rating vector evolves as a Markov decision process. Now, we examine the longitudinal behavior of this process, as summarized in Figure 7. Theorems 1, 2, and 3 reveal that trapping and recurrent class of states may exist in this Markov decision process. From Theorems 1 and 2, there are infinitely many individual "trapping" states on the upper boundary of S^*. Each trapping state represents a "business-as-usual" (low effort) scenario with a characteristic volume allocation determined by the ratio of the two suppliers' ratings. Theorem 3 identifies a "recurrent" class on the lower boundary of S^* under certain conditions. This subset is characterized by high effort from both suppliers and highest value achieved for the manufacturer. From the manufacturer's perspective, this is the most desirable situation. The suppliers however, experience the most intense competition in these states.

Any point from which $\overline{S^*}$ can be reached with a positive probability is a "transient" state. Similarly, when the recurrent class exists on the lower boundary, any point from which a recurrent

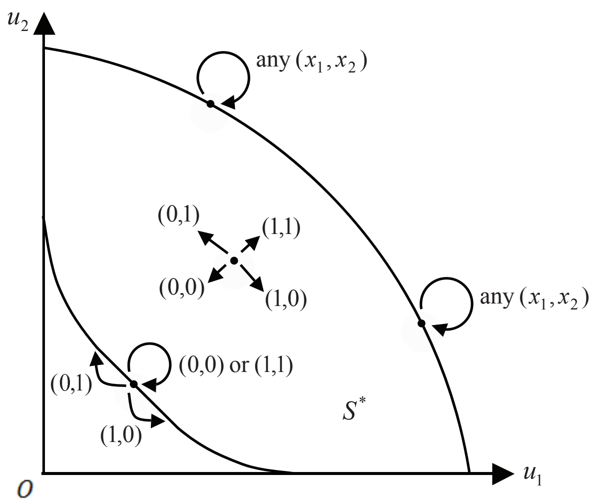


Figure 7: A representative pattern of state evolution under the optimal contract.

state may be reached is also transient. In fact, the majority of the feasible domain S^* comprises of these transient states. Because the analytical solution is intractable in the interior of S^*, a complete characterization of transient states is infeasible. However, Propositions 1-3 and Figures 2-3 shed light on incentive provision and state transitions at those points. In general, the power of incentives at the transient points is less intense than on the lower boundary but stronger than on the upper boundary. More properties of the optimal solution at the interior of S^* are explored numerically in Section 5.

Consider an initial state that is transient, as the point in the interior of S^* in Figure 7. Over time, the state transitions as performance outcomes are observed. The overall trend of such transitions is illustrated in the figure (see also Figure 3(b)). Eventually the state becomes trapped, either to a point on the upper boundary or to some recurrent states on the lower boundary. For instance, the line segment described in Theorem 3 can be reached after both suppliers under-perform for some extended time, in which case the punishment for the suppliers is exactly what characterizes this recurrent class: high effort, intense competition, and low expected payoff. The upper boundary is reached after both or at least one supplier over-perform for some extended time. In this case, the exact location where it is trapped makes a huge difference for the suppliers. As shown in Theorem 1, each point on the upper boundary has a characteristic volume allocation which changes from $0:1$ on one end to $1:0$ on the other, and is solely determined by the ratio of the suppliers' ratings $u_1 : u_2$. Ideally, a supplier prefers the trapping to occur at a location that yields a higher volume for itself (since that volume allocation will persist in all future periods), which provides incentive for the supplier to continually exert high effort in order to influence the direction of the state transition.

In summary, the trapping states on the upper boundary of S^* and the recurrent states on the lower boundary are long-run incentive drivers, as the "carrot" or "stick", for the suppliers to work hard.

The results resonate with some known results in the repeated game literature. The "trapping" states on the upper boundary are reminiscent of the Nash equilibria in a static game in which the manufacturer allocates volume between two suppliers to match each supplier's promised utility. The "recurrent" states on the lower boundary bear some resemblance to the punishment threat in a "trigger strategy" in repeated games (Friedman 1971, Levin 2003, Plambeck and Taylor 2006). While punishment often involves termination of the cooperation and is thus the worst equilibrium for all players, under the optimal contract in our model, the recurrent states impose intense competition and low payoff for the suppliers but result in high effort input and the first-best value to the manufacturer, i.e., they are "punishment" to the suppliers but not to the manufacturer.

5 Numerical Analysis

To further characterize the optimal contract, we resort to numerical analysis. For simplicity, we assume the utility function $\phi(w) = \sqrt{w}$, for $w \geq 0$, but the results can be generalized to other concave utility functions. We examine the optimal solution for a representative example, including the suppliers' continuation utility set, effort choices, and allocated volumes, as well as the manufacturer's value function. We also study the longitudinal evolution of the suppliers' ratings.

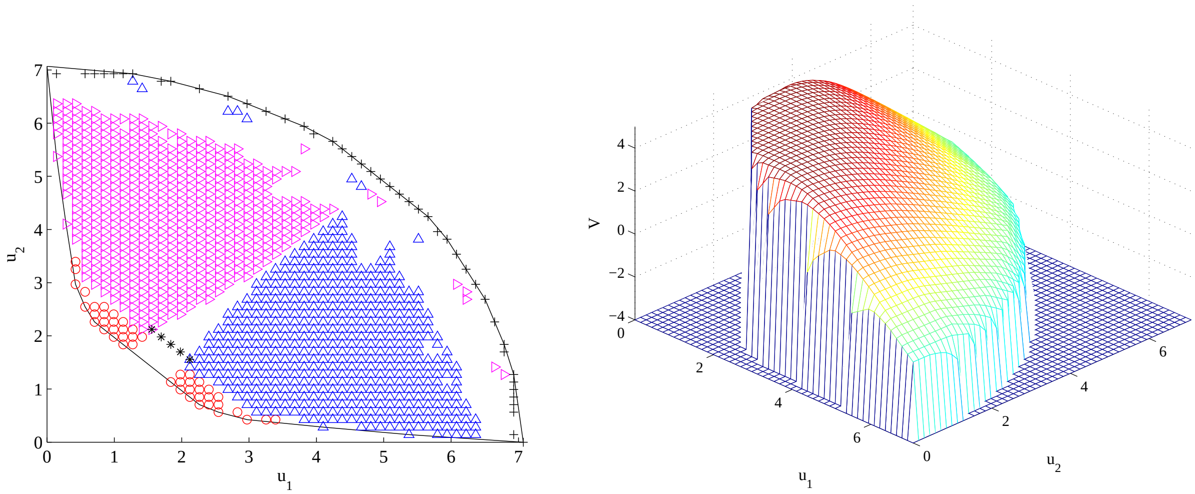
Since we have already provided analytical characterizations of the optimal contract under the conditions given in Theorem 3, in the numerical analysis, we explore the case when such conditions are *not* met. In particular, we consider the example given in Table 1 (the total volume Q is normalized to 1). The results are presented in Figures 8 and 9.[13] We have also conducted a comparative statics analysis, by varying the parameters $p_H(1), p_L(1), \psi_H, \psi_L, \bar{\pi}_H, \bar{\pi}_L, r$, and δ, to verify that the numerical findings are robust; due to space limitation, those results are omitted here but are available from the authors.

Parameter	Q	r	δ	$p_H(1)$	$p_L(1)$	ψ_H	ψ_L	$\bar{\pi}_H$	$\bar{\pi}_L$
Value	1	0.5	0.9	0.7	0.3	0.3	0	1	0.1

Table 1: Parameter Values for the Example

Manufacturer's Optimal Value Function. The domain S^* and function $V^*(\cdot)$ are illustrated

[13]We first identify the minimum and maximum values of each supplier's rating u_i using the results in Theorem 2. We then discretize this interval into 50 points and iteratively search for the two-dimensional self-generating domain S^*, whose upper boundary is specified exactly in Theorem 2 but the lower boundary has to be identified computationally. Next, based on the obtained domain S^* and the benchmark value V_{LL}^{∞} identified in Theorem 1, we iteratively construct the value function $V^*(\cdot)$ through the decomposed problems defined in Propositions 1 and 2.



(a) Domain S^* and Pure Strategy Points

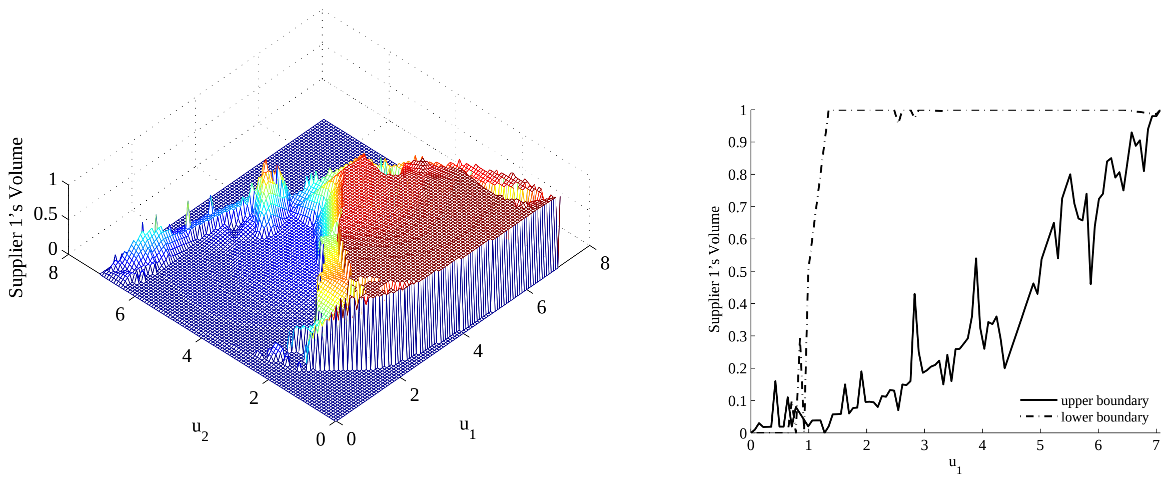
(b) Optimal Continuation Values

Figure 8: Manufacturer's Optimal Value Function $V^*(\cdot)$.

in Figures 8(a) and 8(b), respectively. By the randomized version of equation (4.7), $V^*(\cdot)$ is formed from the upper convex hull of the individual functions $(\Gamma_{a_1a_2}V^*)(\cdot)$, for $a_1a_2 \in \{LL, HL, LH, HH\}$. Hence the vertices on the surface of $V^*(\cdot)$, identified in Figure 8(a), correspond to (the manufacturer's) pure strategies and the non-vertex points in the blank spaces represent mixed strategies. The upper boundary of S^* consists entirely of points with optimal effort pair (L, L), marked with "$+$"; from Theorem 2, we know that these points are "trapping". The lower boundary of S^* consists of points with optimal effort pair (H, H) (marked with "o"), (H, L) (marked with "$\triangle$"), and (L, H) (marked with "$\triangleright$"). In contrast to the cases identified in Theorem 3, the highest manufacturer values, marked with "$*$" in Figure 8(a), are not located *on* the lower boundary, but at the intersection of the three regions where $\mathbf{a} = (H, H)$, (H, L), and (L, H) respectively dominate and near the 45° line. This implies that the manufacturer's value is higher if the suppliers are symmetric with respect to their ratings, and that the highest manufacturer values are achieved with a randomized contract which implements (H, H), (H, L) and (L, H) probabilistically.

Optimal Volume Allocation. As shown in Theorem 1, each point on the upper boundary of S^* (or S^{∞}_{LL}) corresponds to a specific volume allocation, which changes continuously with the ratio u_1/u_2 (such that $\phi(rq_1)/\phi(rq_2) = u_1/u_2$). Figure 9(a) shows supplier 1's volume allocation under the optimal contract over the entire domain S^* (supplier 2's volume is symmetric).[14] Clearly, higher value of u_1 results in higher business volume for supplier 1. The volume drops markedly as

[14]Noise along the upper boundary is due to the computation precision and the fact that the lower-level optimization problems have an objective function that is rather flat near the optimal point.



(a) Across the Feasible Region

(b) On the Boundaries

Figure 9: Volume Allocated to Supplier 1.

the state moves from the area where supplier 1 is stronger (i.e., with a higher rating) to the area where supplier 2 is stronger. We observe that the trend of the volume allocation is interrelated with the optimal effort choices (a_1, a_2). For instance, the allocation along the upper boundary where (L, L) dominates behaves quite differently from the lower boundary where (H, H), (H, L) or (L, H) dominates. Although on both the upper and lower boundaries, the optimal volume allocation for supplier 1 follows a upward trend and changes from 0 to 1 as u_1 increases from its minimum value to its maximum (Figure 9(b)), the change is much more drastic on the lower boundary, where at least one supplier chooses high effort.

State Evolution. The suppliers' ratings form a set of Markov states and evolve over time. We simulate the state path from different starting states, which helps shed light on the behaviors of the transient states between the upper and lower boundaries. We observe in this example that trapping is inevitable and it always occurs on the upper boundary, which is reasonable since the conditions for the recurrent class identified in Theorem 3 are not met. As discussed in Section 4.2, being trapped at a particular point (on the upper boundary) implies that the future "business norm" is represented by a characteristic volume allocation, which serves as the ultimate long-run incentive/disincentive for continuous supplier improvement. Our simulation reveals that the time it takes to reach a trapping state varies with the starting state and so does the exact location where trapping occurs. In particular, when the initial state is farther away from the upper boundary, it takes longer to reach trapping and the initial state (or, the initial ratings of the suppliers) has a weaker impact on the final trapping location.

6 Extensions

In the base model studied in previous sections, we have made some assumptions that simplify our analysis. In this section, we demonstrate that our main results still hold if some of these assumptions are relaxed or altered. We highlight the main findings here and defer the details to Appendix B.

6.1 Asymmetric Suppliers

The basic model (3.6)-(3.9) assumes that the two suppliers are symmetric, with regard to their utility functions, cost functions, unit margins, value contributions, etc. This assumption allows us to concentrate on the most valuable circumstances for dynamic volume allocation. Suppose, for example, the suppliers' unit margins are unequal. Then the manufacturer would tend to allocate less volume to the supplier demanding the higher margin, diminishing the power of volume incentive. Nevertheless, as discussed below, the main results of this paper can be extended to the setting of unequal supplier margins (asymmetries in utility and cost functions can be accommodated similarly).

Suppose supplier i's unit margin is r_i, $i = 1, 2$. The manufacturer's problem (4.1)-(4.6) needs slight modifications – replacing the term rQ in the objective function by $r_1q_1 + r_2q_2$, and replacing the terms rq_1 and rq_2 in the constraints by r_1q_1 and r_2q_2, respectively. It is straightforward to verify that Lemma 1 is still valid and results in Section 4.1 are slightly modified as above.

The set T of one-period utility vectors from deterministic volume allocations, defined in expression (4.40), changes to:

$$\begin{aligned} T &= \{(\phi(r_1q_1), \phi(r_2q_2)) : q_1 + q_2 = Q,\ q_1, q_2 \in [0, Q]\} \\ &= \{(t_1, t_2) : \frac{\phi^{-1}(t_1)}{r_1} + \frac{\phi^{-1}(t_2)}{r_2} = Q,\ t_1 \in [\phi(0), \phi(r_1Q)], t_2 \in [\phi(0), \phi(r_2Q)]\}. \end{aligned} \tag{6.1}$$

As an example, if the utility function is $\phi(w) = \sqrt{w}$, i.e., $\phi^{-1}(t) = t^2$, the new set T would be the north-east quarter of an ellipse with radiuses $\sqrt{r_1Q}$ and $\sqrt{r_2Q}$, as opposed to the circle with radius $\sqrt{rQ}$ in the equal margin case. Equations (4.41) to (4.46) still hold true, and Theorems 1 and 2 only incur minor modifications. The upper boundary of S^∞_{LL} or S^* is still given by $(1 - \delta)^{-1}[conv(T) - (\psi_L, \psi_L)] \cap \mathbb{R}^2_+$, and the optimal volume allocation $\mathbf{q}$ on this boundary is still unique (satisfying $\phi(r_1q_1)/\phi(r_2q_2) = u_1/u_2$), but the manufacturer's expected value function, now given by $V^\infty_{LL}(\mathbf{u}) = (1 - \delta)^{-1}(\overline{\pi}_L Q - r_1q_1 - r_2q_2)$, is not flat any more because the total margin payment $r_1q_1 + r_2q_2$ is not constant. The properties of the optimal solution along the lower boundary of S^*, characterized by Theorem 3 and Proposition 4, can also be generalized except that the slope of the line segment $\overline{\mathbf{u}^l\mathbf{u}^r}$ (or $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$), is no longer $-45°$ when the margins differ and the manufacturer's expected value along that line segment now varies linearly between $V^*(\mathbf{u}^l)$ and $V^*(\mathbf{u}^r)$. Lastly, the longitudinal behavior on the upper and lower boundaries stays unchanged.

6.2 Fixed Total Payment

In the base model of the paper, the unit margin for each supplier is a constant r, and the manufacturer allocates a fixed total volume Q between the suppliers in every period. In this extension, we consider the "opposite" problem, in which the business volume allocated to each supplier is constant at q, and the manufacturer has a fixed total payment W to allocate in each period. The key difference between the two problems lies in the timing of the critical events. Business volumes are usually determined at the beginning of a period, while the payments are often made at the end and thus can be contingent on the performance outcome of that period. Nevertheless, a careful choice of the reference point can suppress this contingency and streamline the latter problem.

We call the time point (in each period) at which the performance outcomes and the manufacturer's payoff have been realized but the payments to the suppliers are yet to be made the *compensation point*. Let $\mathbf{u} = (u_1, u_2)$ be the continuation utility vector promised to the suppliers from the compensation point of the current period onward and $V(\mathbf{u})$ be the manufacturer's corresponding continuation payoff from the compensation point onward (without the current-period payoff). Given $\mathbf{u}$, the manufacturer chooses the current-period payments $\mathbf{w} = (w_1, w_2)$, next-period efforts $\mathbf{a} = (a_1, a_2)$, as well as the suppliers' continuation utilities $\mathbf{U}(\mathbf{x}) = (U_1(\mathbf{x}), U_2(\mathbf{x}))$ (contingent on the next-period performance outcomes $\mathbf{x}$) to maximize its expected value, subject to promise keeping, incentive compatibility, and total payment constraints:

$$V(\mathbf{u}) = \max_{\mathbf{w},\mathbf{a},\{\mathbf{U}(\mathbf{x})\}} E[\delta\pi(x_1)q + \delta\pi(x_2)q + \delta V(U_1(\mathbf{x}), U_2(\mathbf{x}))|\, \mathbf{a}] - W \tag{6.2}$$

$$\text{s.t.} \quad \phi(qw_i) - \delta\psi_{a_i} + \delta E[U_i(\mathbf{x})|\, \mathbf{a}] = u_i, \ i \in \{1, 2\} \tag{6.3}$$

$$\phi(qw_i) - \delta\psi_{a_i} + \delta E[U_i(\mathbf{x})|\, \mathbf{a}] \geq \phi(qw_i) - \delta\psi_{\widehat{a}_i} + \delta E[U_i(\mathbf{x})|\, \widehat{a}_i, a_j], \ \widehat{a}_i \neq a_i, \ j \neq i \in \{1, 2\} \tag{6.4}$$

$$w_1 + w_2 = W/q, \ w_1, w_2 \geq 0. \tag{6.5}$$

The problem is similar to the volume allocation problem (3.6)-(3.9); so the function $V(\mathbf{u})$ and the corresponding optimal contract possess similar properties. The only additional task is to decide for period 1 the optimal effort vector $\mathbf{a}$ and continuation utility vectors $\{\mathbf{U}(\mathbf{x})\}$ (contingent on period 1's outcomes), given an initial state $\mathbf{u}^0 = (u_1^0, u_2^0)$; it is a simple one-shot problem and does not affect the long-term properties of the optimal contract governed by the recursive problem above.

6.3 Flexible Total Volume

In the base model, the manufacturer's total business volume is a constant Q in every period. In this extension, we allow the total volume to vary in an interval, $[Q_m, Q_M]$. We assume that the manufacturer has a target volume $Q_0 \in [Q_m, Q_M]$ and incurs over and under-order penalties. The manufacturer's total cost of procuring Q units is given by $g(Q) = \begin{cases} rQ + \beta_m(Q_0 - Q), & \text{if } Q \in [Q_m, Q_0), \\ rQ + \beta_M(Q - Q_0), & \text{if } Q \in [Q_0, Q_M], \end{cases}$ for some nonnegative coefficients β_m and β_M. When $\beta_m = \beta_M = \infty$, the model reduces to the base model with a fixed total volume Q_0; when $\beta_m = \beta_M = 0$, the model reduces to one without a target volume. To avoid trivial cases, we assume $\overline{\pi}_L < r + \beta_M$, i.e., increasing the total volume beyond Q_0 is not profitable for the manufacturer at least in the low effort scenario; otherwise, the manufacturer would be tempted to push the total volume all the way to Q_M.

The manufacturer's problem (4.1)-(4.6) of inducing a given effort pair (a_1, a_2) only undergoes minor modifications: the manufacturer's total payment rQ in the objective function is replaced by $g(q_1 + q_2)$, and the volume constraint $q_1 + q_2 = Q$ is replaced by $q_1 + q_2 \in [Q_m, Q_M]$. It can be verified that Lemma 1 is intact. Thus, the problems of inducing (L, L), (H, L), (L, H), and (H, H) effort pairs are all similar as before except the above modifications. As a result, the decomposition of these problems is still valid, i.e., Propositions 1 and 2 are still true except for the necessary changes in the objective functions and volume constraints in the upper level problems. Propositions 3 and 4 carry over without any modification. The robustness of these results reveals that the fundamental incentive driver in the problem is unchanged under this generalization.

The flexibility in Q broadens the manufacturer's choices, which enlarges the feasible set of the suppliers' continuation utilities and improves the manufacturer's value function. Due to such changes, Theorems 1, 2, and 3 need to be modified; most notably, the trapping region near the upper boundary of the feasible set S^* and the recurrent set near the lower boundary are both enlarged as a result of the flexibility in Q. Being able to dynamically allocate a larger (as well as smaller) volume makes it easier for the manufacturer to induce high effort from the suppliers. A rigorous analysis can be found in Appendix B.

6.4 Multiple Effort Levels

In the base model, the suppliers' effort level can be either H or L. In this extension, we add an intermediate level, M. More effort levels can be treated similarly.

As in the two-effort-level case, assume that the disutilities of the effort levels and corresponding probabilities of the good outcome are ordered such that $\psi_H > \psi_M > \psi_L$ and $p_H(1) > p_M(1) >$

$p_L(1)$. Define the effective marginal costs of effort as $\mu_{HM} = \delta^{-1}(\psi_H - \psi_M)/(p_H(1) - p_M(1))$, $\mu_{ML} = \delta^{-1}(\psi_M - \psi_L)/(p_M(1) - p_L(1))$, and $\mu_{HL} = \delta^{-1}(\psi_H - \psi_L)/(p_H(1) - p_L(1))$. We assume $\mu_{HM} > \mu_{ML}$; otherwise effort M will never be chosen by the suppliers and the problem becomes trivial. Eliminating symmetric cases, we have six subproblems to solve, which are for effort pairs (H,H), (H,M), (H,L), (M,M), (M,L), and (L,L). For each effort pair (a_1, a_2), the subproblem is largely the same as given in (4.1)-(4.6), with the IC constraints (4.4) and (4.5) each replaced by *two* IC constraints to prevent each supplier from deviating to other effort levels.

Similar to Lemma 1, it can be shown that the IC constraints for supplier i do not bind if $a_i = L$ and only one of them binds if $a_i = M$ (or H). Consequently, the (L,L) subproblem is the same as in the base model and the results about the (L,L)-forever benchmark and the upper boundary of S^* in Theorems 1 and 2 continue to hold. The (H,L) and (M,L) subproblems are similar to the original (H,L) subproblem (with one IC constraint binding for supplier 1); and the (H,H), (H,M), and (M,M) subproblems are similar to the original (H,H) subproblem (with one IC constraint binding for *each* supplier). Since the the recurrent segment on the lower boundary of S^* is driven by the (H,H) subproblem, Theorem 3 and Proposition 4 hold with minor modifications – replacing the constant μ with μ_{HM}, and the effort L with M in the conditions of Theorem 3. Therefore, the main results in the paper withstand the inclusion of more effort levels.

7 Conclusion

We have presented a dynamic contract problem for managing critical suppliers using business volume incentives. Because the manufacturer cannot directly observe or verify each supplier's effort devoted to supplying goods or services that the manufacturer buys from them, a performance-based contract is necessary. In this paper, we solve the repeated moral hazard problem with two agents and characterize the main properties of the optimal contract. We formulate the problem as a Markov decision process, treating the suppliers' continuation utility vector as the state of the system. We have shown that the process comprises of three types of states, each representing unique transition characteristic and longitudinal behavior. The discovery of these states leads to a clear understanding of the dynamic incentive structure embedded in the optimal solution. In particular, we find that individual trapping states with characteristic volume allocations, as well as a trapping region formed by a recurrent class of the Markov states, are the ultimate long-run incentive levers for the manufacturer. Compared to existing literature on dynamic contracts, we are among the very few to give well-characterized solution. In addition, our paper is first to exploit the transition dynamics and longitudinal behaviors of the optimal contract to unveil characteristics of long-run volume incentives, to the best of our knowledge.

We have made some simplifying assumptions in the model, some of which are relaxed in the extensions: we have considered asymmetric suppliers, fixed total payment (instead of volume), flexible total volume, as well as multiple effort levels. In particular, flexibility in the total quantity enables the manufacturer to offer stronger-powered incentives and trapping near the two boundaries where both suppliers exert low effort or high effort is more widespread. It is also possible to generalize our model in other directions. For instance, we have not considered common industry noise in the suppliers' production functions, and therefore, the manufacturer needs to first filter out the common noise when implementing the contract. If, however, common noise is considered in the problem, we expect the reward (punishment) becomes stronger (more severe) for the good (poor) performer because the manufacturer has to rely more on relative performance to infer the effort choice of each supplier.

Finally, although the mathematical model developed in this paper is motivated by a buyer-supplier problem involving two competing suppliers providing the same product or service, the model and solution technique can also be applied to other principal-agent problem settings involving (1) two or more agents, (2) repeated moral hazard issues, and (3) common resource constraints among the agents. This is a promising future research direction.

Acknowledgement

The authors are grateful to Stephen Graves, Donald Rosenfield for providing support and feedback on this research, and to Craig Kirkwood for his advices on the paper. We thank the Associated Editor and three anonymous reviewers for their insightful and constructive suggestions. We also thank the seminar participates at the INFORMS conference, University of California at San Diego, and University of Texas at Dallas for their suggestions.

References

Abreu, D., D. Pearce, and E. Stacchetti. 1986. Optimal Cartel Equilibria with Imperfect Monitoring. *Journal of Economic Theory* 39 (1): 251–269.

Abreu, D., D. Pearce, and E. Stacchetti. 1990. Toward a Theory of Discounted Repeated Games with Imperfect Monitoring. *Econometrica* 58 (5): 1041–1063.

Anton, J., and D. Yao. 1989. Split Awards, Procurement, and Innovation. *The RAND Journal of Economics* 20 (4): 538–552.

Baiman, S., P. Fischer, and M. Rajan. 2001. Performance Measurement and Design in Supply Chains. *Management Science* 47 (1): 173–188.

Belavina, E., and K. Girotra. 2012. The Relational Advantage of Intermediation. *Management Science* forthcoming.

Benjaafar, S., E. Elahi, and K. Donohue. 2007. Outsourcing via Service Competition. *Management Science* 53 (2): 241–259.

Bolton, P., and M. Dewatripont. 2005. *Contract Theory.* Cambridge, MA: MIT Press.

Cachon, G., and F. Zhang. 2007. Obtaining Fast Service in a Queuing System via Performance-Based Allocation of Demand. *Management Science* 53 (3): 408–420.

Chen, Y., S. Seshadri, and M. Sohoni. 2011. Optimal Design of Sales Contracts under Information Asymmetry. *Working Paper*.

Corbett, C., G. DeCroix, and A. Ha. 2005. Optimal Shared Savings Contracts In Supply Chains: Linear Contract And Double Moral Hazard. *European Journal of Operational Research* 163:653–667.

Datta, S. 2004. Adapting Decisions, Optimizing Facts and Predicting Figures. *Working paper. http://dspace.mit.edu/handle/1721.1/41904*.

Doepke, M., and R. M. Townsend. 2006. Dynamic Mechanism Design with Hidden Income and Hidden Actions. *Journal of Economic Theory* 126:235–285.

Duffy, R. 2005. Supply Base Rationalization. *CAPS Research, Critical Issues Report, www.capsresearch.org*.

Farlow, D., G. Schmidt, and A. A. Tsay. 1995. Supplier Management at Sun Microsystems. *Case Study, Graduate School of Business, Stanford University*.

Friedman, J. 1971. A Non-cooperative Equilibrium for Supergames. *Review of Economic Studies* 38 (113): 1–12.

Fudenberg, D., and J. Tirole. 1991. *Game Theory.* Cambridge, MA: MIT Press.

Fuloria, P., and S. Zenios. 2001. Outcomes-Adjusted Reimbursement in a Health-Care Delivery System. *Management Science* 47 (6): 735–751.

Giunipero, L. 1990. Motivating and Monitoring JIT Supplier Performance. *Journal of Purchasing and Materials Management* 26 (3): 19–24.

Gritzmann, P., and B. Sturmfels. 1993. Minkowski Addition of Polytopes: Computational Complexity and Applications to Grobner Bases. *SIAM J. Discrete Math.* 6 (2): 246–269.

Holmstrom, B. 1982. Moral Hazard in Teams. *The Bell Journal of Economics* 13 (2): 324–340.

Judd, K. L., S. Yeltekin, and J. Conklin. 2003. Computing Supergame Equilibria. *Econometrica* 71:1239–1254.

Kaya, M., and Ö. Özer. 2009. Quality Risk in Outsourcing: Noncontractible Product Quality and Private Quality Cost Information. *Naval Research Logistics* 56 (1): 669–685.

Kim, S., M. Cohen, and S. Netessine. 2007. Performance Contracting in After-Sales Service Supply Chains. *Management Science* 53 (12): 1843–1858.

Kim, S., M. Cohen, S. Netessine, and S. Veeraraghavan. 2011. Contracting for Infrequent Restoration and Recovery of Mission-Critical Systems. *Working Paper*.

Klotz, D., and K. Chatterjee. 1995. Dual Sourcing in Repeated Procurement Competitions. *Management Science* 41 (8): 1317–1327.

Krause, D., T. Scannell, and R. Calantone. 2000. A Structural Analysis of the Effectiveness of Buying Firms' Strategies to Improve Supplier Performance. *Decision Sciences* 31 (1): 33–54.

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

Levin, J. 2003. Relational Incentive Contracts. *American Economic Review* 93 (3): 835–857.

Li, C., and L. Debo. 2009a. Second Sourcing versus Sole Sourcing with Capacity Investment and Asymmetric Information. *Manufacturing and Service Operations Management* 11 (3): 448–470.

Li, C., and L. Debo. 2009b. Strategic Dynamic Sourcing from Competing Suppliers with Transferable Capacity Investment. *Naval Research Logistics* 56 (6): 540–562.

Lu, L., and M. Lariviere. 2011. Capacity Allocation over a Long Horizon: The Return on Turn-and-Earn. *Manufacturing and Service Operations Management* forthcoming.

O'Grady, J. 2011. 2011 MacBook Air shipping with (fast) Samsung and (slow) Toshiba SSDs. *http://www.zdnet.com/blog/apple/2011-macbook-air-shipping-with-fast-samsung-and-slow-toshiba-ssds/10701*.

Oh, S., and Ö. Özer. 2012. Mechanism Design for Capacity Planning under Dynamic Evolution of Asymmetric Demand Forecasts. *Working paper. http://ssrn.com/abstract=1529820*.

Phelan, C., and E. Stacchetti. 2001. Subgame-Perfect Equilibria in a Ramsey Tax Model. *Econometrica* 69:1491–1518.

Plambeck, E. L., and T. A. Taylor. 2006. Partnership in a Dynamic Production System with Unobservable Actions and Noncontractible Output. *Management Science* 52 (10): 1509–1527.

Plambeck, E. L., and S. A. Zenios. 2000. Performance-Based Incentives in a Dynamic Principal-Agent Model. *Manufacturing and Service Operations Management* 2 (3): 240–263.

Plambeck, E. L., and S. A. Zenios. 2003. Incentive Efficient Control of A Make-to-Stock Production System. *Operations Research* 51 (3): 371–386.

Porteus, E., and S. Whang. 1991. On Manufacturing/Marketing Incentives. *Management Science* 37 (9): 1166–1181.

Radner, R. 1985. Repeated Principal-Agent Games with Discounting. *Econometrica* 53 (5): 1173–1198.

Ren, Z. J., M. Cohen, T. Ho, and C. Terwiesch. 2010. Information Sharing in a Long-Term Supply Chain Relationship: The Role of Customer Review Strategy. *Operations Research* 58 (1): 81–93.

Sabel, C., D. O'Rourke, and A. Fung. 2000. Ratcheting Labor Standards: Regulation for Continuous Improvement in the Global Workplace. *Working paper. KSG Working Paper No. 00-010*.

Seshadri, S. 1995. Bidding for Contests. *Management Science* 41 (4): 561–576.

Shimpi, A. L., and B. Klug. 2011. Apple iPhone 4S: Thoroughly Reviewed. The Memory Interface. *http://www.anandtech.com/show/4971/apple-iphone-4s-review-att-verizon/5*.

Spear, S., and S. Srivastava. 1987. On Repeated Moral Hazard with Discounting. *Review of Economic Studies* LIV:599–717.

Swinney, R., and S. Netessine. 2009. Long-Term Contracts Under the Threat of Supplier Default. *Manufacturing and Service Operations Management* 11 (1): 109–127.

Taylor, T. A., and E. L. Plambeck. 2007a. Simple Relational Contracts to Motivate Capacity Investment: Price Only vs. Price and Quantity. *Manufacturing and Service Operations Management* 9 (1): 94–113.

Taylor, T. A., and E. L. Plambeck. 2007b. Supply Chain Relationships and Contracts: the Impact of Repeated Interaction on Capacity Investment and Procurement. *Management Science* 53 (10): 1577–1593.

Tezuka, H. 1997. Success as the Source of Failure? Competition and Cooperation in the Japanese Economy.

Sloan Management Review 38 (2): 83–93.

Whitney, L. 2012. Next-gen iPhone, iPad could deliver 20 times the graphics power. *http://news.cnet.com/8301-13579_3-57357766-37/next-gen-iphone-ipad-could-deliver-20-times-the-graphics-power/*.

Zhang, H. 2010. Partially Observable Markov Decision Processes: A Geometric Technique and Analysis. *Operations Research* 58 (1): 214–228.

Zhang, H., M. Nagarajan, and G. Sosic. 2010. Dynamic Supplier Contracts under Asymmetric Inventory Information. *Operations Research* 58 (5): 1380–1397.

A Appendix: Proofs

A.1 Proof of Lemma 1

Proof. (1) Assume $(a_1, a_2) = (L, L)$. Consider any feasible solution $\{q_1, q_2, U_1(\mathbf{x}), U_2(\mathbf{x})\}$. Let $U_1^* = E[U_1(\mathbf{x})|\, L, L]$ and $U_2^* = E[U_2(\mathbf{x})|\, L, L]$. Clearly, U_1^* and U_2^* satisfy (4.2) and (4.3). They also satisfy (4.4) and (4.5) strictly because $\psi_L < \psi_H$. Further, because $V(u_1, u_2)$ is concave, $V(U_1^*, U_2^*) = V(E[U_1(\mathbf{x}), U_2(\mathbf{x})|\, L, L]) \geq E[V(U_1(\mathbf{x}), U_2(\mathbf{x}))|\, L, L]$, by Jensen's inequality. Thus, the set of variables $\{q_1, q_2, U_1^*, U_2^*\}$ is feasible to (4.1)-(4.6) and yields weakly higher expected value for the manufacturer than $\{q_1, q_2, U_1(\mathbf{x}), U_2(\mathbf{x})\}$ does. Therefore, the problem (4.1)-(4.6) for $(a_1, a_2) = (L, L)$ must have an optimal solution that satisfies $U_1(\mathbf{x}) \equiv U_1^*$, $U_2(\mathbf{x}) \equiv U_2^*$, and the IC constraints (4.4) and (4.5) strictly.

(2) Assume $(a_1, a_2) = (H, L)$. Consider any feasible solution $\{q_1, q_2, U_1(\mathbf{x}), U_2(\mathbf{x})\}$. Let $U_i'(x_1) = E[U_i(x_1, x_2)|\, a_2 = L] = \sum_{x_2 \in \{0,1\}} p_L(x_2) U_i(x_1, x_2)$, for $i = 1, 2$, as illustrated in Figure 10(a). We have

$$\begin{aligned} E[U_i(\mathbf{x})|\, H, L] &= \sum_{x_1 \in \{0,1\}} \sum_{x_2 \in \{0,1\}} p_H(x_1) p_L(x_2) U_i(x_1, x_2) \\ &= \sum_{x_1 \in \{0,1\}} p_H(x_1) [\sum_{x_2 \in \{0,1\}} p_L(x_2) U_i(x_1, x_2)] \\ &= E[U_i'(x_1) \big|\, a_1 = H]. \end{aligned}$$

Thus, the menu $\{U_1'(x_1), U_2'(x_1)\}_{x_1 \in \{0,1\}}$ satisfies the PK constraints (4.2) and (4.3). Because $E[U_1(\mathbf{x})|\, L, L] = E[U_1'(x_1)|\, \widehat{a}_1 = L]$, the IC constraint (4.4) implies $u_1 \geq \delta E[U_1(\mathbf{x})|\, L, L] + \phi(rq_1) - \psi_L = \delta E[U_1'(x_1)|\, \widehat{a}_1 = L] + \phi(rq_1) - \psi_L$, and hence (4.4) is satisfied by $\{U_1'(x_1), U_2'(x_1)\}_{x_1 \in \{0,1\}}$. Because $E[U_2'(x_1)|\, H, H] = E[U_2'(x_1)|\, H, L]$, from (4.3) and $\psi_H > \psi_L$ we obtain $u_2 = \delta E[U_2'(x_1)|\, H, L] + \phi(rq_2) - \psi_L > \delta E[U_2'(x_1)|\, H, H] + \phi(rq_2) - \psi_H$, and hence the IC constraint (4.5) is also satisfied (strictly). Because $V(\cdot)$ is concave,

$$\begin{aligned} E[V(\mathbf{U}'(x_1)) \big|\, a_1 = H] &= \sum_{x_1 \in \{0,1\}} p_H(x_1) V(\sum_{x_2 \in \{0,1\}} p_L(x_2) \mathbf{U}(\mathbf{x})) \\ &\geq \sum_{x_1 \in \{0,1\}} p_H(x_1) \sum_{x_2 \in \{0,1\}} p_L(x_2) V(\mathbf{U}(\mathbf{x})) \\ &= E[V(\mathbf{U}(\mathbf{x}))|\, H, L] \end{aligned}$$

by Jensen's inequality. Thus, the set of variables $\{q_1, q_2, U_1'(x_1), U_2'(x_1)\}$ is feasible to the problem (4.1)-(4.6) and yields weakly higher expected value for the manufacturer than $\{q_1, q_2, U_1(\mathbf{x}), U_2(\mathbf{x})\}$ does. Therefore, the problem (4.1)-(4.6) for $(a_1, a_2) = (H, L)$ must have an optimal solution such that $U_1(x_1, x_2) \equiv U_1'(x_1)$, $U_2(x_1, x_2) \equiv U_2'(x_1)$, and the IC constraint (4.5) is strictly satisfied.

If the constraint (4.4) does not bind at $\{\mathbf{U}'(x_1)\}_{x_1 \in \{0,1\}}$, we can find two points $\mathbf{U}''(0)$ and $\mathbf{U}''(1)$ on the line segment $\overline{\mathbf{U}'(0)\mathbf{U}'(1)}$ such that $E[\mathbf{U}''(x_1)|\, H] = E[\mathbf{U}'(x_1)|\, H]$ and (4.4) binds, as illustrated in Figure 10(a). We show below that $\mathbf{U}''(0)$ and $\mathbf{U}''(1)$ must lie between $\mathbf{U}'(0)$ and $\mathbf{U}'(1)$, and hence by the concavity of $V(\cdot)$, $E[V(\mathbf{U}''(x_1))|\, H] \geq E[V(\mathbf{U}'(x_1))|\, H]$.

By the above non-binding assumption,

$$u_1 > \delta E[U_1'(x_1)\big|\, L] + \phi(rq_1) - \psi_L. \tag{A.1}$$

Because $u_1 = \delta E[U_1'(x_1)|\, H] + \phi(rq_1) - \psi_H$ and $\psi_H > \psi_L$, we have

$$u_1 < \delta E[U_1'(x_1)\big|\, H] + \phi(rq_1) - \psi_L, \tag{A.2}$$

i.e., (4.4) is violated at the expected point $E[\mathbf{U}'(x_1)|\, H]$. Inequalities (A.1) and (A.2) imply $p_L(0)U_1'(0) + p_L(1)U_1'(1) < p_H(0)U_1'(0) + p_H(1)U_1'(1)$. Because $p_L(0) - p_H(0) = p_H(1) - p_L(1) > 0$, we obtain $U_1'(0) < U_1'(1)$. For any $U_1''(0)$ and $U_1''(1)$ such that $U_1'(0) < U_1''(0) < E[U_1'(x_1)|\, H] < U_1''(1) < U_1'(1)$ and $E[U_1''(x_1)|\, H] = E[U_1'(x_1)|\, H]$, we have

$$p_H(0)[U_1''(0) - U_1'(0)] = p_H(1)[U_1'(1) - U_1''(1)]$$

and

$$\begin{aligned}
&[p_L(0)U_1''(0) + p_L(1)U_1''(1)] - [p_L(0)U_1'(0) + p_L(1)U_1'(1)] \\
=&p_L(0)[U_1''(0) - U_1'(0)] - p_L(1)[U_1'(1) - U_1''(1)] \\
=&[p_L(0) - p_L(1)\frac{p_H(0)}{p_H(1)}][U_1''(0) - U_1'(0)] > 0,
\end{aligned}$$

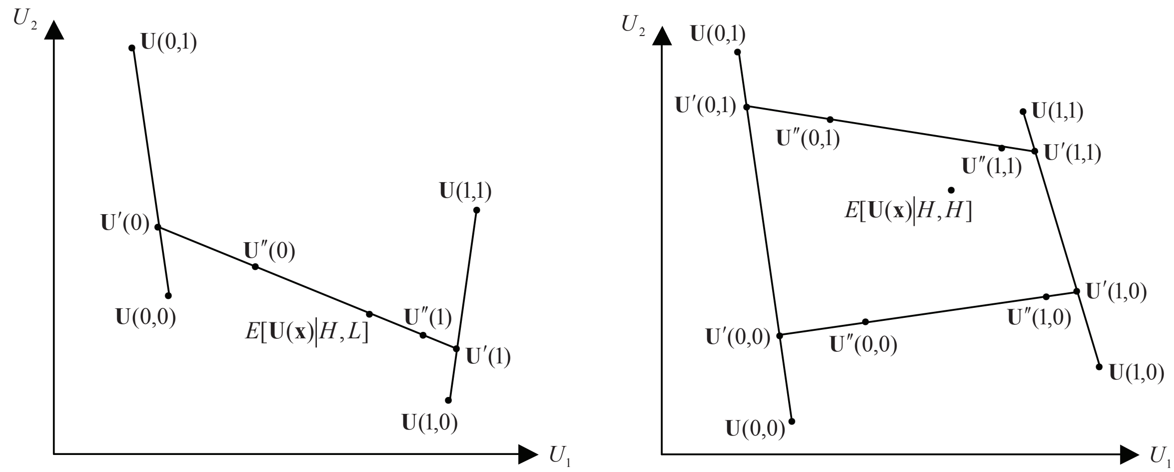
because $p_L(0)p_H(1) - p_L(1)p_H(0) = p_L(0)p_H(1) - (1 - p_L(0))(1 - p_H(1)) = p_L(0) + p_H(1) - 1 = p_H(1) - p_L(1) > 0$. Further,

$$\begin{aligned}
&[p_L(0)U_1''(0) + p_L(1)U_1''(1)] - [p_H(0)U_1''(0) + p_H(1)U_1''(1)] \\
=&[p_L(0) - p_H(0)]U_1''(0) + [p_L(1) - p_H(1)]U_1''(1) \\
=&[p_H(1) - p_L(1)][U_1''(0) - U_1''(1)] < 0.
\end{aligned}$$

Thus, we obtain $E[U_1'(x_1)|\, L] < E[U_1''(x_1)|\, L] < E[U_1''(x_1)|\, H] = E[U_1'(x_1)|\, H]$. By varying the gap between $U_1''(0)$ and $U_1''(1)$ while maintaining $U_1'(0) < U_1''(0) < U_1''(1) < U_1'(1)$ and $E[U_1''(x_1)|\, H] = E[U_1'(x_1)|\, H]$, we can have $E[U_1''(x_1)|\, L]$ anywhere between $E[U_1'(x_1)|\, L]$ and $E[U_1'(x_1)|\, H]$. Then by inequalities (A.1) and (A.2), there must exist a pair of $U_1''(0)$ and $U_1''(1)$ such that $u_1 = \delta E[U_1''(x_1)|\, L] + \phi(rq_1) - \psi_L$, i.e., the IC constraint (4.4) is satisfied with equality.

(3) The case $(a_1, a_2) = (L, H)$ is symmetric to the case (H, L) above and can be proved similarly.

(4) Assume $(a_1, a_2) = (H, H)$. Consider any feasible solution $\{\mathbf{q}, \mathbf{U}(\mathbf{x})\}$ and suppose that the constraint (4.5) does not bind. As illustrated in Figure 10(b), there must exist $\{\mathbf{U}'(\mathbf{x})\}$ such that (i) for any $x_1 \in \{0, 1\}$, $\mathbf{U}'(x_1, 0)$ and $\mathbf{U}'(x_1, 1)$ lie on the line segment $\overline{\mathbf{U}(x_1, 0)\mathbf{U}(x_1, 1)}$ and



(a) The case with effort pair (H,L) (b) The case with effort pair (H,H)

Figure 10: Making IC Constraints Binding.

$E_{x_2}[\mathbf{U}'(x_1,x_2)|\, a_2 = H] = E_{x_2}[\mathbf{U}(x_1,x_2)|\, a_2 = H]$, and (ii) the constraint (4.5) binds (by the same argument as in part 2). Then, we have

$$\begin{aligned} E_{x_1,x_2}[\mathbf{U}'(\mathbf{x})\big|\, a_2 = H, a_1] &= E_{x_1}\{\, E_{x_2}[\mathbf{U}'(x_1,x_2)\big|\, a_2 = H]\big|\, a_1\} \\ &= E_{x_1}\{\, E_{x_2}[\mathbf{U}(x_1,x_2)|\, a_2 = H]|\, a_1\} \\ &= E_{x_1,x_2}[\mathbf{U}(\mathbf{x})|\, a_2 = H, a_1], \quad a_1 \in \{H, L\}. \end{aligned}$$

Consequently, the PK constraints (4.2), (4.3), and the IC constraint (4.4) are unchanged, but the IC constraint (4.5) is now binding (by the assumption about $\{\mathbf{U}'(\mathbf{x})\}$). By the concavity of $V(\cdot)$, for any $x_1 \in \{0,1\}$, $E_{x_2}[V(\mathbf{U}'(x_1,x_2))|\, a_2 = H] \geq E_{x_2}[V(\mathbf{U}(x_1,x_2))|\, a_2 = H]$, and hence $E[V(\mathbf{U}'(\mathbf{x}))|\, H,H] \geq E[V(\mathbf{U}(\mathbf{x}))|\, H,H]$.

If the IC constraint (4.4) binds at $\{\mathbf{U}'(\mathbf{x})\}$, the proof is completed. Suppose (4.4) does not bind. As illustrated in Figure 10(b), there must exist $\{\mathbf{U}''(\mathbf{x})\}$ such that (i) for any $x_2 \in \{0,1\}$, $\mathbf{U}''(0,x_2)$ and $\mathbf{U}''(1,x_2)$ lie on the line segment $\overline{\mathbf{U}'(0,x_2)\mathbf{U}'(1,x_2)}$ and $E_{x_1}[\mathbf{U}''(x_1,x_2)|\, a_1 = H] = E_{x_1}[\mathbf{U}'(x_1,x_2)|\, a_1 = H]$, and (ii) the constraint (4.4) binds. By the same argument as above, we can show that the PK constraints (4.2), (4.3), and the IC constraint (4.5) are all unchanged, and $E[V(\mathbf{U}''(\mathbf{x}))|\, H,H] \geq E[V'(\mathbf{U}(\mathbf{x}))|\, H,H]$. Notice that both IC constraints bind at $\{\mathbf{U}''(\mathbf{x})\}$, and the proof is completed. □

A.2 Proof of Proposition 1

Proof. Define $\widehat{U}_i = E(U_i(x_1)|\, a_1 = H)$, $i = 1,2$. Then constraints (4.13), (4.14), and (4.16) become (4.22), (4.23), and (4.24). Problem (4.12)-(4.16) is transformed into the upper level problem (4.21)-(4.24) as long as $\widehat{\mathbf{U}}$ is created from $\{\mathbf{U}(x_1)\}_{x_1\in\{0,1\}}$ that satisfy

$$p_H(0)U_i(0) + p_H(1)U_i(1) = \widehat{U}_i, \quad i = 1,2 \tag{A.3}$$

and the remaining constraint (4.15).

By constraints (4.13) and (4.15), we have $\delta\widehat{U}_1 + \phi(rq_1) - \psi_H = \delta[p_L(0)U_1(0) + p_L(1)U_1(1)] + \phi(rq_1) - \psi_L$, and hence

$$p_L(0)U_1(0) + p_L(1)U_1(1) = \widehat{U}_1 - \delta^{-1}\Delta_\psi. \tag{A.4}$$

Solving equations (A.3) (for $i = 1$) and (A.4), we obtain constraints (4.18) and (4.19). Constraint (4.20) is equation (A.3) for $i = 2$. The objective (4.17) ensures that for any given $\widehat{\mathbf{U}}$, the variables $\{\mathbf{U}(x_1)\}$ are optimally chosen for the manufacturer. Therefore we obtain the lower level problem (4.17)-(4.20). □

A.3 Proof of Proposition 2

Proof. Define $\widehat{U}_i = E(U_i(\mathbf{x})|\, H, H)$, $i = 1, 2$. Then constraints (4.26), (4.27), and (4.30) become (4.37), (4.38), and (4.39). Problem (4.25)-(4.30) is transformed into the upper level problem (4.36)-(4.39) as long as $\widehat{\mathbf{U}}$ is created from $\{\mathbf{U}(\mathbf{x})\}_{\mathbf{x}\in\{0,1\}^2}$ that satisfy

$$p_H(0)[p_H(0)U_1(0,0) + p_H(1)U_1(0,1)] + p_H(1)[p_H(0)U_1(1,0) + p_H(1)U_1(1,1)] = \widehat{U}_1 \tag{A.5}$$

$$[p_H(0)U_2(0,0) + p_H(1)U_2(1,0)]p_H(0) + [p_H(0)U_2(0,1) + p_H(1)U_2(1,1)]p_H(1) = \widehat{U}_2 \tag{A.6}$$

and the remaining constraints (4.28) and (4.29).

By constraints (4.26) and (4.28), we have $\delta\widehat{U}_1+\phi(rq_1)-\psi_H = \delta\{p_L(0)[p_H(0)U_1(0,0)+p_H(1)U_1(0,1)]+ p_L(1)[p_H(0)U_1(1,0) + p_H(1)U_1(1,1)]\} + \phi(rq_1) - \psi_L$, and hence

$$p_L(0)[p_H(0)U_1(0,0) + p_H(1)U_1(0,1)] + p_L(1)[p_H(0)U_1(1,0) + p_H(1)U_1(1,1)] = \widehat{U}_1 - \delta^{-1}\Delta_\psi. \tag{A.7}$$

Solving equations (A.5) and (A.7), we obtain constraints (4.32) and (4.33). Similarly, by constraints (4.27) and (4.29), we obtain

$$[p_H(0)U_2(0,0) + p_H(1)U_2(1,0)]p_L(0) + [p_H(0)U_2(0,1) + p_H(1)U_2(1,1)]p_L(1) = \widehat{U}_2 - \delta^{-1}\Delta_\psi. \tag{A.8}$$

From (A.6) and (A.8), we obtain constraints (4.34) and (4.35). The objective (4.31) ensures that for any given $\widehat{\mathbf{U}}$, the variables $\{\mathbf{U}(\mathbf{x})\}$ are optimally chosen for the manufacturer. Therefore, the lower level problem is defined by (4.31)-(4.35). □

A.4 Proof of Proposition 3

Proof. Claims (1) and (2) follow (4.32)-(4.35) immediately. Claim (3) is true because

$$\begin{aligned}
&p_H(0)M_1(0) + p_H(1)M_1(1)\\
&= p_H(0)[p_H(0)\mathbf{U}(0,0) + p_H(1)\mathbf{U}(0,1)] + p_H(1)[p_H(0)\mathbf{U}(1,0) + p_H(1)\mathbf{U}(1,1)]\\
&= E[\mathbf{U}(\mathbf{x})|\, H, H] = (\widehat{U}_1, \widehat{U}_2)
\end{aligned}$$

and similarly $p_H(0)M_2(0) + p_H(1)M_2(1) = (\widehat{U}_1, \widehat{U}_2)$.

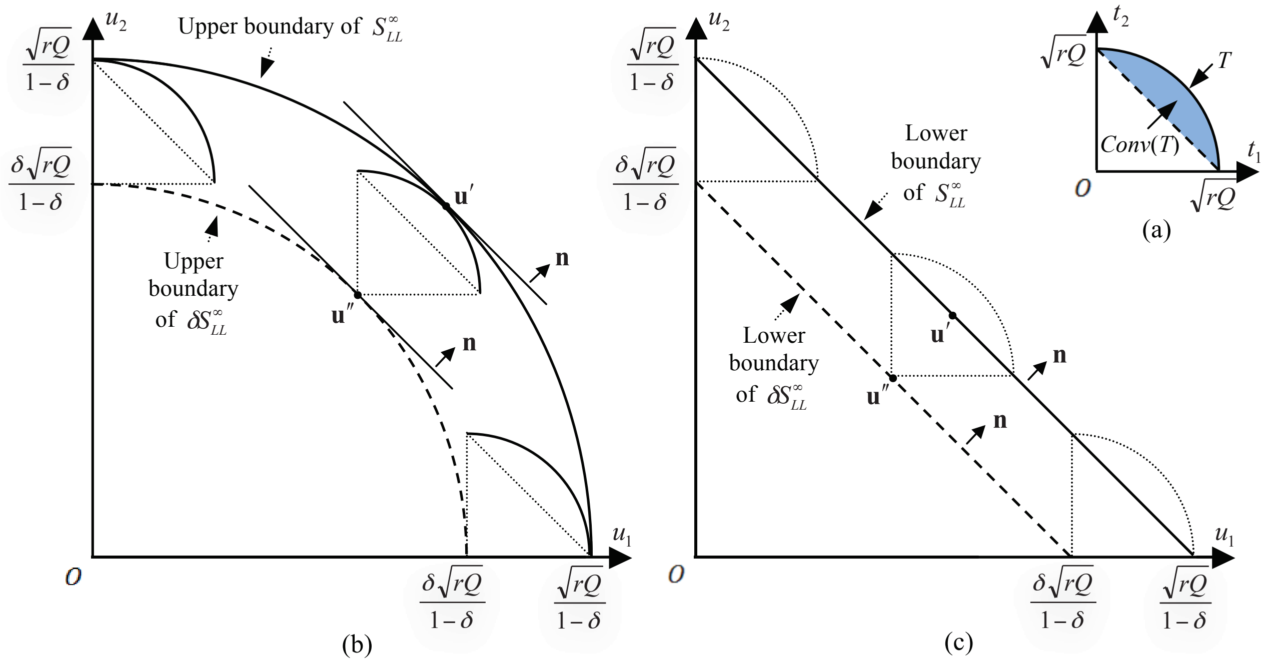


Figure 11: Determining the boundaries of S_{LL}^{∞}: (a) The set T and $conv(T)$; (b) Determining the upper boundary of S_{LL}^{∞}; (c) Determining the lower boundary of S_{LL}^{∞}.

Now, we show claim (4). Because $M_1(0) = p_H(0)\mathbf{U}(0,0)+p_H(1)\mathbf{U}(0,1)$ and $M_2(0) = p_H(0)\mathbf{U}(0,0)+p_H(1)\mathbf{U}(1,0)$, we have $M_1(0)-\mathbf{U}(0,0) = p_H(1)[\mathbf{U}(0,1)-\mathbf{U}(0,0)]$, $M_2(0)-\mathbf{U}(0,0) = p_H(1)[\mathbf{U}(1,0)-\mathbf{U}(0,0)]$, and

$$\frac{l(\overline{\mathbf{U}(0,0)M_1(0)})}{l(\overline{\mathbf{U}(0,0)\mathbf{U}(0,1)})} = p_H(1) = \frac{l(\overline{\mathbf{U}(0,0)M_2(0)})}{l(\overline{\mathbf{U}(0,0)\mathbf{U}(1,0)})}.$$

Thus, $\overline{M_1(0)M_2(0)}$ is parallel to $\overline{\mathbf{U}(0,1)\mathbf{U}(1,0)}$ and $l(\overline{M_1(0)M_2(0)}) = p_H(1)\cdot l(\overline{\mathbf{U}(0,1)\mathbf{U}(1,0)})$. Similarly, because $M_1(1) = p_H(0)\mathbf{U}(1,0)+p_H(1)\mathbf{U}(1,1)$ and $M_2(1) = p_H(0)\mathbf{U}(0,1)+p_H(1)\mathbf{U}(1,1)$, we have

$$\frac{l(\overline{\mathbf{U}(1,1)M_1(1)})}{l(\overline{\mathbf{U}(1,1)\mathbf{U}(1,0)})} = p_H(0) = \frac{l(\overline{\mathbf{U}(1,1)M_2(1)})}{l(\overline{\mathbf{U}(1,1)\mathbf{U}(0,1)})}$$

and hence $\overline{M_2(1)M_1(1)}$ is parallel to $\overline{\mathbf{U}(0,1)\mathbf{U}(1,0)}$ with length $l(\overline{M_2(1)M_1(1)}) = p_H(0)\cdot l(\overline{\mathbf{U}(0,1)\mathbf{U}(1,0)})$.

□

A.5 Proof of Theorem 1

Proof. The proof is by construction. We first derive the two boundaries of S_{LL}^{∞} when $\psi_L = 0$, as illustrated in Figure 11 (in which the utility function is $\phi(w) = \sqrt{w}$). The proof below utilizes the properties of Minkowski sum of convex polytopes.

(1) Let $\overline{S}$ denote the upper boundary of a convex set S. Consider the upper boundary of S_{LL}^{∞}, i.e., $\overline{S_{LL}^{\infty}}$. From equation (4.44), we have $\overline{S_{LL}^{\infty}} = \overline{(\delta S_{LL}^{\infty}) \oplus conv(T)} = (\delta\overline{S_{LL}^{\infty}}) \oplus \overline{conv(T)}$. Notice that $\overline{conv(T)} = T$. Refer to Figure 11(b) and consider any point $\mathbf{u}' \in \overline{S_{LL}^{\infty}}$. Let the normal vector at $\mathbf{u}'$

be $\mathbf{n}$. Clearly, the point on $\delta\overline{S_{LL}^{\infty}}$ with the same normal vector $\mathbf{n}$ is $\mathbf{u}'' = \delta\mathbf{u}'$. By the properties of Minkowski sum (Gritzmann and Sturmfels 1993), $\mathbf{u}' = \mathbf{u}'' + \mathbf{t}'$, where $\mathbf{t}'$ is the point on T with the same normal vector $\mathbf{n}$. Thus, we have $\mathbf{u}' = \delta\mathbf{u}' + \mathbf{t}'$, or, $\mathbf{u}' = (1-\delta)^{-1}\mathbf{t}'$, and consequently $\overline{S_{LL}^{\infty}} = (1-\delta)^{-1}T$.

(2) Let $\underline{S}$ denote the lower boundary of a convex set S. Consider the lower boundary of S_{LL}^{∞}, i.e., $\underline{S_{LL}^{\infty}}$. From equation (4.44), we have $\underline{S_{LL}^{\infty}} = \underline{(\delta S_{LL}^{\infty}) \oplus conv(T)} = (\delta \underline{S_{LL}^{\infty}}) \oplus \underline{conv(T)}$. Notice that $\underline{conv(T)}$ is the line segment $\overline{(\phi(0), \phi(rQ))(\phi(rQ), \phi(0))}$. Refer to Figure 11(c) and consider any point $\mathbf{u}' \in \underline{S_{LL}^{\infty}}$. By the same argument as in part (1) above, we have $\mathbf{u}' = (1-\delta)^{-1}\mathbf{t}'$ for some $\mathbf{t}' \in \underline{conv(T)}$ and thus $\underline{S_{LL}^{\infty}} = (1-\delta)^{-1}\underline{conv(T)}$.

Combining (1) and (2), we obtain $S_{LL}^{\infty} = (1-\delta)^{-1}conv(T)$ when $\psi_L = 0$. Now, assume $\psi_L > 0$ but ignore the constraint $\mathbf{u} \in \mathbb{R}_+^2$ for the moment. Define $S_{LL}^{\infty\prime} = S_{LL}^{\infty} + (1-\delta)^{-1}(\psi_L, \psi_L)$. By the properties of Minkowski sum, $(Y + \mathbf{d}) \oplus Z = (Y \oplus Z) + \mathbf{d}$ for any vector $\mathbf{d}$. Thus, $(\delta S_{LL}^{\infty\prime}) \oplus conv(T) = (\delta S_{LL}^{\infty}) \oplus conv(T) + \delta(1-\delta)^{-1}(\psi_L, \psi_L) = S_{LL}^{\infty} + (\psi_L, \psi_L) + \delta(1-\delta)^{-1}(\psi_L, \psi_L) = S_{LL}^{\infty\prime}$, where the second equality follows from equation (4.44). Therefore, the set $S_{LL}^{\infty\prime}$ is identical to the set S_{LL}^{∞} characterized above when $\psi_L = 0$. As a result, when $\psi_L > 0$ and the constraint $\mathbf{u} \in \mathbb{R}_+^2$ is ignored, we have $S_{LL}^{\infty} = S_{LL}^{\infty\prime} - (1-\delta)^{-1}(\psi_L, \psi_L) = (1-\delta)^{-1}[conv(T) - (\psi_L, \psi_L)]$.

Consider the constraint $\mathbf{u} \in \mathbb{R}_+^2$. As shown above, every $\mathbf{u}' \in (1-\delta)^{-1}[conv(T) - (\psi_L, \psi_L)]$ can be self-generated according to $\mathbf{u}' = \delta\mathbf{u}' + \mathbf{t}'$ (for some $\mathbf{t}' \in conv(T)$). Hence the truncated set $(1-\delta)^{-1}[conv(T) - (\psi_L, \psi_L)] \cap \mathbb{R}_+^2$ can be self-generated as well, and $S_{LL}^{\infty} \supset (1-\delta)^{-1}[conv(T) - (\psi_L, \psi_L)] \cap \mathbb{R}_+^2$ when the constraint $\mathbf{u} \in \mathbb{R}_+^2$ is imposed. Suppose that there exists $\tilde{\mathbf{u}} \in S_{LL}^{\infty} \setminus ((1-\delta)^{-1}[conv(T) - (\psi_L, \psi_L)] \cap \mathbb{R}_+^2)$. Then $\tilde{\mathbf{u}} + (\psi_L, \psi_L)$ must belong to the set S_{LL}^{∞} corresponding to $\psi_L = 0$, which is $(1-\delta)^{-1}conv(T)$. Clearly, such a $\tilde{\mathbf{u}}$ does not exist. Thus, $S_{LL}^{\infty} = (1-\delta)^{-1}[conv(T) - (\psi_L, \psi_L)] \cap \mathbb{R}_+^2$.

Next, we derive the manufacturer's continuation value function $V_{LL}^{\infty}(\cdot)$. Because every point $\mathbf{u}' \in S_{LL}^{\infty}$ can be self-generated as mentioned above (along with certain $\mathbf{t}' \in conv(T)$ or $\mathbf{q}$ satisfying (4.11)), it is feasible to let $\mathbf{U} = \mathbf{u}$ in problem (4.8)-(4.11) and hence $V_{LL}^{\infty}(\mathbf{u}) \geq (\overline{\pi}_L - r)Q + \delta V_{LL}^{\infty}(\mathbf{u})$, i.e., $V_{LL}^{\infty}(\mathbf{u}) \geq (1-\delta)^{-1}(\overline{\pi}_L - r)Q$. It can be further seen that the function $V_{LL}^{\infty}(\mathbf{u}) = (1-\delta)^{-1}(\overline{\pi}_L - r)Q$, for all $\mathbf{u} \in S_{LL}^{\infty}$, is a fixed point of the operator Γ_{LL}, i.e., satisfying $(\Gamma_{LL}V_{LL}^{\infty})(\cdot) = V_{LL}^{\infty}(\cdot)$. Consider the space of continuous and bounded functions with the common domain S_{LL}^{∞}, and equip the space with the supremum norm $\|f\| \equiv \sup_{\mathbf{u} \in S_{LL}^{\infty}} |f(\mathbf{u})|$, for any function $f : S_{LL}^{\infty} \to \mathbb{R}$. Consider any functions $f_1 : S_{LL}^{\infty} \to \mathbb{R}$ and $f_2 : S_{LL}^{\infty} \to \mathbb{R}$ in the space and let $d = \|f_1 - f_2\|$. By the definition (4.8)-(4.11), $\Gamma_{LL}f_1 \geq \Gamma_{LL}f_2$ if $f_1 \geq f_2$ and $\Gamma_{LL}(f + d) = \Gamma_{LL}f + \delta d$ for any constant $d \in \mathbb{R}$. Thus, $\Gamma_{LL}f_2 - \delta d = \Gamma_{LL}(f_2 - d) \leq \Gamma_{LL}f_1 \leq \Gamma_{LL}(f_2 + d) = \Gamma_{LL}f_2 + \delta d$, which implies $\|\Gamma_{LL}f_1 - \Gamma_{LL}f_2\| \leq \delta \|f_1 - f_2\|$. Hence the operator Γ_{LL} is a contraction mapping in this function space and the above fixed point $V_{LL}^{\infty}(\cdot) = (1-\delta)^{-1}(\overline{\pi}_L - r)Q$ is unique under Γ_{LL}.

Finally, according to the construction of $\overline{S_{LL}^{\infty}}$, any $\mathbf{u}' \in \overline{S_{LL}^{\infty}}$ satisfies $\mathbf{u}' = \delta\mathbf{u}' + \mathbf{t}'$, where $\mathbf{t}'$ lies on the curve T and has the same normal vector as $\mathbf{u}'$ does. Because $\mathbf{t}' = (1-\delta)\mathbf{u}'$, it is uniquely determined. By the definition of T, $\mathbf{t}' = (\phi(rq_1'), \phi(rq_2'))$ for some (q_1', q_2'), and therefore $\phi(rq_1')/\phi(rq_2') = u_1'/u_2'$. For any $\mathbf{u}' \in S_{LL}^{\infty} \setminus \overline{S_{LL}^{\infty}}$, we can still have $\mathbf{u}' = \delta\mathbf{u}' + \mathbf{t}'$, with certain $\mathbf{t}' \in conv(T) \setminus T$, corresponding to a randomized volume allocation (recall that any $\mathbf{t}$ in $conv(T) \setminus T$ gives the suppliers' expected utilities from a randomized volume allocation that randomizes between two deterministic volume allocations). However, such a construction is not unique because we can also have $\mathbf{u}' = \mathbf{u}'' + \mathbf{t}''$ for some $\mathbf{u}'' \neq \delta\mathbf{u}'$, as evident from Figure 11(c). $\square$

A.6 Proof of Theorem 2

Proof. We determine the upper boundary of S^*, i.e., $\overline{S^*}$. Because $S^* = conv(S_{LL}^* \cup S_{HL}^* \cup S_{LH}^* \cup S_{HH}^*)$, our main task is to show that the upper boundary of S_{LL}^* dominates those of S_{HL}^*, S_{LH}^*, and S_{HH}^*.

Let $\widehat{S}_{a_1a_2}^*$ represent the feasible parameter set of the lower level problem for the effort pair (a_1, a_2), given the input function $V^*(\cdot)$. By the definition of these problems, we have

$$\widehat{S}_{HL}^* = \{\widehat{\mathbf{U}} : \exists\{\mathbf{U}(x_1) \in S^*\}_{x_1\in\{0,1\}} \text{ s.t. (4.18)-(4.20)}\}, \tag{A.9}$$

$$\widehat{S}_{LH}^* = \{\widehat{\mathbf{U}} : \exists\{\mathbf{U}(x_2) \in S^*\}_{x_2\in\{0,1\}} \text{ s.t. the counterpart of (4.18)-(4.20) for } a_1a_2 = LH\}, \tag{A.10}$$

$$\widehat{S}_{HH}^* = \{\widehat{\mathbf{U}} : \exists\{\mathbf{U}(\mathbf{x}) \in S^*\}_{\mathbf{x}\in\{0,1\}^2} \text{ s.t. (4.32)-(4.35)}\}. \tag{A.11}$$

By the similarity between the upper level problems and the (L, L) problem (4.8)-(4.11), and in analogy to (4.46), we obtain

$$S_{a_1a_2}^* = (\delta\widehat{S}_{a_1a_2}^*) \oplus conv(T) - (\psi_{a_1}, \psi_{a_2}),\ a_1a_2 \in \{HL, LH, HH\}. \tag{A.12}$$

Consider any $\widehat{\mathbf{U}} \in \widehat{S}_{HH}^*$. By definition (as in Proposition 3), $\widehat{\mathbf{U}}$ is the expected continuation utility vector and is the convex combination of some $\{\mathbf{U}(\mathbf{x}) \in S^*\}_{\mathbf{x}\in\{0,1\}^2}$. Hence $\widehat{\mathbf{U}}$ must lie inside the convex hull of $\{\mathbf{U}(\mathbf{x})\}_{\mathbf{x}\in\{0,1\}^2}$ and be dominated by $\overline{S^*}$, by the convexity of S^*. Similarly, any $\widehat{\mathbf{U}} \in \widehat{S}_{HL}^*$ or $\widehat{S}_{LH}^*$ must be dominated by $\overline{S^*}$ as well. Thus, the upper boundary of S^* dominates those of $\widehat{S}_{HL}^*$, $\widehat{S}_{LH}^*$, and $\widehat{S}_{HH}^*$. By equations (4.46), (A.12), the fact $\psi_H > \psi_L$, and the monotonicity of $\overline{S_{LL}^*}$ (assuming that the curve T is monotone), the upper boundary of S_{LL}^* dominates those of S_{HL}^*, S_{LH}^*, and S_{HH}^*.

Therefore, equation (4.45) implies $\overline{S^*} = \overline{S_{LL}^*}$. By equation (4.46), we obtain

$$\overline{S_{LL}^*} = (\overline{\delta S^*}) \oplus \overline{conv(T)} - (\psi_L, \psi_L) = (\overline{\delta S_{LL}^*}) \oplus \overline{conv(T)} - (\psi_L, \psi_L).$$

Because this coincides with the definition of $\overline{S_{LL}^{\infty}}$, the upper boundaries of S^*, S_{LL}^*, and S_{LL}^{∞} are identical. $\square$

A.7 Proof of Theorem 3

The basic idea of the proof is the following: Compare problems (4.8)-(4.11), (4.12)-(4.16), and (4.25)-(4.30). Because the high effort cost ψ_H is incurred by both suppliers in the (H, H) problem

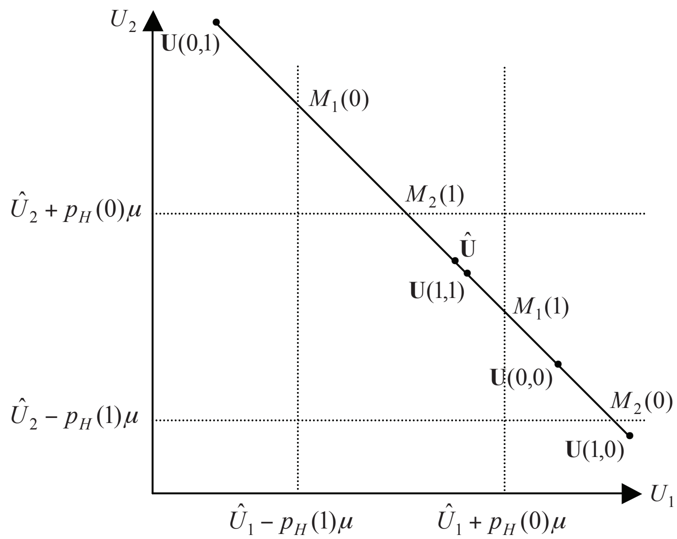


Figure 12: Lined up $\{\mathbf{U}(\mathbf{x})\}$ to Generate a Given $\widehat{\mathbf{U}}$ under (H, H) Efforts.

while by at most one of them in the other problems, the (H, H) effort pair may lead to the lowest continuation utilities for the suppliers. Consider any point $\mathbf{u} \in \underline{S^*}$ that is created by the (H, H) effort pair. According to the geometric structure described in Proposition 3 and illustrated in Figure 3(a), the corresponding expected continuation utility vector $\widehat{\mathbf{U}}$ must lie in the convex hull of the corresponding $\{\mathbf{U}(\mathbf{x}) \in S^*\}_{\mathbf{x}\in\{0,1\}^2}$, and hence $\widehat{\mathbf{U}} \in S^*$ as well. To ensure $\mathbf{u} \in \underline{S^*}$, we must push $\widehat{\mathbf{U}}$ toward $\underline{S^*}$. Ideally, we would have $\widehat{\mathbf{U}} \in \underline{S^*}$. This can be achieved when all $\{\mathbf{U}(\mathbf{x})\}$ lie on the same line and their convex hull degenerates into a line segment. Due to the symmetry between the two suppliers (essentially, the slope of $\underline{conv(T)}$), the line segment should have a $-45°$ slope. We construct the solution rigorously below with the aid of two lemmas.

Lemma A1. *Let $\mu = \delta^{-1}\Delta_\psi/(p_H(1) - p_L(1)) > 0$. If an expected continuation utility vector $\widehat{\mathbf{U}}$ can be generated from a set of $\mathbf{U}(\mathbf{x})$, $\mathbf{x} \in \{0,1\}^2$, that all lie on a $-45°$ line segment, then the line segment is the shortest when $\mathbf{U}(0,0)$ and $\mathbf{U}(1,1)$ lie between $\mathbf{U}(0,1)$ and $\mathbf{U}(1,0)$. The line segment has the following properties, as illustrated in Figure 12: (1) $\overline{\mathbf{U}(0,1)\mathbf{U}(1,0)}$ passes through $\widehat{\mathbf{U}}$; (2) $\mathbf{U}(1,0) - \mathbf{U}(0,1) = (2\mu, -2\mu)$; (3) $\mathbf{U}(0,1)$ lies to the left of the vertical line with horizontal coordinate $\widehat{U}_1 - p_H(1)\mu$ or exactly on it (in which case $\mathbf{U}(0,1)$ coincides with $\mathbf{U}(0,0)$ and $M_1(0)$); and (4) $\mathbf{U}(1,0)$ lies below the horizontal line with vertical coordinate $\widehat{U}_2 - p_H(1)\mu$ or exactly on it (in which case $\mathbf{U}(1,0)$ coincides with $\mathbf{U}(0,0)$ and $M_2(0)$).*

Proof. By Proposition 3 and Figure 3(a), when all $\{\mathbf{U}(\mathbf{x})\}$ lie on the same line, the line must pass through $\widehat{\mathbf{U}}$. As the intersections of this line with the dotted (horizontal or vertical) lines in Figure 12, the points $M_i(x_i)$, $x_i \in \{0,1\}$, $i \in \{1,2\}$, are uniquely determined. Because $l(\overline{\mathbf{U}(0,1)\mathbf{U}(1,0)}) = l(\overline{M_1(0)M_2(0)}) + l(\overline{M_2(1)M_1(1)})$, the distance between $\mathbf{U}(0,1)$ and $\mathbf{U}(1,0)$ is determined as well. To ensure that the line segment that contains all $\{\mathbf{U}(\mathbf{x})\}$ has the shortest

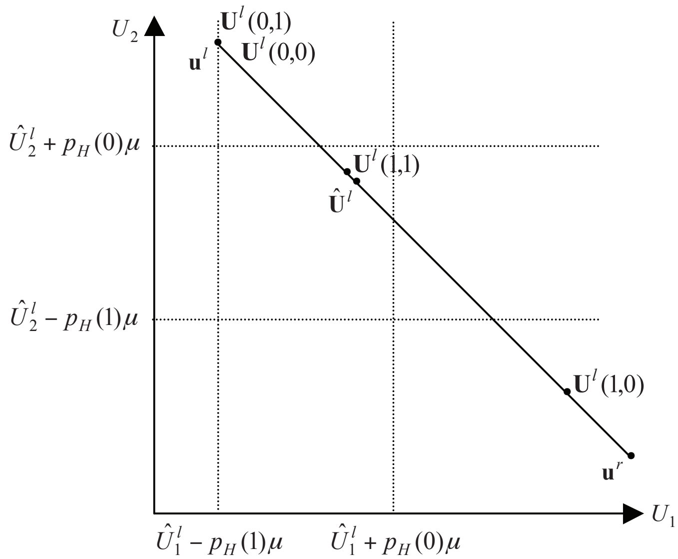


Figure 13: Generating $\mathbf{u}^l$ from $\{\mathbf{U}^l(\mathbf{x})\}$ and $\widehat{\mathbf{U}}^l$ under (H,H) Efforts.

length (so that it is easiest to sustain in the optimal solution), $\mathbf{U}(0,0)$ and $\mathbf{U}(1,1)$ must lie between $\mathbf{U}(0,1)$ and $\mathbf{U}(1,0)$. Then by Proposition 3 and Figure 3(a), $\mathbf{U}(0,1)$ must lie to the left of the vertical line with horizontal coordinate $\widehat{U}_1 - p_H(1)\mu$, and $\mathbf{U}(1,0)$ below the horizontal line with vertical coordinate $\widehat{U}_2 - p_H(1)\mu$. Further, because both $\overline{M_1(0)M_2(0)}$ and $\overline{M_2(1)M_1(1)}$ pass through $\widehat{\mathbf{U}}$, by the geometry illustrated in Figure 12, we obtain $M_2(0) - M_1(0) = (2p_H(1)\mu, -2p_H(1)\mu)$, $M_1(1) - M_2(1) = (2p_H(0)\mu, -2p_H(0)\mu)$, and hence $\mathbf{U}(1,0) - \mathbf{U}(0,1) = (2\mu, -2\mu)$. $\square$

Notice that although the length of $\overline{\mathbf{U}(0,1)\mathbf{U}(1,0)}$ is fixed, the exact locations of $\mathbf{U}(0,1)$ and $\mathbf{U}(1,0)$ (and consequently, $\mathbf{U}(0,0)$ and $\mathbf{U}(1,1)$) are flexible to some extent.

Lemma A2. *(1) If $\frac{\psi_L}{\psi_H} \leq \frac{p_L(1)}{p_H(1)}$ and $\phi(rQ) \geq 2(1-\delta p_H(0))\mu$, the $-45°$ line segment $\overline{\mathbf{u}^l\mathbf{u}^r}$ can be self-generated under the (H,H) effort pair, where $\mathbf{u}^l = (1-\delta)^{-1}(\delta p_H(1)\mu - \psi_H, -\delta p_H(1)\mu + \phi(rQ) - \psi_H)$ and $\mathbf{u}^r = (1-\delta)^{-1}(-\delta p_H(1)\mu + \phi(rQ) - \psi_H, \delta p_H(1)\mu - \psi_H)$. Further, $\overline{\mathbf{u}^l\mathbf{u}^r}$ cannot be extended at either end without losing self-sustainability, and there is no $-45°$ line segment below (to the left of) $\overline{\mathbf{u}^l\mathbf{u}^r}$ that can be self-generated under the (H,H) effort pair.*

(2) If $\frac{\psi_L}{\psi_H} > \frac{p_L(1)}{p_H(1)}$ and $\phi(rQ) \geq 2((1-\delta)\mu + \psi_H)$, the $-45°$ line segment $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$ can be self-generated under the (H,H) effort pair, where $\underline{\mathbf{u}}^l = (1-\delta)^{-1}(0, \phi(rQ) - 2\psi_H)$ and $\underline{\mathbf{u}}^r = (1-\delta)^{-1}(\phi(rQ) - 2\psi_H, 0)$. There is no $-45°$ line segment below (or to the left of) $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$ that can be self-generated under the (H,H) effort pair.

Proof. (1) Assume that the left end point of the line segment, $\mathbf{u}^l$, is generated from the expected continuation utility vector $\widehat{\mathbf{U}}^l$. To push $\mathbf{u}^l$ to the top left, by equations (4.37)-(4.38), we should choose $\mathbf{q} = (0, Q)$, and hence

$$\mathbf{u}^l = \delta\widehat{\mathbf{U}}^l + (0, \phi(rQ)) - (\psi_H, \psi_H). \tag{A.13}$$

The vector $\widehat{\mathbf{U}}^l$ is created from the set of continuation utility vectors $\{\mathbf{U}^l(\mathbf{x})\}$, all lying on the line segment $\overline{\mathbf{u}^l\mathbf{u}^r}$. To push $\mathbf{u}^l$ to the top left, we should push $\widehat{\mathbf{U}}^l$ to the top left as much as possible. From Figure 12 and Lemma A1(3), the minimum horizontal and vertical distance between $\widehat{\mathbf{U}}^l$ and $\mathbf{U}^l(0,1)$ is $p_H(1)\mu$, and the minimum is reached if $\mathbf{U}^l(0,1)$, $\mathbf{U}^l(0,0)$ and $M_1(0)$ are of the same point. Therefore, to generate the leftmost $\widehat{\mathbf{U}}^l$, $\mathbf{U}^l(0,1)$ and $\mathbf{U}^l(0,0)$ must coincide with the left endpoint of the line segment, $\mathbf{u}^l$. Hence, as illustrated in Figure 13,

$$\widehat{\mathbf{U}}^l = \mathbf{u}^l + (p_H(1)\mu, -p_H(1)\mu). \tag{A.14}$$

Substituting (A.14) into (A.13), we obtain

$$\mathbf{u}^l = (1-\delta)^{-1}(\delta p_H(1)\mu - \psi_H, -\delta p_H(1)\mu + \phi(rQ) - \psi_H).$$

Similarly, we can obtain the right end point of the line segment

$$\mathbf{u}^r = (1-\delta)^{-1}(-\delta p_H(1)\mu + \phi(rQ) - \psi_H, \delta p_H(1)\mu - \psi_H).$$

It follows that

$$\mathbf{u}^r - \mathbf{u}^l = (1-\delta)^{-1}[\phi(rQ) - 2\delta p_H(1)\mu](1,-1).$$

By Lemma A1, to generate $\widehat{\mathbf{U}}^l$ from $\{\mathbf{U}^l(\mathbf{x})\}$, we must have $\mathbf{U}^l(1,0) - \mathbf{U}^l(0,1) = (2\mu, -2\mu)$. Thus, to contain all $\{\mathbf{U}^l(\mathbf{x})\}$, $\overline{\mathbf{u}^l\mathbf{u}^r}$ must be long enough, i.e., $(1-\delta)^{-1}[\phi(rQ) - 2\delta p_H(1)\mu] \geq 2\mu$. That is, $\phi(rQ) \geq 2(1-\delta+\delta p_H(1))\mu$, or,

$$\phi(rQ) \geq 2(1-\delta p_H(0))\mu.$$

Here, we implicitly assumed that $u_1^l \geq 0$ (and $u_2^r \geq 0$), i.e., $\delta p_H(1)\mu \geq \psi_H$, which implies $\Delta_\psi p_H(1)/(p_H(1)-p_L(1)) \geq \psi_H$, $(\psi_H - \psi_L)p_H(1) \geq \psi_H(p_H(1)-p_L(1))$, or $\frac{\psi_L}{\psi_H} \leq \frac{p_L(1)}{p_H(1)}$.

Now, we show that there is no $-45°$ line segment below (or to the left of) $\overline{\mathbf{u}^l\mathbf{u}^r}$ that can be self-generated under the (H,H) effort pair. A $-45°$ line is defined by an equation $u_1 + u_2 = k$, for some constant k. A $-45°$ line segment is below or to the left of another $-45°$ line segment if the former has a smaller k in its defining equation. Suppose $\underline{L}$ is the lowest $-45°$ line segment that can be self-generated under the (H,H) effort pair, with a defining equation $u_1 + u_2 = \underline{k}$, for a certain $\underline{k} > 0$. For any feasible $\mathbf{u}$ under the (H,H) efforts, equations (4.37)-(4.38) imply that $\mathbf{u} = \delta\widehat{\mathbf{U}} + \mathbf{t} - (\psi_H, \psi_H)$, for certain $\widehat{\mathbf{U}} \in \widehat{S}^*_{HH}$ and $\mathbf{t} \in conv(T)$ (randomized volume allocation is needed to create a $\mathbf{t}$ vector in $conv(T) \setminus T$). Similar to the situation illustrated in Figure 11(c) for determining $\underline{S^\infty_{LL}}$, any $\mathbf{u} \in \underline{L}$ must be generated from certain $\widehat{\mathbf{U}} \in \underline{L}$ and $\mathbf{t} \in \underline{conv(T)}$. Because $\underline{conv(T)} = \overline{(\phi(0), \phi(rQ))(\phi(rQ), \phi(0))}$, any $\mathbf{t} \in \underline{conv(T)}$ corresponds to a volume allocation $(\widetilde{q}_1, \widetilde{q}_2)$ that randomizes between $(0,Q)$ and $(Q,0)$ and hence satisfies $E[\phi(r\widetilde{q}_1) + \phi(r\widetilde{q}_2)] = \phi(rQ)$. Thus, by (4.37)-(4.38), we have $u_1 + u_2 = \delta(\widehat{U}_1 + \widehat{U}_2) + \phi(rQ) - 2\psi_H$, which implies $\underline{k} = \delta\underline{k} + \phi(rQ) - 2\psi_H$, or $\underline{k} = (1-\delta)^{-1}(\phi(rQ) - 2\psi_H)$. Clearly, both $\mathbf{u}^l$ and $\mathbf{u}^r$ lie on the line segment $\underline{L}$ and $\overline{\mathbf{u}^l\mathbf{u}^r}$ coincides with $\underline{L}$.

(2) When $u_1^l < 0$ (and $u_2^r < 0$), i.e., $\frac{\psi_L}{\psi_H} > \frac{p_L(1)}{p_H(1)}$, because $\mathbf{u} \in \mathbb{R}_+^2$, the line segment $\overline{\mathbf{u}^l\mathbf{u}^r}$ would be truncated by the two axes, becoming $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$. Because $\underline{u}_1^l = 0$ and $\underline{u}_1^l + \underline{u}_2^l = u_1^l + u_2^l = (1-\delta)^{-1}[\phi(rQ) - 2\psi_H]$, we have $\underline{u}_2^l = (1-\delta)^{-1}[\phi(rQ) - 2\psi_H]$. Thus, the two end points of the truncated line segment are $\underline{\mathbf{u}}^l = (1-\delta)^{-1}(0, \phi(rQ) - 2\psi_H)$ and $\underline{\mathbf{u}}^r = (1-\delta)^{-1}(\phi(rQ) - 2\psi_H, 0)$. For the truncated line segment $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$ to be self-sustainable, it must be long enough as well. That is, $(1-\delta)^{-1}[\phi(rQ) - 2\psi_H] \geq 2\mu$, or $\phi(rQ) \geq 2((1-\delta)\mu + \psi_H)$. In addition, because both $\underline{\mathbf{u}}^l$ and $\underline{\mathbf{u}}^r$ lie on the line segment $\underline{L}$ defined above, $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$ is the lowest possible self-sustainable line segment under the (H, H) effort pair. $\square$

Proof. [**Proof of the Theorem**] By Lemma A2, under the (H, H) effort pair, the line segments $\overline{\mathbf{u}^l\mathbf{u}^r}$ and $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$ can be self-generated in the two cases, respectively, and there is no other self-sustainable line segment below (or to the left of) them. Thus, $\overline{\mathbf{u}^l\mathbf{u}^r}$, $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r} \subset S_{HH}^*$ and they can potentially be $\underline{S^*}$ (or part of which) in their respective cases. We verify this by showing that the other effort pairs cannot generate any $\mathbf{u}$ vector below $\overline{\mathbf{u}^l\mathbf{u}^r}$ or $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$.

Note that $\overline{\mathbf{u}^l\mathbf{u}^r}$ and $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$ are both on the $-45°$ line $\underline{L} : u_1 + u_2 = \underline{k}$, for $\underline{k} = (1-\delta)^{-1}(\phi(rQ) - 2\psi_H)$. By equations (4.46) and (A.12), to generate a continuation utility vector with the smallest $u_1 + u_2$ under any effort pair (a_1, a_2), the manufacturer must choose volume allocation $(0, Q)$, $(Q, 0)$, or a randomization between the two, such that $\phi(rq_1) + \phi(rq_2) = \phi(rQ)$ (or $E[\phi(rq_1) + \phi(rq_2)] = \phi(rQ)$). This is similar to the situation illustrated in Figure 11(c) for determining $\underline{S_{LL}^\infty}$.

Consider the (L, L) effort pair. From any $\mathbf{U} \in \underline{L}$, by (4.9)-(4.10), we have $u_1 + u_2 = \delta(U_1 + U_2) + \phi(rq_1) + \phi(rq_2) - 2\psi_L = \delta\underline{k} + \phi(rQ) - 2\psi_L = \underline{k} + 2\Delta_\psi > \underline{k}$, because $\delta\underline{k} + \phi(rQ) - 2\psi_H = \underline{k}$. Thus, the resulting $\mathbf{u}$ must lie above the line $\underline{L}$. Consider the (H, L) effort pair next. By Proposition 1 and Figure 2, when both $\mathbf{U}(0)$ and $\mathbf{U}(1)$ are drawn from $\underline{L}$, we have $\widehat{\mathbf{U}} \in \underline{L}$ as well. By (4.22)-(4.23), we have $u_1 + u_2 = \delta(\widehat{U}_1 + \widehat{U}_2) + \phi(rq_1) + \phi(rq_2) - \psi_H - \psi_L = \delta\underline{k} + \phi(rQ) - \psi_H - \psi_L = \underline{k} + \Delta_\psi > \underline{k}$. Thus, the resulting $\mathbf{u}$ lies above the line $\underline{L}$. Similarly, under the (L, H) effort pair, any $\mathbf{u}$ created from $\mathbf{U}(0)$ and $\mathbf{U}(1)$ on $\underline{L}$ must lie above $\underline{L}$ as well.

Therefore, the line segment $\overline{\mathbf{u}^l\mathbf{u}^r}$ or $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$ can only be sustained under the (H, H) effort pair, and any $\mathbf{u}$ vector on or below the $-45°$ line $\underline{L}$ must be generated by the (H, H) effort pair alone. We can easily verify that no $\mathbf{u}$ vector below $\underline{L}$ can be generated by the (H, H) effort pair. Hence we must have $\overline{\mathbf{u}^l\mathbf{u}^r} \subset \underline{S^*}$. In the case of $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$, it must be $\underline{S^*}$ itself because it extends to the two axes.

Finally, consider the manufacturer's optimal value function. Assume that $V^*(\mathbf{u}) = \overline{V}^*$, for all $\mathbf{u} \in \overline{\mathbf{u}^l\mathbf{u}^r}$, in case (1). Because the corresponding $\{\mathbf{U}(\mathbf{x})\}$ are all on $\overline{\mathbf{u}^l\mathbf{u}^r}$, by (4.25)-(4.30), we have $\overline{V}^* = \delta\overline{V}^* + (\overline{\pi}_H - r)Q$, and $\overline{V}^* = (1-\delta)^{-1}(\overline{\pi}_H - r)Q$. Thus, the function $V^*(\mathbf{u}) = \overline{V}^*$, $\mathbf{u} \in \overline{\mathbf{u}^l\mathbf{u}^r}$, is self-sustainable. Because $\overline{V}^*$ is the highest achievable expected value for the manufacturer given any effort history, we must have $V^*(\mathbf{u}) = \overline{V}^*$, for all $\mathbf{u} \in \overline{\mathbf{u}^l\mathbf{u}^r}$. The same can be shown for case (2). $\square$

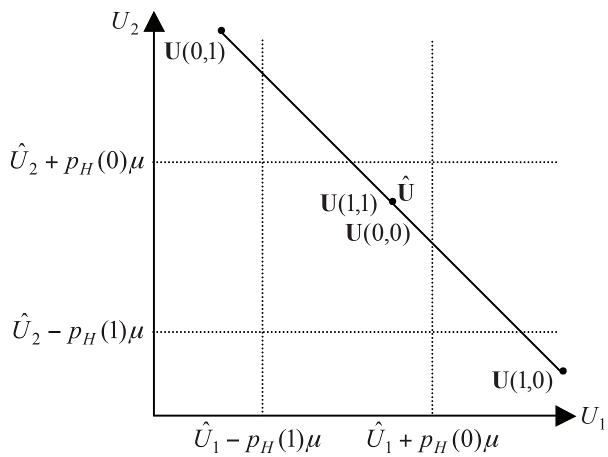


Figure 14: Generating $\widehat{\mathbf{U}}$ in a Symmetric Way.

A.8 Proof of Proposition 4

Proof. By Lemma A1 and Figure 12, an expected continuation utility vector $\widehat{\mathbf{U}}$ can be generated from a set of $\mathbf{U}(\mathbf{x})$ that all lie on a $-45°$ line segment passing through $\widehat{\mathbf{U}}$, with $\mathbf{U}(0,0)$ and $\mathbf{U}(1,1)$ lying between $\mathbf{U}(0,1)$ and $\mathbf{U}(1,0)$ and $\mathbf{U}(1,0) = \mathbf{U}(0,1) + (2\mu, -2\mu)$. By adjusting the positions of $\{\mathbf{U}(\mathbf{x})\}$, we can obtain a symmetric layout such that $\mathbf{U}(0,1) = \widehat{\mathbf{U}} + (-\mu, \mu)$, $\mathbf{U}(1,0) = \widehat{\mathbf{U}} + (\mu, -\mu)$, and $\mathbf{U}(0,0) = \mathbf{U}(1,1) = \widehat{\mathbf{U}}$, as illustrated in Figure 14.

In the first case of Theorem 3, all $\mathbf{U}(\mathbf{x})$ must be drawn from the self-sustainable line segment $\mathbf{u} \in \overline{\mathbf{u}^l\mathbf{u}^r}$, which implies that a $\widehat{\mathbf{U}}$ vector can be generated through the above symmetric layout if and only if $\widehat{\mathbf{U}} \in \overline{\widetilde{\mathbf{u}}^l\widetilde{\mathbf{u}}^r}$, where $\widetilde{\mathbf{u}}^l = \mathbf{u}^l + (\mu, -\mu) = (1-\delta)^{-1}((1-\delta p_H(0))\mu - \psi_H, -(1-\delta p_H(0))\mu + \phi(rQ) - \psi_H)$ and $\widetilde{\mathbf{u}}^r = \mathbf{u}^r + (-\mu, \mu) = (1-\delta)^{-1}(-(1-\delta p_H(0))\mu + \phi(rQ) - \psi_H, (1-\delta p_H(0))\mu - \psi_H)$. Consider any $\mathbf{u} \in \overline{\widetilde{\mathbf{u}}^l\widetilde{\mathbf{u}}^r}$. If we choose $\widehat{\mathbf{U}} = \mathbf{u}$, we would have $\mathbf{u} = \delta\mathbf{u} + (\phi(rq_1), \phi(rq_2)) - (\psi_H, \psi_H)$ and $\mathbf{u} = (1-\delta)^{-1}(\phi(rq_1) - \psi_H, \phi(rq_2) - \psi_H)$. There always exists a random volume allocation $(\widetilde{q}_1, \widetilde{q}_2)$ (randomizing between $(0, Q)$ and $(Q, 0)$) such that $E\phi(r\widetilde{q}_1) + E\phi(r\widetilde{q}_2) = \phi(rQ)$ and $E\phi(r\widetilde{q}_1) \in [(1-\delta p_H(0))\mu, -(1-\delta p_H(0))\mu + \phi(rQ)] \subset [0, \phi(rQ)]$. With this random volume allocation, $\mathbf{u}$ can be self-generated. Thus, for any $\mathbf{u} \in \overline{\widetilde{\mathbf{u}}^l\widetilde{\mathbf{u}}^r}$, it is feasible to choose $\widehat{\mathbf{U}} = \mathbf{u}$. Because any $\widehat{\mathbf{U}} \in \overline{\widetilde{\mathbf{u}}^l\widetilde{\mathbf{u}}^r}$ can be generated through the aforementioned symmetric layout, we have $\mathbf{U}(0,1) = \mathbf{u} + (-\mu, \mu)$, $\mathbf{U}(1,0) = \mathbf{u} + (\mu, -\mu)$, and $\mathbf{U}(0,0) = \mathbf{U}(1,1) = \mathbf{u}$, which proves part (1) of the proposition.

By the proof of Lemma A2 and Figure 13, a $\widehat{\mathbf{U}}$ vector that can be generated from the line segment $\overline{\mathbf{u}^l\mathbf{u}^r}$ must be at least $p_H(1)\mu$ away from each end point horizontally and vertically. Thus, any $\mathbf{u}$ vector close enough to $\mathbf{u}^l$ or $\mathbf{u}^r$ cannot be generated by letting $\widehat{\mathbf{U}} = \mathbf{u}$. These $\mathbf{u}$ vectors can be created from $\widehat{\mathbf{U}}^l$ or $\widehat{\mathbf{U}}^r$, the $\widehat{\mathbf{U}}$ vector corresponding to $\mathbf{u}^l$ or $\mathbf{u}^r$, along with proper volume allocations. By Figure 13, $\mathbf{u}^l$ is generated from $\mathbf{U}^l(0,0) = \mathbf{U}^l(0,1) = \mathbf{u}^l$, $\mathbf{U}^l(1,1) = \mathbf{u}^l + \frac{p_H(1)-p_H(0)}{p_H(1)}(\mu, -\mu)$, and $\mathbf{U}^l(1,0) = \mathbf{u}^l + (2\mu, -2\mu)$. Hence part (2) of the proposition is obtained. Part (3) can be shown similarly. □

B Appendix: Extensions

In the base model studied in previous sections, we have made some assumptions that simplify our analysis. In this appendix, we relax some important assumptions and show that our main results are robust under such extensions. For completeness, most discussions in Section 6 of the main paper are repeated here.

B.1 Asymmetric Suppliers

The basic model (3.6)-(3.9) assumes that the two suppliers are symmetric, with regard to their utility functions, cost functions, unit margins, value contributions, etc. This assumption allows us to concentrate on the most valuable circumstances for dynamic volume allocation. Suppose, for example, the suppliers' unit margins are unequal. Then the manufacturer would tend to allocate less volume to the supplier demanding the higher margin, and hence the power of volume allocation as an incentive lever would diminish. Nevertheless, as shown below, the main results of this paper can be extended to the setting with unequal supplier margins.

Suppose supplier i's unit margin is r_i, $i = 1, 2$. The manufacturer's problem (4.1)-(4.6) of inducing efforts (a_1, a_2) only needs minor modifications: replacing the term rQ in the objective function by $r_1q_1 + r_2q_2$, and replacing the terms rq_1 and rq_2 in the constraints by r_1q_1 and r_2q_2, respectively. It is straightforward to verify that Lemma 1 is still valid. Thus, the problems for inducing efforts (L, L), (H, L), (L, H), and (H, H) in Subsection 4.1 are all valid except for the above modifications in the objective functions and constraints. It implies that the lower level problems in Propositions 1 and 2, when inducing efforts (H, L), (L, H), and (H, H), are the same as before, and hence the relationship between the expected continuation utility vector $\widehat{\mathbf{U}}$ and the set of continuation utility vectors $\{\mathbf{U}(\mathbf{x})\}_{\mathbf{x}\in\{0,1\}^2}$ (or $\{\mathbf{U}(x_i)\}_{x_i\in\{0,1\}}$) is unchanged. Consequently, Figures 2, 3, and Proposition 3 bear no change as well.

The set T of one-period utility vectors generated from deterministic volume allocations, defined in expression (4.40), changes to:

$$\begin{aligned} T &= \{(\phi(r_1q_1), \phi(r_2q_2)) : q_1 + q_2 = Q,\ q_1, q_2 \in [0, Q]\} \\ &= \{(t_1, t_2) : \frac{\phi^{-1}(t_1)}{r_1} + \frac{\phi^{-1}(t_2)}{r_2} = Q,\ t_1 \in [\phi(0), \phi(r_1Q)], t_2 \in [\phi(0), \phi(r_2Q)]\}. \end{aligned} \tag{B.1}$$

As an example, if the utility function is $\phi(w) = \sqrt{w}$, i.e., $\phi^{-1}(t) = t^2$, the new set T would be the north-east quarter of an ellipse with radiuses $\sqrt{r_1Q}$ and $\sqrt{r_2Q}$, as opposed to the circle with radius $\sqrt{rQ}$ in the equal margin case. Equations (4.41) to (4.46) still hold true. Theorems 1 and 2 are also valid, after minor modifications. They can be combined as follows.

Theorem B1. *Suppose that $r_1 \neq r_2$ and both suppliers' reservation utility is 0. To induce efforts (L, L) forever, the set of suppliers' continuation utility vectors is $S_{LL}^{\infty} = (1-\delta)^{-1}[conv(T) - (\psi_L, \psi_L)] \cap \mathbb{R}_+^2$. At any $\mathbf{u} \in \overline{S_{LL}^{\infty}}$, the optimal $\mathbf{U}$ equals $\mathbf{u}$, the optimal volume allocation $\mathbf{q}$ satisfies*

$\phi(r_1q_1)/\phi(r_2q_2) = u_1/u_2$, and the manufacturer's expected value is $V^{\infty}_{LL}(\mathbf{u}) = (1-\delta)^{-1}(\overline{\pi}_L Q - r_1q_1 - r_2q_2)$. In the optimal solution to the manufacturer's problem, the upper boundary of the continuation utility set S^* coincides with $\overline{S^{\infty}_{LL}}$, and the manufacturer's optimal value at any $\mathbf{u} \in \overline{S^*}$ is given by $V^*(\mathbf{u}) = V^{\infty}_{LL}(\mathbf{u})$.

The proof of the theorem repeats those of Theorems 1 and 2 (and hence is omitted). Here, the manufacturer's expected value is only derived along the upper boundary of S^{∞}_{LL}, where the optimal volume allocation is unique but the value function $V^{\infty}_{LL}(\cdot)$ is not flat any more because the total margin $r_1q_1 + r_2q_2$ is not constant. The theorem implies that the trapping behavior of the upper boundary of S^* extends to the unequal margin case.

The properties of the optimal solution along the lower boundary of S^*, characterized by Theorem 3 and Proposition 4, can be generalized as well. However, due to space limitation, a rigorous analysis is omitted. The main modification required is that the line segment $\overline{\mathbf{u}^l\mathbf{u}^r}$ (or $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$), self-generated under the (H, H) efforts, is parallel to the line segment $\underline{conv(T)} = \overline{(\phi(0), \phi(r_2Q))(\phi(r_1Q), \phi(0))}$, whose slope is no longer $-45°$ when the margins differ. Figures 6 and 14 need be modified as well, by tilting the lines along the direction of $\underline{conv(T)}$. In addition, the manufacturer's expected value along the line segment $\overline{\mathbf{u}^l\mathbf{u}^r}$ (or $\overline{\underline{\mathbf{u}}^l\underline{\mathbf{u}}^r}$) now varies linearly between $V^*(\mathbf{u}^l)$ and $V^*(\mathbf{u}^r)$. Despite these changes, the trapping behavior of the lower boundary of S^* remains the same.

We remark that asymmetries in utility and cost functions can also be accommodated similarly, by replacing $\phi(\cdot)$ and ψ in problem (4.1)-(4.6) with $\phi_i(\cdot)$ and $\phi_i, i = 1, 2$, and the same solution approach applies. The feasible region will not be symmetric along the 45° line but the results are similar to those under the base model.

B.2 Fixed Total Payment

The additional problem at the start of period 1, given continuation utilities $\mathbf{u}^0 = (u^0_1, u^0_2)$ promised to the suppliers at the beginning, is the following:

$$V^0(\mathbf{u}^0) = \max_{\mathbf{a}, \{\mathbf{U}(\mathbf{x})\}} E[\pi(x_1)q + \pi(x_2)q + V(U_1(\mathbf{x}), U_2(\mathbf{x}))|\, \mathbf{a}] \tag{B.2}$$

$$\text{s.t.} \quad u^0_i = E[U_i(\mathbf{x})|\, \mathbf{a}] - \psi_{a_i},\ i \in \{1, 2\}. \tag{B.3}$$

This is a simple one-shot problem. The manufacturer's optimal value function $V^0(\mathbf{u}^0)$ retains the structural properties of the function $V(\mathbf{u})$ obtained from the recursive problem (6.2)-(6.5), and the optimal contract has similar properties as in the volume allocation case.

B.3 Flexible Total Volume

In the base model, the manufacturer's total business volume is a constant Q in every period. In this extension, we allow the total volume to vary in an interval, $[Q_m, Q_M]$. We assume that the manufacturer has a target volume $Q_0 \in [Q_m, Q_M]$ and incurs overorder and underorder penalties.

Thus, the manufacturer's total cost of procuring Q units, including the margins paid to the suppliers, is described by a function $g(Q) = \begin{cases} rQ + \beta_m(Q_0 - Q), & \text{if } Q \in [Q_m, Q_0), \\ rQ + \beta_M(Q - Q_0), & \text{if } Q \in [Q_0, Q_M], \end{cases}$ for some nonnegative coefficients β_m and β_M. When $\beta_m = \beta_M = \infty$, the model reduces to the base model with fixed total volume; when $\beta_m = \beta_M = 0$, the model reduces to one without a target volume. To avoid trivial cases, we assume $\overline{\pi}_L < r + \beta_M$, i.e., increasing the total volume beyond Q_0 is not profitable for the manufacturer at least in the low effort scenario; otherwise, the manufacturer would be tempted to push the total volume all the way to Q_M.

The manufacturer's problem (4.1)-(4.6) of inducing a given effort pair (a_1, a_2) only undergoes minor modifications: the manufacturer's total payment rQ in the objective function is replaced by $g(q_1 + q_2)$, and the volume constraint $q_1 + q_2 = Q$ is replaced by $q_1 + q_2 \in [Q_m, Q_M]$. It can be verified that Lemma 1 is intact. Thus, the problems of inducing (L, L), (H, L), (L, H), and (H, H) effort pairs are all similar as before except the above modifications. As a result, the decomposition of these problems is still valid, i.e., Propositions 1 and 2 are still true except for the necessary changes in the objective functions and volume constraints in the upper level problems. Propositions 3 and 4 carry over without any modification. The robustness of these results reveals that the fundamental incentive driver in the problem is unchanged under this generalization.

The flexibility in Q broadens the manufacturer's choices, which enlarges the feasible set of the suppliers' continuation utilities and improves the manufacturer's value function. Due to such changes, Theorems 1, 2, and 3 need to be modified. The main result is that the trapping region near the upper boundary of the feasible set S^* and the recurrent set near the lower boundary are both enlarged in general, as shown below.

Benchmark Contract: Inducing (L, L) Efforts Forever. The benchmark problem of inducing efforts (L, L) forever can be solved similarly as in the base model. For convenience, define the one-period utility set $T(Q)$ given total volume $Q \in [Q_m, Q_M]$ (under deterministic volume allocation) as:

$$\begin{aligned} T(Q) &= \{(\phi(rq_1), \phi(rq_2)) : q_1 + q_2 = Q,\ q_1, q_2 \geq 0\} \\ &= \{(t_1, t_2) : \phi^{-1}(t_1) + \phi^{-1}(t_2) = rQ,\ t_1, t_2 \geq \phi(0)\}, \end{aligned} \tag{B.4}$$

which is a concave curve in the (t_1, t_2) space but not a convex set. Let $conv(T(Q))$ be the convex hull of $T(Q)$, whose lower left boundary is the line segment $\overline{(0, \phi(rQ))(\phi(rQ), 0)}$ (recall that $\phi(0) = 0$). Define the one-period utility set for an interval of volumes $[Q_a, Q_b]$ as

$$T([Q_a, Q_b]) = \cup_{Q \in [Q_a, Q_b]} T(Q).$$

Theorem 1 can be generalized as follows (as illustrated in Figures 15 and 16):

Theorem B2. *Suppose both suppliers' reservation utility is* 0*. To induce efforts* (L, L) *forever, the set of suppliers' continuation utility vectors is* $S^{\infty}_{LL} = (1-\delta)^{-1}[conv(T([Q_m, Q_M])) - (\psi_L, \psi_L)] \cap \mathbb{R}^2_+$

and the manufacturer's value function $V_{LL}^{\infty}(\mathbf{u})$ *is given by: (1) If* $\overline{\pi}_L - r + \beta_m \geq 0$,

$$V_{LL}^{\infty}(\mathbf{u}) = \begin{cases} \frac{(1-\delta)(u_1+u_2)-\phi(rQ_m)}{\phi(rQ_0)-\phi(rQ_m)} \cdot \frac{(\overline{\pi}_L-r+\beta_m)(Q_0-Q_m)}{1-\delta} \\ \quad +\frac{(\overline{\pi}_L-r+\beta_m)Q_m-\beta_m Q_0}{1-\delta}, & \text{if } \mathbf{u} \in \mathbb{R}_+^2 \text{ and } u_1+u_2 \in [\frac{\phi(rQ_m)}{1-\delta}, \frac{\phi(rQ_0)}{1-\delta}), \\ \frac{(\overline{\pi}_L-r)Q_0}{1-\delta}, & \text{if } \mathbf{u} \in \frac{conv(T(Q_0))-(\psi_L,\psi_L)}{1-\delta} \cap \mathbb{R}_+^2, \\ \frac{(\overline{\pi}_L-r-\beta_M)Q+\beta_M Q_0}{1-\delta}, & \text{if } \mathbf{u} \in \frac{T(Q)-(\psi_L,\psi_L)}{1-\delta} \cap \mathbb{R}_+^2 \text{ for } Q \in (Q_0, Q_M]; \end{cases}$$

(2) If $\overline{\pi}_L - r + \beta_m < 0$,

$$V_{LL}^{\infty}(\mathbf{u}) = \begin{cases} \frac{(\overline{\pi}_L-r+\beta_m)Q_m-\beta_m Q_0}{1-\delta}, & \text{if } \mathbf{u} \in \frac{conv(T(Q_m))-(\psi_L,\psi_L)}{1-\delta} \cap \mathbb{R}_+^2, \\ \frac{(\overline{\pi}_L-r+\beta_m)Q-\beta_m Q_0}{1-\delta}, & \text{if } \mathbf{u} \in \frac{T(Q)-(\psi_L,\psi_L)}{1-\delta} \cap \mathbb{R}_+^2 \text{ for } Q \in (Q_m, Q_0], \\ \frac{(\overline{\pi}_L-r-\beta_M)Q+\beta_M Q_0}{1-\delta}, & \text{if } \mathbf{u} \in \frac{T(Q)-(\psi_L,\psi_L)}{1-\delta} \cap \mathbb{R}_+^2 \text{ for } Q \in (Q_0, Q_M]. \end{cases}$$

At any $\mathbf{u} \in S_{LL}^{\infty}$*, an optimal choice of* $\mathbf{U}$ *is* $\mathbf{u}$*. The optimal* $\mathbf{U}$ *is unique and the optimal volume allocation* $\mathbf{q}$ *satisfies* $\phi(rq_1)/\phi(rq_2) = u_1/u_2$*, for* $\mathbf{u} \in (1-\delta)^{-1}[T([Q_0, Q_M]) - (\psi_L, \psi_L)] \cap \mathbb{R}_+^2$ *if* $\overline{\pi}_L - r + \beta_m \geq 0$*, or* $\mathbf{u} \in (1-\delta)^{-1}[T([Q_m, Q_M]) - (\psi_L, \psi_L)] \cap \mathbb{R}_+^2$ *if* $\overline{\pi}_L - r + \beta_m < 0$.

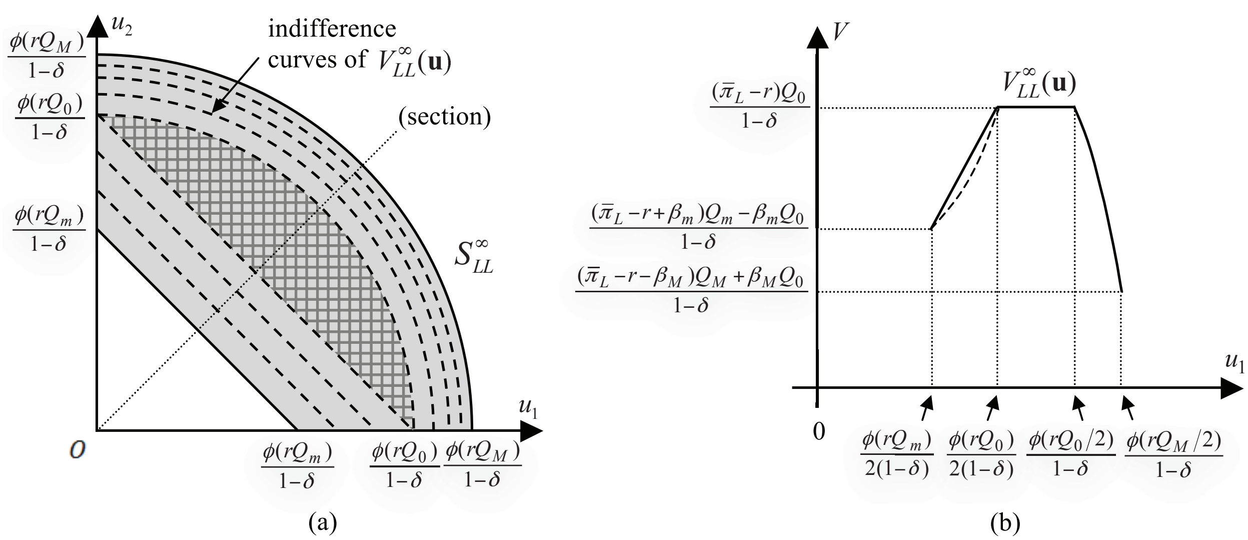


Figure 15: (a) The set S_{LL}^{∞} and indifference curves of $V_{LL}^{\infty}(\cdot)$, and (b) the section of $V_{LL}^{\infty}(\cdot)$ at $u_1 = u_2$, when $\overline{\pi}_L - r + \beta_m > 0$ and $\psi_L = 0$.

When $\overline{\pi}_L - r + \beta_m > 0$, the indifference curves of the manufacturer's value function $V_{LL}^{\infty}(\cdot)$ are illustrated in Figure 15(a) and the section of the function along the 45° ray in the $\mathbf{u}$ plane (such that $u_1 = u_2$) is illustrated in Figure 15(b). In this case, $(\overline{\pi}_L - r)Q_0 > \max\{(\overline{\pi}_L - r + \beta_m)Q_m - \beta_m Q_0, (\overline{\pi}_L - r - \beta_M)Q_M + \beta_M Q_0\}$, and thus $V_{LL}^{\infty}(\cdot)$ has a flat top over the middle subset of S_{LL}^{∞} highlighted in Figure 15(a) (which is the set S_{LL}^{∞} in the base model). $V_{LL}^{\infty}(\cdot)$ decreases as $\mathbf{u}$ moves away from the middle. Over the lower left subset of S_{LL}^{∞}, $V_{LL}^{\infty}(\cdot)$ is a convex combination of $(1-\delta)^{-1}[(\overline{\pi}_L - r + \beta_m)Q_m - \beta_m Q_0]$ and $(1-\delta)^{-1}(\overline{\pi}_L - r)Q_0$ with proper weights. When $\overline{\pi}_L - r + \beta_m < 0$, we have $(\overline{\pi}_L - r + \beta_m)Q_m - \beta_m Q_0 > (\overline{\pi}_L - r)Q_0 > (\overline{\pi}_L - r - \beta_M)Q_M + \beta_M Q_0$, and thus $V_{LL}^{\infty}(\cdot)$ has a flat top over the lower left subset of S_{LL}^{∞} illustrated in Figure 16(a). As shown in Figure

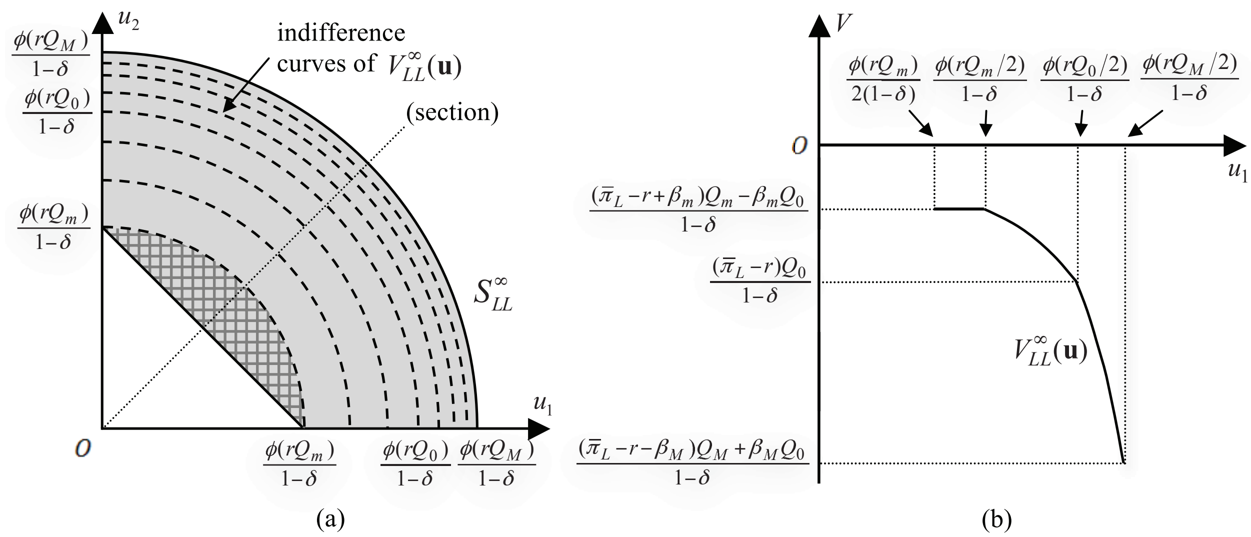


Figure 16: (a) The set S_{LL}^{∞} and indifference curves of $V_{LL}^{\infty}(\cdot)$, and (b) the section of $V_{LL}^{\infty}(\cdot)$ at $u_1 = u_2$, when $\overline{\pi}_L - r + \beta_m < 0$ and $\psi_L = 0$.

16(b), the manufacturer's value declines as $\mathbf{u}$ moves toward the upper boundary of S_{LL}^{∞} and the slope is steeper when $Q \in (Q_0, Q_M]$ than when $Q \in (Q_m, Q_0]$ because $\overline{\pi}_L - r - \beta_M < \overline{\pi}_L - r + \beta_m < 0$.

By Theorem B2, every point on the declining part of $V_{LL}^{\infty}(\cdot)$ is a trapping point, created from a total volume in $[Q_0, Q_M]$ or $[Q_m, Q_M]$, depending on the sign of $\overline{\pi}_L - r + \beta_m$. Therefore, the set of potential trapping points is enlarged as a result of the flexibility of Q.

To prove the theorem, we first show the following lemmas:

Lemma B1. *In the (t_1, t_2) utility plane, the curve $T(Q)$ is decreasing and concave everywhere.*

Proof. By definition, $T(Q) = \{(\phi(rq_1), \phi(rq_2)) : q_1 + q_2 = Q,\ q_1, q_2 \geq 0\} = \{(\phi(z), \phi(rQ - z)) : z \in [0, rQ]\}$. In the (t_1, t_2) plane (where t_1 is on the horizontal axis and t_2 on the vertical axis), the upper left endpoint of $T(Q)$ corresponds to $z = 0$ and the lower right endpoint corresponds to $z = rQ$. The slope of $T(Q)$ at $z \in [0, rQ]$ is given by:

$$s(z; Q) = \frac{-\phi'(rQ - z)}{\phi'(z)} < 0.$$

The inequality follows from the fact that $\phi'(\cdot) > 0$. Thus, $T(Q)$ is decreasing everywhere. The derivative of the slope at $z \in [0, rQ]$ is given by:

$$s'(z; Q) = \frac{\phi''(rQ - z)\phi'(z) + \phi'(rQ - z)\phi''(z)}{[\phi'(z)]^2} < 0.$$

The inequality follows from the fact that $\phi'(\cdot) > 0$ and $\phi''(\cdot) < 0$. Thus, the slope of $T(Q)$ is decreasing everywhere and $T(Q)$ is concave. $\square$

Lemma B2. *Suppose $\mathbf{t} \in T(Q)$, $\mathbf{t}' \in T(Q')$, and $\mathbf{t}'' = \lambda\mathbf{t} + (1 - \lambda)\mathbf{t}'$ for some $Q > 0$, $Q' > 0$, and $\lambda \in (0, 1)$. Then, $\mathbf{t}''$ lies below $T(Q'')$, where $Q'' = \lambda Q + (1 - \lambda)Q'$.*

Proof. By definition, $(t_1,t_2)=(\phi(rq_1),\phi(rq_2))$ for some $q_1,q_2\geq 0$ such that $q_1+q_2=Q$; $(t_1',t_2')=(\phi(rq_1'),\phi(rq_2'))$ for some $q_1',q_2'\geq 0$ such that $q_1'+q_2'=Q'$; and $(t_1'',t_2'')=(\lambda\phi(rq_1)+(1-\lambda)\phi(rq_1'),\lambda\phi(rq_2)+(1-\lambda)\phi(rq_2'))$.

Because $\phi(\cdot)$ is strictly concave, we have $\lambda\phi(rq_1)+(1-\lambda)\phi(rq_1')<\phi(\lambda rq_1+(1-\lambda)rq_1')$ and $\lambda\phi(rq_2)+(1-\lambda)\phi(rq_2')<\phi(\lambda rq_2+(1-\lambda)rq_2')$. Because $\phi^{-1}(\cdot)$ is increasing, we have

$$\begin{aligned}\phi^{-1}(t_1'')+\phi^{-1}(t_2'')&=\phi^{-1}(\lambda\phi(rq_1)+(1-\lambda)\phi(rq_1'))+\phi^{-1}(\lambda\phi(rq_2)+(1-\lambda)\phi(rq_2'))\\&<\phi^{-1}(\phi(\lambda rq_1+(1-\lambda)rq_1'))+\phi^{-1}(\phi(\lambda rq_2+(1-\lambda)rq_2'))\\&=\lambda rq_1+(1-\lambda)rq_1'+\lambda rq_2+(1-\lambda)rq_2'\\&=\lambda rQ+(1-\lambda)rQ'=rQ''.\end{aligned}$$

By the definition of $T(Q'')$, the point (t_1'',t_2'') lies below the curve $T(Q'')$ in the (t_1,t_2) plane. ☐

Now, we prove the theorem:

Proof. [**Proof of Theorem B2**] Given the total volume Q, Theorem 1 states that: (1) The manufacturer's continuation value $(1-\delta)^{-1}(\overline{\pi}_LQ-g(Q))$ can be achieved over the suppliers' continuation utility set $S_{LL}^{\infty}(Q)=(1-\delta)^{-1}[conv(T(Q))-(\psi_L,\psi_L)]\cap\mathbb{R}_+^2$, i.e., $V_{LL}^{\infty}(\mathbf{u};Q)=(1-\delta)^{-1}(\overline{\pi}_LQ-g(Q))$ for $\mathbf{u}\in S_{LL}^{\infty}(Q)$; (2) For any $\mathbf{u}\in S_{LL}^{\infty}(Q)$, an optimal choice of the future utility vector $\mathbf{U}=\mathbf{u}$; (3) For any $\mathbf{u}\in\overline{S_{LL}^{\infty}}(Q)=(1-\delta)^{-1}[T(Q)-(\psi_L,\psi_L)]\cap\mathbb{R}_+^2$, the optimal $\mathbf{U}$ is unique and the optimal volume allocation $\mathbf{q}$ satisfies $\phi(rq_1)/\phi(rq_2)=u_1/u_2$.

Now, let Q vary in the interval $[Q_m,Q_M]$. Define $S_{LL}^{\infty}([Q_a,Q_b])=\cup_{Q\in[Q_a,Q_b]}S_{LL}^{\infty}(Q)$, $\overline{S_{LL}^{\infty}}([Q_a,Q_b])=\cup_{Q\in[Q_a,Q_b]}\overline{S_{LL}^{\infty}}(Q)$, $\underline{S_{LL}^{\infty}}([Q_a,Q_b])=\cup_{Q\in[Q_a,Q_b]}\underline{S_{LL}^{\infty}}(Q)$, etc. Then, the suppliers' continuation utility set is given by $S_{LL}^{\infty}([Q_m,Q_M])=(1-\delta)^{-1}[conv(T([Q_m,Q_M]))-(\psi_L,\psi_L)]\cap\mathbb{R}_+^2$. Because $g(Q)=\begin{cases}rQ+\beta_m(Q_0-Q), & \text{if } Q\in[Q_m,Q_0),\\ rQ+\beta_M(Q-Q_0), & \text{if } Q\in[Q_0,Q_M],\end{cases}$ the manufacturer's one-period profit from total volume Q is $\overline{\pi}_LQ-g(Q)=\begin{cases}(\overline{\pi}_L-r+\beta_m)Q-\beta_mQ_0, & \text{if } Q\in[Q_m,Q_0),\\ (\overline{\pi}_L-r-\beta_M)Q+\beta_MQ_0, & \text{if } Q\in[Q_0,Q_M].\end{cases}$ Thus, given a fixed Q from the interval $[Q_m,Q_M]$, the manufacturer's value function $V_{LL}^{\infty}(\cdot;Q)$ has three cases:

$$V_{LL}^{\infty}(\mathbf{u};Q)=\begin{cases}(1-\delta)^{-1}[(\overline{\pi}_L-r+\beta_m)Q-\beta_mQ_0], & \text{if } Q\in[Q_m,Q_0),\\ (1-\delta)^{-1}(\overline{\pi}_L-r)Q_0, & \text{if } Q=Q_0,\\ (1-\delta)^{-1}[(\overline{\pi}_L-r-\beta_M)Q+\beta_MQ_0], & \text{if } Q\in(Q_0,Q_M],\end{cases}\tag{B.5}$$

for all $\mathbf{u}\in S_{LL}^{\infty}(Q)$. When Q varies in $[Q_m,Q_M]$, the manufacturer's value function $V_{LL}^{\infty}(\cdot)$ is formed by the upper convex hull of the collection of functions $\{V_{LL}^{\infty}(\cdot;Q)\}_{Q\in[Q_m,Q_M]}$. The shape of this convex hull depends on the sign of $\overline{\pi}_L-r+\beta_m$ as follows:

(1) Assume $\overline{\pi}_L-r+\beta_m>0$. Consider three regions of $\mathbf{u}$.

(i) Recall that $\overline{\pi}_L-r-\beta_M<0$. Thus, we have $V_{LL}^{\infty}(\cdot;Q_0)>\max\{V_{LL}^{\infty}(\cdot;Q_m),V_{LL}^{\infty}(\cdot;Q_M)\}$. It follows that $V_{LL}^{\infty}(\cdot)$ has a flat top over the set $S_{LL}^{\infty}(Q_0)$, i.e., $V_{LL}^{\infty}(\mathbf{u})=V_{LL}^{\infty}(\mathbf{u};Q_0)$ for all $\mathbf{u}\in S_{LL}^{\infty}(Q_0)$, as illustrated in Figure 15.

(ii) Consider the region $\overline{S^{\infty}_{LL}}((Q_0, Q_M]) = (1-\delta)^{-1}[T((Q_0, Q_M]) - (\psi_L, \psi_L)] \cap \mathbb{R}^2_+$. Any $\mathbf{u}$ in $\overline{S^{\infty}_{LL}}((Q_0, Q_M])$ must belong to the set $\overline{S^{\infty}_{LL}}(Q) = (1-\delta)^{-1}[T(Q) - (\psi_L, \psi_L)] \cap \mathbb{R}^2_+$ for some (unique) $Q \in (Q_0, Q_M]$, which can be denoted by $Q(\mathbf{u})$. Intuitively, the continuation utility vector $\mathbf{u}$ is created by splitting the total volume $Q(\mathbf{u})$ in a specific (deterministic) way forever. We show that $V^{\infty}_{LL}(\mathbf{u}) = V^{\infty}_{LL}(\mathbf{u}; Q(\mathbf{u}))$ for all $\mathbf{u} \in \overline{S^{\infty}_{LL}}((Q_0, Q_M])$. It suffices to verify that the function $V^{\infty}_{LL}(\cdot)$ so defined is concave over $\overline{S^{\infty}_{LL}}((Q_0, Q_M])$. Consider any points $\mathbf{u}_A, \mathbf{u}_B \in \overline{S^{\infty}_{LL}}((Q_0, Q_M])$ and $\mathbf{u}_\lambda = \lambda \mathbf{u}_A + (1-\lambda)\mathbf{u}_B$ for some $\lambda \in (0,1)$. By Lemma B2, $\mathbf{u}_\lambda$ lies below the curve $\overline{S^{\infty}_{LL}}(\widehat{Q})$ in the $\mathbf{u}$ plane, where $\widehat{Q} = \lambda Q(\mathbf{u}_A) + (1-\lambda)Q(\mathbf{u}_B)$. Thus, $Q(\mathbf{u}_\lambda) < \widehat{Q}$. Because $V^{\infty}_{LL}(\cdot; Q)$ is decreasing and linear in $Q \in (Q_0, Q_M]$ (by equation (B.5), $V^{\infty}_{LL}(\mathbf{u}; Q)$ is flat in $\mathbf{u}$ for a given Q), we have $V^{\infty}_{LL}(\mathbf{u}_\lambda; Q(\mathbf{u}_\lambda)) > V^{\infty}_{LL}(\cdot; \widehat{Q}) = \lambda V^{\infty}_{LL}(\mathbf{u}_A; Q(\mathbf{u}_A)) + (1-\lambda)V^{\infty}_{LL}(\mathbf{u}_B; Q(\mathbf{u}_B))$. By the definition of concave functions, the function $V^{\infty}_{LL}(\mathbf{u}) = V^{\infty}_{LL}(\mathbf{u}; Q(\mathbf{u}))$ is concave in the domain $\overline{S^{\infty}_{LL}}((Q_0, Q_M])$. (If $\mathbf{u}_\lambda$ lies below $\overline{S^{\infty}_{LL}}(Q_0)$, the total volume $Q(\mathbf{u}_\lambda)$ is out of the range $(Q_0, Q_M]$. But in that case we must have $\mathbf{u}_\lambda \in S^{\infty}_{LL}(Q_0)$ and, as shown before, $V^{\infty}_{LL}(\mathbf{u}_\lambda) = V^{\infty}_{LL}(\mathbf{u}_\lambda; Q_0)$. We just need to replace $Q(\mathbf{u}_\lambda)$ with Q_0 in the above argument.)

(iii) The lower boundary of each set $S^{\infty}_{LL}(Q)$ is the line segment $\underline{S^{\infty}_{LL}}(Q) = \{\mathbf{u} : u_1 + u_2 = (1-\delta)^{-1}\phi(rQ); u_1, u_2 \geq 0\}$. For each $\mathbf{u}$, there is a unique Q such that $\mathbf{u} \in \underline{S^{\infty}_{LL}}(Q)$, or $Q = \phi^{-1}((1-\delta)(u_1+u_2))/r$. Thus, over the set $\underline{S^{\infty}_{LL}}([Q_m, Q_0))$, which consists of the lower boundaries when $Q \in [Q_m, Q_0)$, the surface of the collection of functions $\{V^{\infty}_{LL}(\cdot; Q)\}_{Q \in [Q_m, Q_0)}$ is given by $\widetilde{V^{\infty}_{LL}}(\mathbf{u}) = (1-\delta)^{-1}[(\overline{\pi}_L - r + \beta_m)Q - \beta_m Q_0] = (1-\delta)^{-1}[(\overline{\pi}_L - r + \beta_m)\phi^{-1}((1-\delta)(u_1+u_2))/r - \beta_m Q_0]$. Because $\overline{\pi}_L - r + \beta_m > 0$ and $\phi^{-1}(\cdot)$ is convex increasing, the function $\widetilde{V^{\infty}_{LL}}(\mathbf{u})$ is convex increasing in $u_1 + u_2$, as illustrated by the dashed line in Figure 15(b). Thus, over the set $\underline{S^{\infty}_{LL}}([Q_m, Q_0))$, the convex hull of $\widetilde{V^{\infty}_{LL}}(\cdot)$, which gives $V^{\infty}_{LL}(\cdot)$, is the convex combination of $\widetilde{V^{\infty}_{LL}}(\mathbf{u})$'s at the two edges $\underline{S^{\infty}_{LL}}(Q_m)$ and $\underline{S^{\infty}_{LL}}(Q_0)$, with weights $\frac{\phi(rQ_0) - (1-\delta)(u_1+u_2)}{\phi(rQ_0) - \phi(rQ_m)}$ and $\frac{(1-\delta)(u_1+u_2) - \phi(rQ_m)}{\phi(rQ_0) - \phi(rQ_m)}$, respectively. As a result,

$$V^{\infty}_{LL}(\mathbf{u}) = \frac{(1-\delta)(u_1+u_2) - \phi(rQ_m)}{\phi(rQ_0) - \phi(rQ_m)} \cdot \frac{(\overline{\pi}_L - r + \beta_m)(Q_0 - Q_m)}{1-\delta} + \frac{(\overline{\pi}_L - r + \beta_m)Q_m - \beta_m Q_0}{1-\delta}$$

for any $\mathbf{u} \in \mathbb{R}^2_+$ such that $u_1 + u_2 \in (1-\delta)^{-1}[\phi(rQ_m), \phi(rQ_0))$.

(2) Assume $\overline{\pi}_L - r + \beta_m < 0$. Then, we have $V^{\infty}_{LL}(\cdot; Q_m) > V^{\infty}_{LL}(\cdot; Q_0) > V^{\infty}_{LL}(\cdot; Q_M)$, and thus $V^{\infty}_{LL}(\cdot)$ has a flat top over the set $S^{\infty}_{LL}(Q_m)$, i.e., $V^{\infty}_{LL}(\mathbf{u}) = V^{\infty}_{LL}(\mathbf{u}; Q_m)$ for all $\mathbf{u} \in S^{\infty}_{LL}(Q_m)$, as illustrated in Figure 16. By equation (B.5), $V^{\infty}_{LL}(\cdot; Q)$ is decreasing in $Q \in (Q_m, Q_M]$ and linear in both $(Q_m, Q_0]$ and $(Q_0, Q_M]$. Thus, similar to the case (1.ii) above, the function $V^{\infty}_{LL}(\mathbf{u}; Q(\mathbf{u}))$ is concave in the sub-domains $\overline{S^{\infty}_{LL}}((Q_m, Q_0])$ and $\overline{S^{\infty}_{LL}}((Q_0, Q_M])$ separately, where $Q(\mathbf{u})$ is the (unique) total volume Q such that $\mathbf{u} \in \overline{S^{\infty}_{LL}}(Q)$. Because $\overline{\pi}_L - r - \beta_M < \overline{\pi}_L - r + \beta_m < 0$, the slope of $V^{\infty}_{LL}(\mathbf{u}; Q(\mathbf{u}))$ in the sub-domain $\overline{S^{\infty}_{LL}}((Q_0, Q_M])$ is steeper than that in the sub-domain $\overline{S^{\infty}_{LL}}((Q_m, Q_0])$, as illustrated in Figure 16(b). So, when pieced together, $V^{\infty}_{LL}(\mathbf{u}; Q(\mathbf{u}))$ is still concave, over the entire set $\overline{S^{\infty}_{LL}}((Q_m, Q_M])$. Therefore, $V^{\infty}_{LL}(\mathbf{u})$ equals $V^{\infty}_{LL}(\mathbf{u}; Q(\mathbf{u}))$ over $\overline{S^{\infty}_{LL}}((Q_m, Q_M])$.

The proof is complete. $\square$

Optimal Solution near the Upper and Lower Boundaries of S^*. Now, consider the manufacturer's optimal value function $V^*(\cdot)$ and its domain S^*. Theorem 2 still holds after a minor modification: the upper boundary of S^* coincides with the upper boundaries of S^*_{LL} and S^∞_{LL}, i.e., $\overline{S^*} = (1-\delta)^{-1}[T(Q_M) - (\psi_L, \psi_L)] \cap \mathbb{R}^2_+$, and the manufacturer's optimal value $V^*(\mathbf{u}) = (1-\delta)^{-1}[(\overline{\pi}_L - r - \beta_M)Q_M + \beta_M Q_0]$ for any $\mathbf{u} \in \overline{S^*}$. Recall that $V^*(\cdot)$ is the convex hull of the optimal objective functions of the four subproblems, $V^*_{LL}(\cdot)$, $V^*_{HL}(\cdot)$, $V^*_{LH}(\cdot)$, and $V^*_{HH}(\cdot)$, with domains S^*_{LL}, S^*_{HL}, S^*_{LH}, and S^*_{HH}, respectively. If the sets S^*_{HL}, S^*_{LH}, and S^*_{HH} are relatively far away from $\overline{S^*}$ (e.g., when $\psi_H \gg \psi_L$), $V^*(\cdot)$ may contain a substantial portion of the declining part of $V^\infty_{LL}(\cdot)$ discussed above. Thus, on (and possibly near) $\overline{S^*}$, $V^*(\cdot)$ is made up of all trapping points.

Next, consider the lower boundary of S^*, $\underline{S^*}$. Define the set (line segment)

$$L^\infty_{HH}(Q) = \begin{cases} \{\mathbf{u} : u_1 + u_2 = \frac{\phi(rQ)-2\psi_H}{1-\delta}, u_1, u_2 \geq \frac{\delta p_H(1)\mu - \psi_H}{1-\delta}\}, & \text{if } \frac{\psi_L}{\psi_H} \leq \frac{p_L(1)}{p_H(1)},\ \phi(rQ) \geq 2(1-\delta p_H(0))\mu, \\ \{\mathbf{u} : u_1 + u_2 = \frac{\phi(rQ)-2\psi_H}{1-\delta}, u_1, u_2 \geq 0\}, & \text{if } \frac{\psi_L}{\psi_H} > \frac{p_L(1)}{p_H(1)},\ \phi(rQ) \geq 2((1-\delta)\mu + \psi_H), \\ \emptyset, & \text{otherwise.} \end{cases}$$

Theorem 3 can be summarized as follows: given a fixed total volume Q, the (H, H) effort pair can be sustained and the manufacturer's value $V^\infty_{HH}(\mathbf{u}; Q) = (1-\delta)^{-1}(\overline{\pi}_H Q - g(Q))$ can be achieved over the set $L^\infty_{HH}(Q)$ (if it is nonempty). The result still holds when Q varies in $[Q_m, Q_M]$, and the (H, H) effort pair can be sustained over the set $L^\infty_{HH}([Q_m, Q_M]) = \cup_{Q\in[Q_m,Q_M]} L^\infty_{HH}(Q)$. In addition, the manufacturer's values can be improved by randomization. Let $Q(\mathbf{u})$ be the (unique) Q such that $\mathbf{u} \in L^\infty_{HH}(Q)$. The value function $V^\infty_{HH}(\mathbf{u}; Q(\mathbf{u}))$ is not concave in $\mathbf{u}$ and can be improved by taking its upper convex hull, denoted by $V^\infty_{HH}(\mathbf{u})$. The set $L^\infty_{HH}([Q_m, Q_M])$ and function $V^\infty_{HH}(\mathbf{u})$ are illustrated in Figure 17 (assuming $L^\infty_{HH}(Q_m) \neq \emptyset$). The dashed curves in panels (b) and (c) represent the function $V^\infty_{HH}(\mathbf{u}; Q(\mathbf{u}))$ along the 45°-section.

We have the following result:

Theorem B3. *(1) If $\overline{\pi}_H - r - \beta_M < 0$, the manufacturer's optimal value is $V^*(\mathbf{u}) = (1-\delta)^{-1}(\overline{\pi}_H - r)Q_0$ for $\mathbf{u} \in L^\infty_{HH}(Q_0)$ and the line segment $L^\infty_{HH}(Q_0)$ is self-generated under the (H, H) effort pair. If, further, $L^\infty_{HH}(Q_m)$ is nonempty, it must belong to $\underline{S^*}$ and $V^*(\mathbf{u}) = V^\infty_{HH}(\mathbf{u})$ for $\mathbf{u} \in L^\infty_{HH}([Q_m, Q_0])$. (2) If $\overline{\pi}_H - r - \beta_M \geq 0$, $V^*(\mathbf{u}) = (1-\delta)^{-1}[(\overline{\pi}_H - r - \beta_M)Q_M + \beta_M Q_0]$ for $\mathbf{u} \in L^\infty_{HH}(Q_M)$ and $L^\infty_{HH}(Q_M)$ is self-generated under the (H, H) effort pair. If, further, $L^\infty_{HH}(Q_m)$ is nonempty, it must belong to $\underline{S^*}$ and $V^*(\mathbf{u}) = V^\infty_{HH}(\mathbf{u})$ for $\mathbf{u} \in L^\infty_{HH}([Q_m, Q_M])$.*

The flexibility in Q enlarges the recurrent region near the lower boundary of S^*. When $\overline{\pi}_H - r - \beta_M < 0$, as under the base model, the line segment $L^\infty_{HH}(Q_0)$ (if nonempty) is a recurrent set, although it may lie in the interior of S^* now; if the line segment $L^\infty_{HH}(Q_m)$ is nonempty, it must be part of $\underline{S^*}$ and the larger set $L^\infty_{HH}([Q_m, Q_0])$ is recurrent. When $\overline{\pi}_H - r - \beta_M \geq 0$, the existence of a recurrent set is implied by a weaker condition that $L^\infty_{HH}(Q_M) \neq \emptyset$; if $L^\infty_{HH}(Q_m) \neq \emptyset$ in addition, the whole set $L^\infty_{HH}([Q_m, Q_M])$ is a recurrent set. Intuitively, when $\overline{\pi}_H - r - \beta_M \geq 0$, a larger volume leads to higher profit for the manufacturer and, in the meantime, dynamically allocating a larger

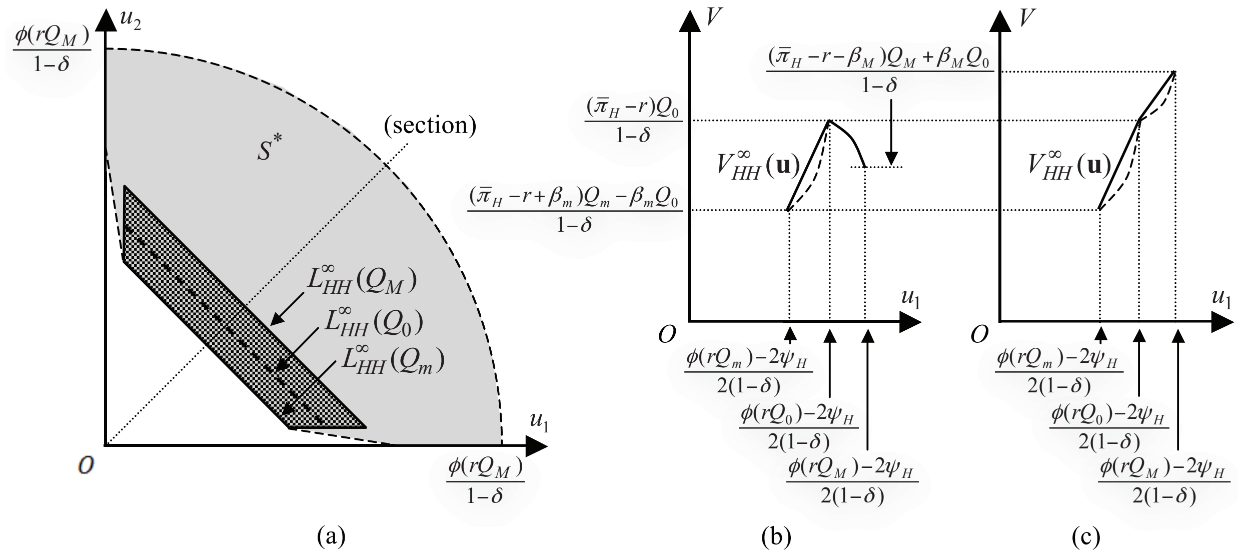


Figure 17: (a) The set $L^\infty_{HH}([Q_m, Q_M])$, (b) the 45°-section of function $V^\infty_{HH}(\cdot)$ when $\overline{\pi}_H - r - \beta_M < 0$, and (c) the 45°-section of $V^\infty_{HH}(\cdot)$ when $\overline{\pi}_H - r - \beta_M > 0$.

volume can create stronger incentives for the suppliers, so it is not only more appealing but also easier for the manufacturer to induce high effort from both suppliers.

A proof of the theorem is provided below.

Proof. **[Proof of Theorem B3]** We prove part (1) of the theorem below, i.e., assuming $\overline{\pi}_H - r - \beta_M < 0$. Part (2) can be shown in the same way.

By Theorem 3, for a given total volume Q in the interval $[Q_m, Q_M]$, if the set $L^\infty_{HH}(Q)$ is nonempty, the (H, H) effort pair can be sustained and the manufacturer's value $V^\infty_{HH}(\mathbf{u}; Q) = (1-\delta)^{-1}(\overline{\pi}_H Q - g(Q))$ can be achieved over $L^\infty_{HH}(Q)$. More specifically, conditional on Q,

$$V^\infty_{HH}(\mathbf{u}; Q) = \begin{cases} (1-\delta)^{-1}[(\overline{\pi}_H - r + \beta_m)Q - \beta_m Q_0], & \text{if } Q \in [Q_m, Q_0), \\ (1-\delta)^{-1}(\overline{\pi}_H - r)Q_0, & \text{if } Q = Q_0, \\ (1-\delta)^{-1}[(\overline{\pi}_H - r - \beta_M)Q + \beta_M Q_0], & \text{if } Q \in (Q_0, Q_M], \end{cases} \tag{B.6}$$

for $\mathbf{u} \in L^\infty_{HH}(Q)$. Let $Q(\mathbf{u})$ be the (unique) total volume Q such that $\mathbf{u} \in L^\infty_{HH}(Q)$.

Because $\overline{\pi}_H - r + \beta_m > 0$ (implied by $\overline{\pi}_H - r > 0$) and $\overline{\pi}_H - r - \beta_M < 0$ (the assumption), the manufacturer's value $(1-\delta)^{-1}(\overline{\pi}_H Q - g(Q))$ is maximized at $Q = Q_0$. Thus, $(1-\delta)^{-1}(\overline{\pi}_H - r)Q_0$ is the highest achievable value for the manufacturer and his optimal value function must satisfy $V^*(\mathbf{u}) = (1-\delta)^{-1}(\overline{\pi}_H - r)Q_0$ for $\mathbf{u} \in L^\infty_{HH}(Q_0)$.

If, in addition, the line segment $L^\infty_{HH}(Q_m)$ is nonempty (as in Figure 17), the set $L^\infty_{HH}([Q_m, Q_0])$ is included in the domain S^* of the optimal value function. Following the proof of Theorem 3, we can show that if the total volume is fixed at Q_m, no $\mathbf{u}$ vector can be sustained below (or to the left of) the line $L^\infty_{HH}(Q_m)$ by any effort pair. Because a lower volume reduces the suppliers' utilities, the minimum value of $u_1 + u_2$ must be created from the minimum volume Q_m. Thus, even when the total volume varies in $[Q_m, Q_M]$, no $\mathbf{u}$ vector can be sustained below (or to the left of) $L^\infty_{HH}(Q_m)$. As a result, $L^\infty_{HH}(Q_m)$ must be part of $\underline{S^*}$. To show that the optimal value function $V^*(\cdot)$ coincides with $V^\infty_{HH}(\cdot)$ (the convex hull of the function $V^\infty_{HH}(\mathbf{u}; Q(\mathbf{u}))$ over the set $L^\infty_{HH}([Q_m, Q_0])$, we need to show that for any $\mathbf{u} \in L^\infty_{HH}([Q_m, Q_0])$ the highest value obtainable from any other effort pair, (L, L), (H, L) or (L, H), cannot exceed $V^\infty_{HH}(\mathbf{u})$.

The argument is similar to the proof of Theorem 3. Define a line segment $L(k) = \{\mathbf{u} : u_1 + u_2 = k, u_1, u_2 \geq 0\}$, indexed by k. Let $k_m = (1-\delta)^{-1}(\phi(rQ_m) - 2\psi_H)$ and $k_0 = (1-\delta)^{-1}(\phi(rQ_0) - 2\psi_H)$. Then $L(k_m)$ and $L(k_0)$ contain the line segments $L^\infty_{HH}(Q_m)$ and $L^\infty_{HH}(Q_0)$, respectively. Consider any vector $\mathbf{u}'$ created under the (L, L) effort pair from a total volume $Q' \in [Q_m, Q_0]$ in the first period and a continuation utility vector $\mathbf{U}' \in L(k')$ from the second period onward, for some $k' \in [k_m, k_0]$. (The case $Q' \in (Q_0, Q_M]$ can be shown similarly.) By equations (4.9)-(4.10), we have

$$
\begin{aligned}
u'_1 + u'_2 &= \delta(U'_1 + U'_2) + \phi(rq'_1) + \phi(rq'_2) - 2\psi_L \\
&\geq \delta k' + \phi(rQ') - 2\psi_L \\
&> \delta k' + (1-\delta)\frac{\phi(rQ') - 2\psi_H}{1-\delta},
\end{aligned}
\tag{B.7}
$$

where the first inequality follows from the concavity of $\phi(\cdot)$ and the assumption $\phi(0) = 0$ (the inequality still holds when randomized allocation $(\tilde{q}'_1, \tilde{q}'_2)$ is considered). By expression (4.8) (with the cost rQ replaced by $g(Q)$), the manufacturer's continuation value at $\mathbf{u}'$ is

$$
\begin{aligned}
V_{LL}(\mathbf{u}') &= \delta V^\infty_{HH}(\mathbf{U}') + (\overline{\pi}_L - r + \beta_m)Q' - \beta_m Q_0 \\
&< \delta V^\infty_{HH}(\mathbf{U}') + (1-\delta)\frac{(\overline{\pi}_H - r + \beta_m)Q' - \beta_m Q_0}{1-\delta} \\
&= \delta V^\infty_{HH}(\mathbf{U}') + (1-\delta)V^\infty_{HH}(\cdot; Q').
\end{aligned}
\tag{B.8}
$$

Therefore, the point $(\mathbf{u}', V_{LL}(\mathbf{u}'))$ is dominated by the convex combination of the points $(\mathbf{U}', V^\infty_{HH}(\mathbf{U}'))$ (with weight δ) and $(\mathbf{w}', V^\infty_{HH}(\mathbf{w}'; Q'))$ (with weight $1-\delta$), for some $\mathbf{w}' \in L^\infty_{HH}(Q')$. Because $V^\infty_{HH}(\mathbf{w}'; Q') \leq V^\infty_{HH}(\mathbf{w}')$, $(\mathbf{u}', V_{LL}(\mathbf{u}'))$ is dominated by the convex combination of $(\mathbf{U}', V^\infty_{HH}(\mathbf{U}'))$ and $(\mathbf{w}', V^\infty_{HH}(\mathbf{w}'))$. Because $V^\infty_{HH}(\cdot)$ is concave, $(\mathbf{u}', V_{LL}(\mathbf{u}'))$ lies below the graph of $V^\infty_{HH}(\cdot)$.

Similarly, under the (H, L) or (L, H) effort pair, any point created by a future continuation utility vector $\mathbf{U}' \in L(k')$ for some $k' \in [k_m, k_0]$ lie below the graph of $V^\infty_{HH}(\cdot)$ as well. Thus, the optimal value function $V^*(\mathbf{u}) = V^\infty_{HH}(\mathbf{u})$ for all $\mathbf{u} \in L^\infty_{HH}([Q_m, Q_0])$. ☐

B.4 Multiple Effort Levels

In the base model, the suppliers' effort level can be either H or L. In this extension, we add an intermediate level, M. More effort levels can be treated similarly.

As in the two-effort-level case, assume that the disutilities of the effort levels and corresponding probabilities of the good outcome are ordered such that $\psi_H > \psi_M > \psi_L$ and $p_H(1) > p_M(1) >$

$p_L(1)$. Define the effective marginal costs of effort as $\mu_{HM} = \delta^{-1}(\psi_H - \psi_M)/(p_H(1) - p_M(1))$, $\mu_{ML} = \delta^{-1}(\psi_M - \psi_L)/(p_M(1) - p_L(1))$, and $\mu_{HL} = \delta^{-1}(\psi_H - \psi_L)/(p_H(1) - p_L(1))$.

Now we have nine possible effort pairs. After eliminating symmetric cases, six pairs are left, which are (H, H), (H, M), (H, L), (M, M), (M, L), and (L, L). As a result, we have more subproblems to solve. For each effort pair (a_1, a_2), the manufacturer's subproblem (4.1)-(4.6) is more complex as well because there are two IC constraints for each supplier. For instance, to induce $a_1 = M$, the IC constraints for supplier 1, denoted by $(IC_{1,MH})$ and $(IC_{1,ML})$, would prevent the supplier from deviating to effort H or L. Nevertheless, the subproblems can be simplified through the following generalized version of Lemma 1:

Lemma B3. *If $\mu_{HM} < \mu_{ML}$, effort M will never be chosen by the suppliers and can be removed from the problem formulation without loss of optimality. If $\mu_{HM} > \mu_{ML}$, given any concave function $V(\cdot)$ and continuation utility vector $\mathbf{u}$, there exists an optimal solution to problem (4.1)-(4.6) such that: (1) if $a_i = L$, the IC constraints for supplier i do not bind; (2) if $a_i = M$, constraint $(IC_{i,ML})$ binds while $(IC_{i,MH})$ does not; (3) if $a_i = H$, constraint $(IC_{i,HM})$ binds while $(IC_{i,HL})$ does not. If $\mu_{HM} = \mu_{ML}$, the two IC constraints mentioned in case (2) or (3) above bind simultaneously. Furthermore, in all circumstances, the future continuation utility vectors $\{\mathbf{U}(\mathbf{x})\}$ are independent of x_i if and only if $a_i = L$, for $i \in \{1, 2\}$.*

Proof. Because of the symmetry between the suppliers, it suffices to consider $i = 1$. Without loss of generality, suppose that the manufacturer wants to induce effort a_1 from supplier 1 through continuation utility vectors $\{\mathbf{U}(\mathbf{x})\}$ that depends on both x_1 and x_2. Let a_2 be the effort exerted by supplier 2. Define $\overline{U}_1(x_1) = p_{a_2}(0)U_1(x_1, 0) + p_{a_2}(1)U_1(x_1, 1)$, for $x_1 \in \{0, 1\}$. The expected continuation utility for supplier 1 is given by $\delta E[\overline{U}_1(x_1) \big| a_1] + \phi(rq_1) - \psi_{a_1} = \delta[p_{a_1}(0)\overline{U}_1(0) + p_{a_1}(1)\overline{U}_1(1)] + \phi(rq_1) - \psi_{a_1} = \delta[\overline{U}_1(0) + p_{a_1}(1)(\overline{U}_1(1) - \overline{U}_1(0))] + \phi(rq_1) - \psi_{a_1}$. The variable part of the continuation utility related to effort a_1 is $\delta p_{a_1}(1)(\overline{U}_1(1) - \overline{U}_1(0)) - \psi_{a_1}$. Supplier 1's continuation utilities under efforts a_1 and $\hat{a}_1$ differ by $\delta(p_{a_1}(1) - p_{\hat{a}_1}(1))(\overline{U}_1(1) - \overline{U}_1(0)) - (\psi_{a_1} - \psi_{\hat{a}_1}) = \delta(p_{a_1}(1) - p_{\hat{a}_1}(1))(\overline{U}_1(1) - \overline{U}_1(0) - \mu_{a_1\hat{a}_1})$. The constraint $(IC_{1,a_1\hat{a}_1})$ that ensures that supplier 1 prefers effort a_1 to $\hat{a}_1$ is equivalent to $\overline{U}_1(1) - \overline{U}_1(0) \geq \mu_{a_1\hat{a}_1}$ when $p_{a_1}(1) > p_{\hat{a}_1}(1)$ or $\overline{U}_1(1) - \overline{U}_1(0) \leq \mu_{a_1\hat{a}_1}$ when $p_{a_1}(1) < p_{\hat{a}_1}(1)$.

Now, assume $\mu_{HM} < \mu_{ML}$. Because $p_L(1) < p_M(1) < p_H(1)$, the constraints $(IC_{1,MH})$ and $(IC_{1,ML})$, which induce effort M, imply that $\mu_{MH}(= \mu_{HM}) \geq \overline{U}_1(1) - \overline{U}_1(0) \geq \mu_{ML}$. But this contradicts the assumption and therefore, supplier 1 will never choose effort M.

Next, assume $\mu_{HM} > \mu_{ML}$. (1) The case $a_i = L$ can be shown by the same argument as in Lemma 1. (2) Consider the case $a_i = M$. By the argument above, the constraints $(IC_{1,MH})$ and $(IC_{1,ML})$ are equivalent to $\mu_{MH} \geq \overline{U}_1(1) - \overline{U}_1(0) \geq \mu_{ML}$. According to the proof of Lemma 1, the gap $\overline{U}_1(1) - \overline{U}_1(0)$ should be minimized at optimality. Thus, we have $\mu_{MH} > \overline{U}_1(1) - \overline{U}_1(0) = \mu_{ML}$ at optimality, which implies that $(IC_{1,ML})$ binds and $(IC_{1,MH})$ holds with strict inequality. (3)

Consider the case $a_i = H$. The assumption $\mu_{HM} > \mu_{ML}$, or $\frac{\psi_H-\psi_M}{p_H(1)-p_M(1)} > \frac{\psi_M-\psi_L}{p_M(1)-p_L(1)}$, implies that $\frac{\psi_H-\psi_M}{p_H(1)-p_M(1)} > \frac{(\psi_H-\psi_M)+(\psi_M-\psi_L)}{(p_H(1)-p_M(1))+(p_M(1)-p_L(1))} > \frac{\psi_M-\psi_L}{p_M(1)-p_L(1)}$, or $\mu_{HM} > \mu_{HL} > \mu_{ML}$. According to the result at the beginning of the proof, $(IC_{1,HM})$ and $(IC_{1,HL})$ are equivalent to $\overline{U}_1(1)-\overline{U}_1(0) \geq \mu_{HM}$ and $\overline{U}_1(1)-\overline{U}_1(0) \geq \mu_{HL}$, respectively. From $\mu_{HM} > \mu_{HL}$ and the fact that the gap $\overline{U}_1(1)-\overline{U}_1(0)$ is minimized at optimality, the two constraints imply that $\overline{U}_1(1) - \overline{U}_1(0) = \mu_{HM} > \mu_{HL}$, i.e., $(IC_{1,HM})$ binds and $(IC_{1,HL})$ holds with strict inequality.

When $\mu_{HM} = \mu_{ML}$, we have $\mu_{HM} = \mu_{HL} = \mu_{ML}$. It follows that $(IC_{1,MH})$ and $(IC_{1,ML})$ binds simultaneously in case (2) and $(IC_{1,HM})$ and $(IC_{1,HL})$ binds simultaneously in case (3).

By the same argument as in Lemma 1, we can show that to induce $a_i = L$, $\{\mathbf{U}(\mathbf{x})\}$ should not depend on x_i at optimality, since no incentive is needed for supplier i; but to induce $a_i = M$ or H, $\{\mathbf{U}(\mathbf{x})\}$ should be positively related to x_i, to provide necessary incentive for supplier i. $\square$

The intuition behind the lemma is similar to the one in the base model: the future continuation utility vector $\mathbf{U}(\mathbf{x})$ should increase with x_i to motivate supplier i to exert non-trivial effort, and the gap between supplier i's expected continuation utilities $\overline{U}_i(1)$ and $\overline{U}_i(0)$ should be large enough to overcome the pertinent effective marginal cost of effort, where $\overline{U}_i(x_i) = \sum_{x_j \in \{0,1\}} p_{a_j}(x_j) U_i(x_i, x_j)$ for $j \neq i$.

By Lemma B3, the manufacturer's subproblem (4.1)-(4.6) for effort pair (a_1, a_2) can be simplified as follows: for $a_i = L$, no IC constraint is present for supplier i; and for $a_i = M$ or H, the IC constraint $(IC_{i,ML})$ or $(IC_{i,HM})$ is present. Due to this simplification, the (L, L) subproblem is the same as in the base model; the (H, L) and (M, L) subproblems are similar to the original (H, L) subproblem; and the (H, H), (H, M), and (M, M) subproblems are similar to the original (H, H) subproblem. The decomposition of these subproblems, and hence Propositions 1, 2, and 3, are also similar as before, except that effort H for supplier i in the original propositions can be H or M now and the constant μ should be μ_{HM} or μ_{ML}, correspondingly. Because the (L, L) subproblem does not change, the results about the (L, L)-forever benchmark and the upper boundary of S^*, i.e., Theorems 1 and 2, still hold true. Because the recurrent segment along the lower boundary of S^* is driven by the (H, H) subproblem, which only bears a minor modification by replacing constraints $(IC_{i,HL})$ with $(IC_{i,HM})$, Theorem 3 and Proposition 4 only need minor modifications as well: the constant μ becomes μ_{HM}, and the effort L in the conditions of Theorem 3 becomes M.

In conclusion, the main results in the paper withstand the inclusion of more effort levels.