Competition or Compensation: Supplier Incentives Under the American and Japanese Subcontracting Systems

By Curtis R. Taylor and Steven N. Wiggins *

Two fundamentally different subcontracting systems arise as distinct solutions to the quality control problem facing an input buyer. The "American" system involves competitive bidding on each contract, large orders, and inspections. The "Japanese" system involves repeat purchases from a supplier who earns a premium, small orders, and no inspections. Both systems may coexist as local solutions, but the global optimum is determined by the ratio of set-up to inspection costs. This suggests that the adoption of flexible manufacturing equipment and rising product complexity may be responsible for the shift from the American to the Japanese system observed in many industries. (*JEL* L14, L15, and L22)

This paper investigates two procurement systems, the competitive-bidding system traditionally practiced by firms in the United States and other Western countries and the Just-In-Time (JIT) purchasing system that originated in Japan. We call these two forms of procurement "systems" because they are associated with sharply different sets of behavior for buyers and suppliers and, therefore, lead to different procurement outcomes. For ease of reference, we identify the competitive-bidding system below with "American-style" procurement and the JIT purchasing system with "Japanese-style" procurement, bearing in mind that these labels reflect only historical practice and that there is nothing inherently "American" or "Japanese" about either mode of purchasing.

American-style procurement typically exhibits the following features. A large lot is ordered from the supplier making the lowest bid with little regard to past or future purchases. The buyer often exercises its right to inspect a shipment and may decide to reject delivery and withhold payment if unsatisfied with lot quality. (These rights are formalized in the Uniform Commercial Code Sections 2-513 and 2-601.) If a shipment is accepted and quality subsequently turns out to be poor, the buyer often has little legal recourse. Japanese-style procurement features almost the reverse set of characteristics. Under JIT purchasing, small lots are ordered repeatedly from the same supplier over long periods. Competitive bidding is not used for individual lots, and the incumbent supplier may not even be the least cost source. Deliveries are rarely inspected by the buyer, but firms that supply low quality are ultimately cut off.[1] Our emphasis on the differences between competitive bidding and JIT procurement is in the spirit of Paul Milgrom and John Roberts (1990, 1995) and Bengt Holmstrom and Milgrom (1994). These authors argue that it is important to focus on an entire set of characteristics in analyzing institutional alternatives. Holmstrom and Milgrom (1994 p. 972), for example, ask: "Why does

* Department of Economics, Texas A&M University, College Station, TX 77843. We thank Kenneth Judd, Paul Milgrom, William Neilson, Canice Prendergast, two anonymous referees, and many seminar participants for helpful comments and suggestions. Financial support from the Texas Higher Education Coordinating Board's Advanced Research Program is gratefully acknowledged. Taylor also is grateful for the support of the Hoover Institution, the Alfred P. Sloan Foundation, and the National Science Foundation (Grant SBR94-22830); Wiggins thanks the Jordan Professorship of the Private Enterprise Research Center at Texas A&M.

[1] The absence of incoming inspections throughout Japanese industry is widely documented. For example, see John McMillan (1994 p. 216), who portrays Japanese subcontracting much as it is modeled here. For a general discussion of the theory of the Japanese firm, see the survey by Masahiko Aoki (1990) and the works cited there.

outside procurement tend to involve purchases from a worker who chooses his or her own methods and hours and owns the tools used and is paid only for quantities supplied?'' They answer this question by identifying the ''complementarities'' among these traits. In a similar vein, we show that competitive bidding among numerous suppliers, large orders, and inspections constitute one set of complementary characteristics, while long-term business relationships with limited suppliers, small orders, and a lack of inspections form another.

Our analysis hinges on a basic contractual difference between the competitive bidding and JIT systems, namely, the means for inducing contractual performance. Specifically, under the American system, incentives to supply high quality are supported by an up-front inspection and an associated threat to reject delivery. Under the Japanese system, however, the incentive to furnish high quality is governed by an implicit promise of contract renewals. This key contractual difference is firmly grounded in the literature.

American firms tend to deal with faulty supplies on a lot-by-lot basis (Richard Schonberger, 1982 Ch. 7). Deliveries are typically inspected and returned to the supplier without payment if they are found unsatisfactory. The Japanese, in contrast, back their long-run view of a supply partnership with a long-run threat to permanently ''break off'' a supplier who performs unsatisfactorily. (See Seiichi Kawasaki and McMillan, 1987 p. 337; Russel Johnston and Paul R. Lawrence, 1988 p. 101; Mari Sako, 1992 p. 159.) Of course, this threat carries weight only if it is both credible and would harm the supplier. Credibility stems in part from the fact that Japanese manufacturers tend to be much larger than their suppliers, and often have more-or-less costless access to several qualified replacement firms (Banri Asanuma, 1992). A cutoff hurts the supplier in several ways. One source of harm is that average risk adjusted profits to Japanese subcontractors are positive so that severance jeopardizes these returns (Kawasaki and McMillan, p. 339). Perhaps more important, the loss of a contract harms the reputation of the supplier in the community so that ''a firm which gets 'broken off' earns a bad reputation and would have difficulty finding other partners'' (Sako, p. 159).[2]

The American and Japanese automobile industries provide an excellent illustration of the two systems of managing suppliers. In the American automobile industry it was customary until recently to award supply contracts to the lowest bidder on an annual basis. The Japanese, in contrast, award a contract for the model life of the car (usually four years), and give an incumbent supplier who performs well preferential treatment at renewal. Not surprisingly, in 1986 General Motors (GM) dealt with more than 5,000 potential parts and component suppliers for its North American production, while Toyota dealt with fewer than 250 suppliers for production in Japan. A typical GM assembly plant deals with approximately 800 potential suppliers, while a typical Toyota plant (which has twice the capacity of its GM counterpart) deals with only 125 suppliers (Asanuma, 1988 pp. 5–6). These large differences are explained in part by the fact that American auto producers are more vertically integrated and, hence, purchase more raw materials and basic parts. (In 1993 GM produced more than 65 percent of the parts used in its cars, while Toyota produced less than 30 percent [McMillan, 1994 p. 204].) Another important reason, however, is that Japanese auto producers reward the satisfactory performance of their suppliers with contract renewals rather than putting each new job out for bid. This view is supported by Jeffrey H. Dyer and William G. Ouchi (1993 p. 54) who observe that,

> While a supplier may have a contract for four years, in reality Toyota's and Nissan's partners have essentially open-ended contracts that according to a Nissan purchasing manager ''have no real termination date.'' As one Toyota supplier said, ''Once we win the business, it is basically our business unless

[2] Japanese-style procurement contracts also correspond to what Oliver E. Williamson (1979, 1985) describes as a ''relational contract,'' and Victor Goldberg (1976) describes as an ''administered contract.''

we don't perform. It is our business to lose.''

In fact, the practice of awarding open-ended, long-term supply contracts is widespread throughout Japanese industry. A survey of subcontracting firms reported in 1984 found that 84 percent had not had a change of primary buyer in the previous five years. In 1988 a similar study found that 68.2 percent of all small- and medium-sized subcontractors had apparently never changed their largest parental company (Sako, 1992 p. 170). The longevity of Japanese supply relationships stands in stark contrast to the short-term trading horizon occasioned by the competitive bidding system traditionally used by American firms.

The two methods for providing incentives for suppliers (up-front inspections and contract renewals) translate naturally into two fundamentally different systems for playing a dynamic procurement game of incomplete information. We analyze such a game below, and discover theoretical support for many commonly observed features of the two procurement systems. We first determine that the two modes of procurement are true ''systems'' in the sense that it never pays for a buyer to simultaneously use both up-front inspections and contract renewals to provide incentives. Hence, it is appropriate to focus our analysis on the pure American system (in which inspections provide incentives for quality assurance and contract renewals play no role) and the pure Japanese system (in which inspections play no role and contract renewals provide incentives for quality assurance).

Investigating these two systems, we find that in equilibrium under American-style procurement contracts: (i) the buyer may use a large number of suppliers over time; (ii) suppliers make zero expected profit; (iii) buyer costs are higher than first best because inspections are costly; and (iv) lot sizes and inventory levels are higher than first best and the frequency of orders is correspondingly low. We find, in contrast, that in equilibrium under Japanese-style contracts: (i) the buyer selects a single supplier and never switches; (ii) the supplier earns positive expected profit; (iii) buyer costs are higher than the first-best level because the supplier receives a quality premium; and (iv) lot sizes and inventory levels are lower than first best and the frequency of orders is correspondingly high. Finally, we find that equilibrium quality under either procurement system is typically quite close to the efficient standard.

These results are of considerable interest because shifting from the American-style to the Japanese-style system of procurement is increasingly fashionable. Asanuma (1988) reports that American automakers began to forge long-run Japanese-style partnerships with many of their suppliers in the mid-1980's. McMillan (1994) documents similar recent changes at Xerox, Boeing, and Motorola and throughout the American textile and machine tool industries. Indeed, Faye W. Gilbert et al. (1994) surveyed manufacturing firms capable of implementing JIT purchasing systems located in the southern United States. Of the 107 respondents, 68 claimed to be using some facets of JIT purchasing. Milgrom and Roberts (1990, 1995) also note that firms worldwide appear to be adopting Japanese-style procurement practices as part of a general modern economic strategy.

Although our focus and results differ from those of Milgrom and Roberts (1990, 1995), our findings underscore the link these authors draw between the implementation of flexible manufacturing equipment and the adoption of Japanese-style procurement practices. Numerically controlled machine tools that can dramatically reduce the cost of setting up a production run were first used widely in Japan, and have spread to the United States and other Western countries only in recent years. The results we present below indicate that falling set-up costs from the implementation and improvement of such equipment will eventually lead firms to switch from the American-style to the Japanese-style mode of procurement in a ''bang-bang'' fashion.[3] Thus, our analysis points to particular characteristics that influence the desirability of the two systems, identifies complementarities in implementation, and shows that it is important to carefully eval-

[3] Susan Athey and Armin Schmutzler (1995) analyze the product and process research strategy of a firm engaged in ''modern manufacturing.''

uate the entire system rather than individual components. We also show that even when systems differ radically, costs and quality may not differ much, and we identify several specific hypotheses that can be used to test the theory.[4]

I. The Model

Consider two firms, a buyer and a supplier.[5] The supplier produces output in batches and has costs of $f + c(\theta)x$, where x is the number of units produced, and θ is the proportion of nondefective units (the "quality" of the lot), which is chosen deterministically by the supplier. The nonnegative parameter f represents the fixed costs of setting up and delivering a production run. The marginal production cost $c(\cdot)$ is nonnegative, increasing, convex, and twice continuously differentiable in $\theta \in [\underline{\theta}, \bar{\theta}]$, where $0 \leq \underline{\theta} < \bar{\theta} \leq 1$. Both firms are expected profit maximizers, have reservation profit levels of zero, and discount in continuous time at interest rate r.[6] The buyer is assumed to possess all the bargaining power.

Suppose the buyer draws down its inventory of supplies at an exogenous rate q, and that it wishes the proportion of nondefective supplies to be at least $\theta > \underline{\theta}$.[7] Further assume the actual proportion of defectives in a lot is nonverifiable and, hence, cannot be contracted upon directly.

We consider two incentive mechanisms for inducing the desired level of quality in this setting, "American-style" and "Japanese style" procurement. Each of these systems is defined fundamentally in our analysis by the buyer's decision to inspect deliveries; incoming shipments are inspected under American-style procurement while under Japanese-style procurement they are not. This difference in the decision to inspect in turn leads to the other disparities in the procurement arrangements, including differences in lot sizes and inventory levels, supplier profitability, and the longevity of trading relationships.

Under American-style procurement the buyer inspects a shipment on delivery and either accepts or rejects the lot without payment depending on the outcome of the inspection. Acceptance sampling of a lot with an unknown proportion of defectives was first studied by Abraham Wald (1947), and has been refined by such authors as Anders Hald (1981) and Edward G. Schilling (1982). Following this literature, if the buyer conducts an inspection, it is assumed to sample units at random sequentially from a given lot and to classify each unit it examines as either defective or nondefective. After examination of each unit, the buyer applies the probability ratio test suggested by Wald (p. 90) that indicates whether the lot should be accepted or rejected or whether sampling should continue. In particular, the buyer chooses a tolerance band (θ_0, θ_1) around the desired quality level θ. The error probabilities associated with accepting the lot when its true quality is less than θ_0 or rejecting the lot when its true quality is greater than θ_1 are also chosen by the buyer, and define the power of the test.

For simplicity, assume that the tolerance band is narrow and the error probabilities are negligible; i.e., if the buyer inspects a shipment, it employs a powerful test and essentially learns the true proportion of defectives. Also, suppose that each lot is large enough that examination of successive units can be viewed as essentially independent; i.e., the difference between sampling with and without replacement is negligible. This common assumption implies that the expected sample size at which the inspection is terminated does not depend on lot size.[8] Hence, the expected cost of

[4] Two technical lemmas (2 and 3) and proofs of most of the main results appear in the Appendix. To save space, several straightforward proofs have been omitted but are available from the authors in a technical Appendix.

[5] The analysis can be applied to any two firms (or any two divisions within a firm) situated vertically on the value-added chain, such as an original equipment manufacturer and parts supplier, or a retail firm and wholesaler.

[6] Kawasaki and McMillan (1987) reject statistically the hypothesis that Japanese subcontractors are risk neutral. Adding risk aversion to our model would strengthen the case for Japanese-style procurement because of the random inspection policy used under American-style procurement.

[7] The choice of θ is considered in Section III below.

[8] Expounding on this point, Hald (1981 p. 38) writes: "The discriminating power of the sampling plan is thus determined essentially by the number of items in the sample, not by the percentage of items inspected."

inspecting a shipment is also independent of lot size and is denoted here by m.[9] Finally, it will generally be unnecessary for the buyer to inspect every shipment. Instead, it is assumed the buyer performs random inspections and commits to an inspection probability (intensity) of $\alpha \in [0, 1]$.[10]

Although inspection outcomes are not contractible, payment to the supplier can be made contingent on the buyer accepting delivery. Once the buyer accepts a shipment, however, it is assumed to have no meaningful legal recourse if it subsequently discovers the lot to contain an unsatisfactory proportion of defectives.[11] If the buyer rejects a shipment, the seller forfeits current payment on the output produced.

The second method for inducing supplier performance is associated with Japanese-style procurement. Specifically, the buyer can make an informal promise to award the supplier contract renewals so long as it performs satisfactorily. It is assumed that if the buyer does not perform an inspection, it learns quality shortly after accepting delivery of a shipment. If the supplier earns positive profit on each contract, then the threat of nonrenewal will provide an incentive to furnish the desired level of quality.

Assume the buyer has access to a pool of equally qualified suppliers and the time associated with writing a contract and producing a lot is negligible. Thus, the buyer is not injured if it rejects delivery of a shipment or terminates a supply relationship. This, however, is not the case for the supplier. The scrap value of a rejected shipment to the supplier is taken to be zero, so that rejection of delivery means the supplier receives no revenue on the lot.[12] In addition, if the buyer fails to renew a contract the supplier is assumed to suffer a loss of reputation and any associated future quasi-rent stream; i.e., it receives its reservation profit of zero from then on.[13] Hence, under the American system, incentives are provided primarily by a threat to withhold payment and give a supplier caught shirking negative *current* profit, while under the Japanese system incentives are provided by the threat to cut off a shirker and eliminate its *future* reputational returns.

We consider a general model where the two procurement systems just described are distinct solutions. Any supply contract in this setting consists of five elements: an inspection probability α, a lot size x, an implied order frequency (q/x), a quality level θ, and a delivery-contingent payment $w(x, \theta)$. Because the buyer possesses all the bargaining power, it will choose x, α, and $w(x, \theta)$ so as to minimize the expected present value of its procurement and inspection costs subject to the supplier's incentive compatibility and individual rationality constraints.

Note that the time between orders (the inverse of the order frequency) is (x/q), and hence the discount factor is

$$\delta \equiv e^{-hx},$$

[9] Although m does not depend on x, it does depend on θ. The properties of this dependence are governed by technological factors and the form of the buyer's loss function, which are suppressed here. Note, however, that none of our results, including Proposition 6, requires any properties other than $m \geq 0$ and $|dm/d\theta| < \infty$.

[10] The assumption of commitment to an inspection intensity greatly simplifies the analysis and is standard in the auditing literature. The American system may still dominate even if commitment to α is not possible, but then one would expect suppliers to cheat with positive probability. Some insight into this problem is provided by our earlier working paper (Taylor and Wiggins, 1994), which analyzes a conceptually similar model where there is no commitment.

[11] It is relatively straightforward for a court to verify acceptance of a shipment, whereas verifying its quality and whether it was handled appropriately by the buyer after delivery is apt to be prohibitively costly. This is undoubtedly the reason Sako (1992 p. 11), commenting on American-style procurement, notes that: "The principle of caveat emptor predominates."

[12] In fact, a supplier might receive some residual benefit from a rejected shipment; e.g., it might sell the lot or some fraction of it to another buyer at a later date. The critical feature is that rejection hurts the supplier as compared with acceptance. This is well known among the proponents of acceptance sampling. For instance, Hald (1981 p. 6) writes: "It is of utmost importance that the rejection of a lot should be of economic consequence to the supplier ... ." If the injury from rejection is small, then the buyer will be compelled to inspect more intensively or to give more rent to the supplier on each contract.

[13] Again, it is not essential that the supplier suffer a complete reputation loss, as is assumed here for simplicity.

where $h \equiv (r/q)$. It is incentive compatible for the supplier to deliver shipments of quality θ, if it does better by doing this and receiving contract renewals than by delivering shipments of any other quality. The supplier obviously has no incentive to furnish quality higher than θ. Moreover, there is no point in attempting to shirk just a little because an inspection (if one is performed) reveals the true quality of a shipment. This means the supplier will either comply with the contract, or attempt to deliver a shipment of the lowest possible quality $\underline{\theta}$. As a result, the incentive compatibility constraint is given by

$$(1) \quad \frac{w(x,\theta) - f - c(\theta)x}{1-\delta} \geq (1-\alpha)w(x,\theta) - f - c(\underline{\theta})x.$$

The left side of this inequality is the net present value of continually complying with the contract and receiving renewals. The right side is the expected benefit of attempting to deliver a shipment of quality $\underline{\theta}$. Specifically, the supplier bears the production cost $f + c(\underline{\theta})x$, and receives the payment $w(x, \theta)$ if the buyer does not perform an inspection; i.e., with probability $(1 - \alpha)$. Even if the buyer does not perform an inspection, however, it ultimately observes the low quality of the shipment and severs its relationship with the supplier before placing another order—though this threat is only meaningful to the supplier if there is a quasi-rent flow under the contract.

The individual rationality constraint is simply that the supplier not earn negative profit by complying with the contract

$$(2) \quad w(x,\theta) - f - c(\theta)x \geq 0.$$

For a given θ, the buyer's contract design problem is thus given by

$$\min_{\langle x,\alpha,w(\cdot,\theta)\rangle} \frac{w(x,\theta) + \alpha m}{1-\delta},$$

subject to (1) and (2).

The problem can be simplified by noting that the buyer will pay the supplier no more than required by the incentive compatibility constraint (1), making it possible to write this constraint as an equality which can then be used to eliminate $w(x, \theta)$ from the problem.[14] Substitution yields the following cost-minimization problem:

$$(3) \quad \min_{\langle x,\alpha\rangle} K(x,\alpha,\theta) \equiv \frac{f + c(\theta)x}{1-\delta} + \frac{\alpha m}{1-\delta} + \frac{(\bar{\alpha}(x,\theta) - \alpha)(f + c(\theta)x)}{1-(1-\alpha)(1-\delta)},$$

subject to

$$(4) \quad \alpha \geq 0$$

and

$$(5) \quad \alpha \leq \bar{\alpha}(x,\theta),$$

where

$$(6) \quad \bar{\alpha}(x,\theta) \equiv \frac{(c(\theta) - c(\underline{\theta}))x}{f + c(\theta)x}.$$

The first term on the right side of (3) is the present value of production costs; the second term is the expected present value of inspection costs which can be thought of as the agency cost associated with American-style procurement; and the third term is the expected present value of the incentive rent paid to the supplier on each contract. This rent stream can be thought of as the agency cost associated with Japanese-style procurement. The nonnegativity condition (4) simply indicates that zero is the lowest probability with which an inspection can be performed.[15] Condition (5) is a reformulation of the individual

[14] To see why (1) can always be written as an equality, note that if (2) binds but (1) does not, the buyer can reduce α unless it already equals zero. Moreover, if $\alpha = 0$, then (2) cannot bind, or else (1) would boil down to the contradiction $c(\theta) \leq c(\underline{\theta})$. Hence, the buyer can always reduce α until (1) binds.

[15] Technically, there is also a nonnegativity condition for x, but such a constraint never binds when $f > 0$ because the present value of production costs becomes infinite as x approaches zero.

rationality constraint (2) when (1) binds. Specifically, if $\alpha > \bar{\alpha}(x, \theta)$ and (1) binds, then the supplier earns negative profit, as is clear from the third term on the right side of (3). The supplier receives positive rent so long as $\bar{\alpha}(x, \theta) - \alpha > 0$, and just breaks even when (5) binds. In other words, if the buyer pays the supplier the minimal incentive compatible rent, then this imposes an upper bound on its inspection intensity. The boundary, $\bar{\alpha}(x, \theta)$, starts at the origin and increases monotonically in x toward its asymptotic value $[c(\theta) - c(\underline{\theta})]/c(\theta)$. The set of points $\{(x, \alpha)\}$ satisfying (4) and (5) is referred to below as the "feasible region."

Note that the present value of inspection costs $[\alpha m/(1 - \delta)]$ is increasing in α and decreasing in x due to economies of scale in inspections. On the other hand, the expected present value of agency rents

$$\frac{(\bar{\alpha}(x, \theta) - \alpha)(f + c(\theta)x)}{1 - (1 - \alpha)(1 - \delta)}$$

is increasing in x and decreasing in α because the incentive to shirk on quality is positively related to lot size and negatively related to inspection intensity. Hence, the expected present value of inspection costs will be high precisely when the expected present value of agency rents is low and vice versa. This suggests a tension between the two incentive mechanisms (inspections and contract renewals), which is formalized in the following lemma.

LEMMA 1: *Lot size and inspection intensity are complements over the feasible region,*

$$K_{x,\alpha}(x, \alpha, \theta) \leq 0, \qquad \text{for } \alpha \in [0, \bar{\alpha}(x, \theta)].$$

Remark.—For differentiable functions, nonpositive cross-partial derivatives are equivalent to submodularity. Moreover, the feasible region is a sublattice of $\Re^2$ under the conventional Euclidean ordering. Hence, the monotone methods for analyzing comparative statics first articulated by Donald M. Topkis (1978) and introduced into economics by Milgrom and Roberts (1990, 1995) and Milgrom and Christina Shannon (1994) could be applied here. We, however, do not employ this methodology because our cost-minimization problem (while nonconvex) is differentiable. Also, several of our most important results are "limit theorems" for which monotone methods are ill suited.

Consider the equilibrium contract solving (3). If this contract relies more on inspections than on contract renewals to regulate supplier behavior (i.e., if it looks more like an American-style procurement contract), then Lemma 1 indicates that it will involve comparatively large orders and inventory levels and low supplier profit. Conversely, if the equilibrium contract relies more on renewals to provide incentives (i.e., if it looks more like a Japanese-style procurement contract), then it will involve comparatively small orders and inventory levels and high supplier profit. The rather complicated form of the objective function makes further analysis of the equilibrium contract problematic in general.

In order to develop further results from the model, we consider cases in which the interest rate divided by the use rate h is taken to be small. This assumption seems appropriate in all reasonable applications and makes it possible to simplify greatly the buyer's contract design problem.[16] Specifically Lemma 2 in the Appendix shows that if (x, α) is a potential solution to the buyer's problem, then $\lim_{h \to 0} hx = 0$. Hence although any potentially optimal lot size x becomes arbitrarily large as h becomes small, the divergence is slow enough that hx becomes arbitrarily small. Moreover, when hx is small, δ is approximated very closely by the second-order Padé expansion[17]

$$\tilde{\delta} = \frac{2 - hx}{2 + hx}.$$

[16] Technically, the unit-free number hx is required to be small because the magnitude of h alone depends on the units in which the supplier's output is measured. This is somewhat awkward because x is a choice variable, not a parameter. Fortunately, Lemma 2 (in the Appendix) resolves this problem and results below are stated in terms of small h with the understanding that this implies small hx.

[17] In this setting (as is frequently the case) a Padé expansion is algebraically more convenient and often more accurate than a Taylor-series expansion (see Kenneth Judd, 1997 Ch. 6). The Padé expansion is used here only for algebraic convenience and to obtain closed-form solutions.

Substituting $\tilde{\delta}$ into (3) transforms the buyer's problem into

$$(7)\quad \min_{\langle x,\alpha\rangle} \tilde{K}(x, \alpha, \theta) \equiv \phi(x)(f + c(\theta)x + \alpha m) + \gamma(x, \alpha)(\bar{\alpha}(x, \theta) - \alpha) \times (f + c(\theta)x),$$

subject to (4) and (5), where

$$\phi(x) \equiv (1 - \tilde{\delta})^{-1}$$

and

$$\gamma(x, \alpha) \equiv [1 - (1 - \alpha)(1 - \tilde{\delta})]^{-1}.$$

In the analysis that follows, we sometimes require h not only to be small, but *arbitrarily* small in order to prove limit theorems concerning solutions to the cost-minimization problem. We identify these situations by writing "for h sufficiently small."[18]

An extreme point of $\tilde{K}(x, \alpha, \theta)$ is any point (x_0, α_0) satisfying the first-order conditions

$$(8)\quad \tilde{K}_x(x_0, \alpha_0, \theta) = \phi'(x_0)(f + \alpha_0 m) + (c(\theta)/2) + \gamma(x_0, \alpha_0) \times [(1 - \alpha_0)c(\theta) - c(\underline{\theta})] + \gamma_x(x_0, \alpha_0)(\bar{\alpha}(x_0, \theta) - \alpha_0) \times (f + c(\theta)x_0) = 0$$

and

$$(9)\quad \tilde{K}_\alpha(x_0, \alpha_0, \theta) = \phi(x_0)m + \gamma(x_0, \alpha_0)(f + c(\theta)x_0) + \gamma_\alpha(x_0, \alpha_0) \times (\bar{\alpha}(x_0, \theta) - \alpha_0)(f + c(\theta)x_0) = 0.$$

Lemma 3 in the Appendix shows that as h approaches zero, any x_0 must go to infinity at rate $(1/\sqrt{h})$ and any α_0 must converge to the finite value $\hat{\alpha}_0$. These facts make it possible to prove a key result of the paper.

PROPOSITION 1: *For h sufficiently small, the equilibrium procurement contract solving (7) must be either a purely American-style contract with $\alpha = \bar{\alpha}(x, \theta)$ or a purely Japanese-style contract with $\alpha = 0$.*

PROOF:

The claim is established by showing that when h is sufficiently small $\tilde{K}(x, \alpha, \theta)$ possesses a unique extreme point (x_0, α_0). Moreover, (x_0, α_0) is a saddlepoint, which implies that a corner solution involving either $\alpha = \bar{\alpha}(x, \theta)$ or $\alpha = 0$ must obtain. See the Appendix for the algebraic details.

The intuition behind Proposition 1 is straightforward. The complementarity between lot size and inspection intensity identified in Lemma 1 is so strong that it never pays to use inspections and contract renewals simultaneously to provide incentives. The reason is that the marginal returns to inspections are highest when lots are large, and the marginal returns to contract renewals are highest when lots are small. Hence, using both inspections and renewals and ordering intermediate-sized lots is dominated by moving to one extreme or the other, exclusively using one incentive system.

To illustrate the significance of the complementarity between the choice variables, suppose h is small and the saddlepoint (x_0, α_0) lies in the feasible region. Then, x_0 is the cost-minimizing lot size holding inspection intensity fixed at α_0, and α_0 is the cost-minimizing inspection intensity holding the lot size fixed at x_0. In other words, it is not possible to

[18] While arbitrarily small h dramatically simplifies the analysis, to our knowledge, it is a necessary condition only for the second part of Proposition 3.

reduce costs below $\tilde{K}(x_0, \alpha_0, \theta)$ by changing either lot size or inspection intensity alone. Proposition 1, however, indicates that it is possible to reduce costs by exploiting the complementarity between the choice variables and decreasing or increasing lot size and inspection intensity together. In fact, it is possible to continue to reduce costs in either of these ways until a boundary of the feasible region is reached; i.e., until (4) or (5) binds. Of course, if (x_0, α_0) does not occur in the feasible region, then a corner solution to (7) still obtains along the boundary furthest from the saddlepoint. Hence, cost minimization involves either heavy inspections, larger less frequent orders, high inventory levels, supplier profit of zero, and a job-by-job relationship (the pure American system), or no inspections, smaller more frequent orders, low inventory levels, positive supplier profit, and an ongoing relationship (the pure Japanese system).

Remark.—In practice, American-style procurement is usually implemented via competitive bidding for orders by several potential suppliers. This is consistent with Proposition 1 because the American system does not use contract renewals to provide incentives and hence there is no particular reason to use the same supplier over time. In fact, if there is cost heterogeneity among potential suppliers that varies over time (not modeled here), then awarding a supply contract on the basis of competitive bidding is desirable for the buyer and also enhances efficiency. This contrasts sharply with the Japanese system where incentives are provided exclusively through the promise of contract renewals, and as a result a long-term supply relationship is an essential ingredient. The importance of contract renewal creates an incentive to restrict the number of suppliers and avoid bidding once the relationship has begun (see the discussion of multiple-sourcing under Japanese-style procurement in Section IV).[19]

II. Comparing the Procurement Systems

It turns out that neither the American nor the Japanese system is generally first-best efficient. In order to demonstrate this, the following benchmark is necessary.

PROPOSITION 2: *The first-best lot size is given by*

$$x_{\mathrm{FB}} = \sqrt{\frac{2qf}{rc(\theta)}}. \tag{10}$$

Expression (10) is the celebrated square-root order-quantity rule known by management scientists since the 1920's and introduced to mainstream economics by William J. Baumol (1952) and James Tobin (1956), who studied the inventory-theoretic demand for cash. In the present context, the square-root rule balances the frequency with which the fixed set-up/delivery cost is borne against the financial cost of carrying inventory. Accordingly, a rise in q or f causes the first-best lot size to increase, while a rise in r or $c(\theta)$ causes it to decrease.

PROPOSITION 3: *The equilibrium lot size under the Japanese system is smaller than first best, $x_{\mathrm{J}} < x_{\mathrm{FB}}$, and the equilibrium lot size under the American system is larger than first best for $m > 0$ and h sufficiently small, $x_{\mathrm{A}} > x_{\mathrm{FB}}$.*

Both Japanese-style and American-style procurement typically involve inefficient lot sizes. Under Japanese-style contracting, there are two reasons for placing smaller than first-best orders. To see this, consider the agency cost under this system

$$\gamma(x, 0)\bar{\alpha}(x, \theta)(f + c(\theta)x) = \frac{[c(\theta) - c(\underline{\theta})]x}{\tilde{\delta}}. \tag{11}$$

[19] Of course, a buyer using the Japanese system would like to capture the capitalized value of the agency rents at the time it selects a supplier. Requiring candidates to bid for the right to be a long-term supplier, however, is problematic because this may provide the buyer with an incentive to opportunistically sever the relationship in order to resell the right. In practice, Japanese firms appear to compete for the right to be a long-term supplier by offering to make relationship-specific investments such as "... significant customized investments in plant, equipment, and personnel ..." often including "... building a supplier plant within 15 miles of the customer plant ... ." (Dyer and Ouchi, 1993 pp. 53, 55).

The expression on the right is familiar from the reputation literature pioneered by Benjamin Klein and Keith B. Leffler (1981) and Carl Shapiro (1982, 1983). It is the present value of the stream of quality premiums earned by the seller of an experience good. The two reasons for the buyer to place inefficiently small orders can be identified respectively with the numerator and the denominator. The numerator $(c(\theta) - c(\underline{\theta}))x$ is increasing in x because the smaller the order, the smaller the gain to the supplier from shirking, and the smaller the required incentive payment. The denominator $\tilde{\delta}$ is decreasing in x because the smaller the order, the shorter the period between contract renewals, and the less heavily the supplier discounts its rent stream. In other words, if the buyer renews the contract more frequently, it can reduce the per-unit agency rent it must pay to the supplier. Thus, although x_{FB} minimizes the present value of production cost, the buyer finds it advantageous to place smaller orders under Japanese-style contracting. The associated productive inefficiency is more than offset by savings in agency cost.

Next, consider the agency cost under American-style procurement

$$\text{(12)} \quad \phi(x)\bar{\alpha}(x,\theta)m = \frac{[(c(\theta)-c(\underline{\theta}))x/(f+c(\theta)x)]m}{1-\tilde{\delta}}.$$

Again, there are two effects, one associated with the numerator of the ratio on the right and one associated with the denominator. This time, however, the effects work in opposite directions. Specifically, the denominator, $1 - \tilde{\delta}$ is increasing in x because the larger the lot, the longer between orders and potential inspections. On the other hand, the numerator, $[(c(\theta) - c(\underline{\theta}))x/(f + c(\theta)x)]m$, is also increasing in x because the larger the order, the larger the gain to the supplier from shirking, and the higher the required inspection intensity. Proposition 3 indicates that when h is sufficiently small (as it generally is) the former effect dominates, and it behooves the buyer to order larger than efficient lots. This is because when h is small, x_{A} is large and the inspection intensity $\bar{\alpha}(x_{\text{A}}, \theta)$ is close to its asymptotic value $[c(\theta) - c(\underline{\theta})]/c(\theta)$; i.e., it is nearly flat and, hence, changes very little with lot size. In other words, large x_{A} implies that expected inspection costs depend very little on lot size because

$$\bar{\alpha}_x(x_{\text{A}}, \theta)m = \frac{f\,(c(\theta) - c(\underline{\theta}))}{(f + c(\theta)x_{\text{A}})^2}\,m$$

is very small (on the order of $1/x_{\text{A}}^2$).

One implication of Proposition 3 is that the two methods of procurement cannot generally be ranked in terms of efficiency. For instance, in the limit as the expected inspection cost m goes to zero, American-style procurement becomes fully efficient. If m is large, however, the American system will involve significant deadweight loss due both to the distortion in lot size and the expected cost of performing inspections. Under the Japanese system, in contrast, the only source of inefficiency derives from inappropriately small orders because the agency cost is merely a transfer from the buyer to the supplier.

Observe from (11) and (12) that both systems are fully efficient when the incentive to commit moral hazard disappears; i.e., when $\theta = \underline{\theta}$. In this case, of course, it is not necessary to pay a quality premium or to perform inspections. A less obvious and more important point concerns the role of the set-up/delivery cost f.

PROPOSITION 4: *In the limit as f goes to zero, the Japanese procurement system becomes fully efficient while the American system does not. In particular, for $m > 0$*

$$\lim_{f\to 0} \tilde{K}(x_{\text{J}}, 0, \theta) = \lim_{f\to 0} \tilde{K}(x_{\text{FB}}, 0, \theta) = \frac{c(\theta)q}{r} < \lim_{f\to 0} \tilde{K}(x_{\text{A}}, \bar{\alpha}(x_{\text{A}}, \theta), \theta).$$

Recall the reason for periodic production runs, to balance the frequency with which the fixed set-up/delivery cost is incurred against the cost of carrying inventory. As f becomes small, the frequency with which it is incurred becomes less important, and the lot sizes x_{J}, x_{FB}, and x_{A} all accordingly shrink to economize on carrying cost. In fact, x_{J} and x_{FB} both go to zero in the limit and production becomes a continuous flow that matches the use rate of

q. This, however, is not the case under the American procurement system. As f becomes small, the inspection intensity $\bar{\alpha}(x, \theta)$ converges to its limiting value $[c(\theta) - c(\underline{\theta})]/c(\theta)$, which does not depend on x. Hence the expected inspection cost $\bar{\alpha}(x, \theta)m$ is a fixed cost in the limit, and the optimal lot size x_A is therefore strictly positive.[20]

Remark.—A closely related result is that buyers that use the Japanese-style procurement system have an additional incentive at the margin to seek out suppliers with low set-up/delivery costs, while firms that use the American-style system have less incentive to seek out such suppliers. This is easily verified by applying the envelope theorem as follows:

$$\frac{\partial \tilde{K}(x_J, 0, \theta)}{\partial f} = \phi(x_J)$$

and

$$\frac{\partial \tilde{K}(x_A, \bar{\alpha}(x_A, \theta), \theta)}{\partial f} = \phi(x_A)\left(1 - \frac{[c(\theta) - c(\underline{\theta})]x_A}{[f + c(\theta)x_A]^2}\right) < \phi(x_A).$$

Thus, $x_J < x_A$ implies that a reduction in f is more valuable to a buyer procuring inputs via the Japanese system than to one using the American system. This is because the smaller lots associated with Japanese-style procurement generate a shorter inventory cycle and more frequent orders.

So far, it has been shown that as m becomes small, the American system becomes efficient and is the preferable mode of procurement for the buyer. Alternatively, as f becomes small, the Japanese system becomes efficient and is the preferable mode. The following result sharpens these observations and provides some valuable additional insights.

PROPOSITION 5: *For h sufficiently small, the buyer strictly prefers Japanese-style to American-style procurement if, and only if,*

$$\frac{f}{m} < \frac{1}{2}.$$

This result is noteworthy in part because it provides a simple criterion for a buyer to evaluate the two systems. Specifically, if $f/m < 1/2$, then the Japanese system minimizes the present value of procurement costs; if $f/m > 1/2$, then the American system is optimal; and if $f/m = 1/2$, then the two systems are equally costly. This argument is valid, moreover, regardless of whether the saddlepoint (x_0, α_0) lies above, below, or within the feasible region.

To investigate this further, note that as h goes to zero, $\bar{\alpha}(x_0, \theta)$ converges to $[c(\theta) - c(\underline{\theta})]/c(\theta)$. Set this limiting value of $\bar{\alpha}(x_0, \theta)$ less than the limiting value $\hat{\alpha}_0$ given in (A1) (in Lemma 3 in the Appendix) and collect terms to get the following necessary and sufficient condition for the saddlepoint to lie "above" the feasible region when h is small,

$$\text{(13)} \qquad \frac{f}{m} < \frac{1}{2} - \frac{c(\theta) - c(\underline{\theta})}{c(\theta)}.$$

Similarly, set the limiting value $\hat{\alpha}_0$ from (A1) less than zero and rearrange terms to get the following necessary and sufficient condition for the saddlepoint to lie "below" the feasible region when h is small,

$$\text{(14)} \qquad \frac{f}{m} > \frac{1}{2} + \frac{c(\theta) - c(\underline{\theta})}{c(\theta)}.$$

When h is small and (13) holds (so that the saddlepoint is above the feasible region), the Japanese system, $x = x_J$ and $\alpha = 0$, is the unique local (and hence the global) solution to the buyers contract design problem (7). A buyer using American-style procurement in this situation would be better off switching to

[20] Two caveats are warranted here. First, as x_J becomes small, the lag between when the buyer accepts a lot and when it ultimately learns quality is relevant because the supplier may be able to shirk on several small shipments before the buyer terminates the contract. As stated, Proposition 4 assumes the lag to be zero, but it is straightforward (although messy) to generalize this. Second, when x_A is small, the assumption that m is independent of lot size may be inappropriate. However, so long as American-style procurement involves costly monitoring of some kind, the gist of Proposition 4 holds.

the Japanese system, but even a small reduction in lot size and inspection intensity would lower its costs. In other words, local incentives for change coincide with the global incentives. Similarly, when h is small and (14) holds (so that the saddlepoint is below the feasible region), then the American system, $x = x_A$ and $\alpha = \bar{\alpha}(x_A, \theta)$, is the unique local solution to (7), so that a buyer using Japanese-style procurement would lower its costs even if it only increased lot size and inspection intensity slightly. When h is small, however, but neither (13) nor (14) holds, then (x_0, α_0) lies in the feasible region and both systems are local solutions to (7). In this instance, local incentives for change will not be aligned with the global incentives. In other words, even though a buyer might be better off switching from the American procurement system to the Japanese system (or vice versa), a slight decrease (increase) in lot size and inspection intensity will raise costs; i.e., large-scale shifts in both behavior and organizational form are required in order to make the buyer better off.

These arguments highlight the existence of an intriguing discontinuous comparative static. Specifically, if f/m moves from slightly above to slightly below (or from slightly below to slightly above) $1/2$, then the optimal mode of procurement will switch from the American to the Japanese system (or vice versa). In other words, a modest change in parameter values can give rise to a large shift in cost-minimizing behavior and organizational form that involves switching between heavy inspections and no inspections, between large orders and small orders, between competitive bidding on each job, and a long-term supply partnership. These changes, however, will be associated with only small cost savings so that firms may be slow to implement them.

As an example, consider Table 1, which was generated under the parameter values: $q = 200$ units per day, $r = 0.04$ percent per day, $m =$ \$2,000, $c(\theta) =$ \$20 per unit, and $c(\underline{\theta}) =$ \$10 per unit. The table compares the American and Japanese systems in a case where there are two local minima before and after a technological innovation giving rise to a 2-percent fall in f from \$1,010 to \$990. The top row shows the expected present value of procurement cost (minus 10 million dollars) in each case; the second row gives the agency cost which is either the expected present value of inspection costs or agency rents. The third and fourth rows show the lot sizes and inspection intensities, respectively.

It is easy to check that neither (13) nor (14) holds before or after the innovation in f. Thus, both systems are local minima of the cost function, although only one system corresponds to the global minimum in each case. Specifically, when set-up/delivery costs are \$1,010 on each production run (so that $f/m = 0.505$), then American-style procurement is \$830 *less* expensive than Japanese-style procurement. Suppose that set-up/delivery costs fall to \$990 due to technological innovation (so that $f/m =$ 0.495), then American-style procurement becomes \$250 *more* expensive than Japanese-style procurement. If the buyer switches modes of procurement in response to this innovation in f, the expected present value of its total cost savings is only \$1,260 or about 0.012 percent. Although these savings are trivial, the buyer's behavior and organizational form must change dramatically. In particular, it cuts the size of its orders from 10,000 to 4,950 units and the period between orders from 50 to 24.75 days. Instead of soliciting competitive zero-profit bids on each order, it forms a long-term partnership with a single supplier and pays it about \$490 profit on each order. Instead of inspecting nearly half the shipments it receives, it stops performing inspections altogether.

As the example suggests, when the saddlepoint lies in the feasible region (so that both systems are local minima), it may be difficult for a buyer using one procurement system to evaluate the merits of switching to the other dramatically different system. Indeed, making a switch requires large-scale changes in behavior and organizational form, while the procurement cost savings resulting from a switch may be very modest. For this reason, it should not be surprising to observe, for example, buyers in the same industry (but different countries) using different procurement systems.

Market conditions, moreover, may increase the incentives of buyers to coalesce around a single system, raising the costs of switching. If most buyers are using the American system, it might be difficult for a buyer to identify and educate potential suppliers about the merits of the

TABLE 1—A DISCONTINUOUS COMPARATIVE STATIC

	f = \$1,010		f = \$990	
	American	Japanese	American	Japanese
$\tilde{K} - 10^7$	\$201,250	\$202,080	\$200,240	\$199,990
Agency cost	\$50,250	\$50,500	\$50,510	\$49,990
x	10,000	5,000	9,950	4,950
α	0.4975	0.0000	0.4975	0.0000

Japanese system and how its incentives operate, which would be important if inspections were not undertaken. Hence conditions in the supplier market may create persistence in a system even if it would be profitable for a given buyer to switch. Indeed, "Japanese-owned plants in the U.S. purchase more than 50% of their auto components from Japanese suppliers." (Dyer and Ouchi, 1993 p. 51.) If, however, set-up/delivery costs decline over time because of technological innovation, as we discuss further below, then it seems likely that there may be a period where the systems coexist but a shift to Japanese-style procurement will ultimately occur.

Remark.—Inspection costs have also fallen over time due to technological innovation. However, the new inspection technologies such as "machine vision" are better adapted for continuous process inspection of simple parts and components rather than spot inspections of incoming shipments possibly containing complex components or subassemblies. (See Jose A. Ventura et al., 1988; Kenneth W. Chapman et al., 1990; Ventura and Sencer Yeralan, 1990; Ehsan Asoudegi, 1992.) Hence, these new inspection technologies may well reduce the cost of quality control at the supplier's plant, but probably impact the buyer's inspection cost, m, very little.

III. The Quality Standard

For expositional reasons, the preceding analysis exogenously fixed the acceptable proportion of defectives at some level θ. In general, however, it is important to endogenize θ and determine how this quality standard varies under the two procurement systems.

In order to analyze the buyer's choice of quality standard, it is necessary to consider its output market.[21] For brevity this is done with a minimum of additional modeling. Specifically, assume that a buyer that has not tarnished its reputation in the output market earns flow revenue of $p(\theta)q$ per unit of time. The price function $p(\theta)$ represents the willingness of consumers to pay for units of output that contain defective inputs with probability θ, and is assumed to be positive, increasing, strictly concave, and twice continuously differentiable.

Consumers cannot observe the buyer's procurement contracts or the supplier's behavior. Instead, in addition to their own experiences, they periodically observe market data regarding θ; e.g., from consumer reports, certified advertizing claims, or word of mouth.[22] As usual, if a buyer with a reputation for quality level θ ever sells output of a lower quality, then consumers learn this after a brief interval, and refuse to purchase the buyer's output from that time forward. Hence, a buyer must initially decide where on the price-quality spectrum it would like to operate. To ease notation, the buyer's own production costs are assumed to be netted out of the price function so that its profit-maximization problem can be written

$$(15) \qquad \max_{\langle x,\alpha,\theta\rangle} \Pi(x, \alpha, \theta) \equiv p(\theta)/h - \tilde{K}(x, \alpha, \theta)$$

subject to (4) and (5).

[21] For an analysis of the links between flexible manufacturing and market structure, see B. Curtis Eaton and Nicolas Schmitt (1994).

[22] If consumers used only their own experiences with the product, then demand would steadily erode over time unless $\theta = 1$.

Obviously, the buyer's choice of x and α concern only cost minimization. Hence, Proposition 1 applies, and when h is small, it can be assumed that the buyer either employs the pure Japanese or the pure American procurement system. Accordingly, let (x_J, θ_J) denote the solution to (15) when $\alpha = 0$ and let (x_A, θ_A) denote the solution when $\alpha = \bar{\alpha}(x, \theta)$. Also, let (x_{FB}, θ_{FB}) denote the first-best solution (when $m = 0$).

PROPOSITION 6: *For h sufficiently small, both θ_J and θ_A are arbitrarily close to θ_{FB}, and $x_J < x_{FB} < x_A$.*

Thus, when h is small, the quality standard under either system is approximately the same as would be set in the first-best world, although the distortions in lot size identified in the previous section persist. The intuition is straightforward. The direct effects under each system arising from $\Pi_{x,x}\Pi_{\theta,\theta} < 0$ favor $\theta_J < \theta_{FB}$, $\theta_A < \theta_{FB}$, and $x_J < x_{FB} < x_A$. Because $\Pi_{x,\theta} \neq 0$, however, there is an interaction or cross effect between lot size and quality that can potentially swamp the direct effects under either system. Proposition 6 indicates that as h becomes small, both the direct effects regarding quality and the cross effects become negligible, while the direct effects concerning lot size do not.

There is a final noteworthy point regarding quality. It is a straightforward comparative statics exercise to show that the equilibrium quality standard under both systems is decreasing in f, which accords nicely with the central theme of this paper. Specifically, as f drifts down over time due to technological innovation, quality will improve, lot sizes will decrease, and all firms in an industry will eventually switch from the American to the Japanese procurement system.

IV. Applications and Empirical Implications

Our analysis of the two procurement systems provides insight concerning a number of sourcing issues. In particular, it helps explain why procurement systems may differ across countries, and provides specific empirical predictions about associated cost conditions. The model predicts that set-up costs for individual production run together with inspection costs drive the optimal procurement system. Hence, there should be differences in manufacturing cost structures across countries associated with differences in procurement systems. The model also predicts that firms within an economy will exhibit similar cost differences associated with differences in procurement.

The model also holds some implications for product variety. Like Milgrom and Roberts (1990, 1995) who argue that variety and JIT procurement are complementary, we find that the American system facilitates large orders of identical goods. Because lots are apt to shrink as variety increases and because optimal lot sizes are smaller under the Japanese system, it will be less costly to provide more variety under this system or procurement. Hence, declining set-up costs should lead both to increased product variety and a switch to this system.

Another point concerns the number of suppliers and multiple-sourcing. A purchasing system akin to the Japanese method has been used in the United States for long periods in defense procurement.[23] Proposals for second-sourcing have frequently been advanced. These proposals seem sensible in that second-sourcing can keep costs down, and provide greater buyer flexibility in dealing with quality problems or supplier shirking. Michael H. Riordan and David E. M. Sappington (1989) have shown, however, that when the first source has a cost advantage, second-sourcing may be inefficient. Our analysis shows that in an environment where quality premiums address the moral hazard problem, introduction of a second source may be undesirable even under cost parity. For instance, if the buyer attempts to maintain Japanese-style partnerships concurrently with n identical suppliers, then its agency cost is

$$(\gamma(x, 0))^n \bar{\alpha}(x, \theta)(f + c(\theta)x) = \frac{(c(\theta) - c(\underline{\theta}))x}{\tilde{\delta}^n},$$

[23] See, for example, the discussion of quality premiums contained in William P. Rogerson (1989), who argues both that there are premiums over marginal cost in military procurement and that these premiums serve an efficiency purpose in enhancing development of effective weapons systems.

which is increasing in n. Because suppliers must wait longer between orders under multiple-sourcing, they discount future profits more heavily according to the discount factor $\tilde{\delta}^n < \tilde{\delta}$. This requires payment of a higher premium. The buyer will ameliorate this to some extent by further reducing the size of its orders below x_J. Hence, in this context multiple-sourcing hurts the buyer and further impairs efficiency.[24]

At a more fundamental level, multiple-sourcing through competitive bidding will undermine the solution to the agency problem provided by the Japanese system. This system uses quasi-rent streams to discipline sellers, whereas a competitive bidding system extracts rents. Moreover, there does not appear to be any ready solution to this rent extraction problem since suppliers' bids will be distorted by any effort to restore rents in the final pricing of the lot.

These problems may also help explain the apparent reluctance of Japanese firms to open their supplier markets to U.S. companies. American companies have repeatedly claimed that it is difficult for them to participate in Japanese supplier markets (most notably auto parts). The analysis here provides at least three reasons why such penetration could be costly apart from any Japanese protectionist policies. First, multiple suppliers increase the per-unit payment to solve the agency problem. Second, a bidding system could undermine Japanese-style procurement. Third, American firms could have higher set-up/delivery costs that would make it difficult for them to participate in a JIT purchasing arrangement.

The analysis indicates in contrast that firms using American-style procurement contracts can more readily deal with multiple suppliers. Our results indicate that competitive pricing is the key advantage of the American system. McMillan (1994 pp. 213–14) corroborates this point when he writes,

> If the gains from a long-term relationship are small, as in the case of a standardized item for which the quality is easy to check and which is produced by many firms, then short-term relationships may be optimal. The costs of forgoing bidding competition can be large.

This also underscores the point that cost uncertainty, which is outside the scope of our model, creates an incentive for multiple suppliers. Hence while our formal analysis shows only an incentive for sole-sourcing under the Japanese system, it points toward the importance of numerous suppliers in American-style procurement.

The model also provides some insight regarding different levels of vertical integration in the United States and Japan. The inefficiency of the American system is implicitly that a transaction requires two inspections—a process inspection by the seller and a second inspection by the buyer. In the terms used by Ronald H. Coase (1937) there are high transaction costs to using the "market." Since inspection costs undoubtedly rise with the complexity of inputs as one moves from procuring more basic inputs to complex subassemblies and components, these rising transactions costs create an incentive for buyers using the American system to integrate upstream. Equivalently, because inspections are not performed by Japanese buyers, there is no quality control reason for producing complex intermediate inputs in house.[25] This incentive for integration is also present if buyers are using the American mode of procurement either because market conditions make it difficult to find appropriate Japanese-style suppliers or because of organizational inertia. In contrast, the relative cost of internal versus external procurement does not rise under the Japanese

[24] An alternative to multiple-sourcing used by Toyota and other Japanese companies is parallel-sourcing, under which different firms are granted renewable contracts to supply similar (but not identical) items; e.g., axil assemblies for different models of car. This practice captures many of the benefits of multiple-sourcing without incurring many of its costs (see James Richardson and James Roumasset, 1995).

[25] The argument here applies to differences in inspection costs that depend on the level in the chain of procurement. Changes in inspection costs across time will lead firms to switch procurement systems as predicted by the model, but within the American system there appears to be a greater incentive for upstream integration.

system because there is only one quality inspection under that system or internal procurement. In each of these cases the line of reasoning provided by the model offers one explanation for the higher degree of vertical integration found in the United States.

The model also sheds new light on existing empirical studies of integration. For example, it suggests that inspection costs may help explain the degree of vertical integration found in the empirical studies of the make-or-buy decision by Kirk Monteverde and David J. Teece (1982) and Scott E. Masten (1984). These authors find greater integration for complex components and ones with high levels of engineering and relate that finding to Williamsonian sunk costs. The inspection cost hypothesis advanced here can be tested directly by determining whether complex components that lead to greater integration in the United States are also likely to be produced by integrated firms in Japan, where sunk costs may be comparable but inspection costs are lower. Our analysis suggests that complex items produced in integrated organizations in the United States would be procured externally in Japan.

Finally, the analysis provides a number of empirical implications that can be used to test the theory. The theory predicts correlations between the use of inspections and lot size, between the use of inspections and the number of suppliers, between the use of inspections and supplier profit, and between average lot size and the number of suppliers. The theory also predicts an impact of changing set-up costs on the optimal procurement system, though in this case one must control for differing inspection costs. The procurement systems in turn can be measured by the commitment to repeated dealing and the use of competitive bidding, as well as lot size and inspections. Hence there are a number of direct empirical tests of the basic model.

V. Conclusion

In this paper we have presented and analyzed a formal model of two methods of industrial procurement, American-style competitive bidding and Japanese-style relational contracts. Specifically, we identified each of these supply institutions with particular ways of playing a dynamic procurement game. Although we took the methods for inducing contractual performance as given, all of the other features associated with these modes of procurement arose endogenously. For instance, equilibrium lot sizes were shown to be larger under the American-style competitive bidding system than under Japanese-style partnerships. In fact, "The hallmark of JIT purchasing is the steady purchase of parts in small lot sizes rather than in large batches as is traditional under U.S. purchasing practices." (Ahsanuddin Ansari, 1986 p. 45.)

A key finding is that these procurement arrangements represent unified systems. Our analysis shows that competitive bidding, large lot sizes and inventories, numerous suppliers, and the need for quality inspections go together. On the other hand, long-term commitments and price premiums, small lots, few suppliers, and the absence of inspections also go together. The chief disadvantage of the American system is the reliance on inspection for quality assurance, and our analysis suggests that such a system will fare worse as quality becomes a more central focus and products become more technologically complex.

Our model also provides a number of testable implications. The results predict that incentive mechanisms such as inspections and quality premia should be correlated with order size and inventories, supplier profitability, and the durability of supply relationships. Cost structures such as set-up and inspection costs should also drive differences in organizational form. There should be differences in the degree of vertical integration into the production of complex intermediate inputs across firms using different procurement systems. And finally, product variety should be greater under Japanese-style procurement.

In this paper, we have not addressed cost heterogeneity or endogenous process innovation. Cost heterogeneity obviously can create a significant incentive to use bidding to sort across suppliers. Investigation of process innovation requires a separate analysis with attention to the details of the mechanisms used to adjust prices over time wed to an inventory-theoretic quality model of the type studied here. Still, intuition suggests there might be greater incentives for process innovation under American-style procurement where the winning bidder pockets the difference between its cost and that of its closest rival, while under Japanese procurement the buyer is likely to receive more of the incremental surplus from cost reductions.

The analysis here also focuses on the remedies the buyer might undertake to assure quality and ignores possible seller remedies. In many manufacturing contexts it would be impractical for the seller to offer warranties since defects could injure the buyer's reputation. Still, the use of warranties and other mechanisms the seller might use to assure quality remains an important issue for future research.

Appendix

PROOF OF LEMMA 1:

Differentiating first with respect to α gives

$$K_{\alpha}(x, \alpha, \theta) = \frac{m}{1-\delta} - \frac{f + c(\theta)x - (1-\delta)(f + c(\underline{\theta})x)}{[1-(1-\alpha)(1-\delta)]^2}.$$

Differentiate the first term on the right with respect to x to establish

$$\frac{\partial}{\partial x}\left(\frac{m}{1-\delta}\right) = \frac{-h\delta m}{(1-\delta)^2} < 0.$$

Now differentiate the second term on the right and rearrange to get

$$-\frac{\partial}{\partial x}\frac{f + c(\theta)x - (1-\delta)(f + c(\underline{\theta})x)}{[1-(1-\alpha)(1-\delta)]^2}$$

$$= -\frac{[c(\theta) - (1-\delta)c(\underline{\theta})][1-(1-\alpha)(1-\delta)] + h\delta\{2(1-\alpha)(f + c(\theta)x) - [1 + (1-\alpha)(1-\delta)](f + c(\underline{\theta})x)\}}{[1-(1-\alpha)(1-\delta)]^3}.$$

The first term in the numerator is evidently positive. Hence, the claim will follow if it can be shown that the expression in braces is also positive. First, note that

$$2 > [1 + (1-\alpha)(1-\delta)].$$

Finally (5) and (6) are equivalent to

$$(1-\alpha)(f + c(\theta)x) \geq f + c(\underline{\theta})x.$$

LEMMA 2: *Let* (x_0, α_0) *be either an extreme point of* $K(x, \alpha, \theta)$ *or a potential solution to* (3) *in which one of the constraints binds, then*

$$\lim_{h \to 0} hx_0 = 0.$$

LEMMA 3: *Let* (x_0, α_0) *be an extreme point of* $\tilde{K}(x, \alpha, \theta)$, *then*

$$\lim_{h \to 0} h(x_0)^2 = \frac{m}{c(\theta)}$$

and

$$\lim_{h\to 0} \alpha_0 = \frac{1}{2}\left(\frac{1}{2} + \frac{c(\theta) - c(\underline{\theta})}{c(\theta)} - \frac{f}{m}\right) \equiv \hat{\alpha}_0. \tag{A1}$$

PROOF OF PROPOSITION 1:

Recall that an extreme point (x_0, α_0) is a saddlepoint of $\tilde{K}(x, \alpha, \theta)$ if cross effects outweigh direct effects. To show that this is true when h is sufficiently small, begin by writing

$$\tilde{K}_{x,x}(x_0, \alpha_0, \theta) = \frac{2(f + \alpha_0 m)}{h(x_0)^3} + 2\gamma_x(x_0, \alpha_0)[(1 - \alpha_0)c(\theta) - c(\underline{\theta})]$$

$$+ \gamma_{x,x}(x_0, \alpha_0)[(1 - \alpha_0)(f + c(\theta)x_0) - f - c(\underline{\theta})x_0].$$

It is easy to check that the second two terms go to zero in the limit as h approaches zero by Lemma 2. The first term also goes to zero because the denominator can be written as $[h(x_0)^2]x_0$. The bracketed term goes to a constant $[m/c(\theta)]$ in the limit by Lemma 3, while x_0 itself goes to infinity. Next, compute

$$\tilde{K}_{\alpha,\alpha}(x_0, \alpha_0, \theta) = \frac{4hx_0[2 + hx_0][2c(\theta)x_0 + h(x_0)^2(c(\theta) - 2c(\underline{\theta})) - hx_0 f - 2f]}{[2 + hx_0(2\alpha_0 - 1)]^3}.$$

Applying Lemma 2 and then Lemma 3 yields

$$\lim_{h\to 0} \tilde{K}_{\alpha,\alpha}(x_0, \alpha_0, \theta) = \lim_{h\to 0} 2c(\theta)h(x_0)^2 = 2m.$$

Finally,

$$\tilde{K}_{x,\alpha}(x_0, \alpha_0, \theta) = -\frac{m}{h(x_0)^2} - \gamma(x_0, \alpha_0, \theta)c(\theta) - \gamma_x(x_0, \alpha_0, \theta)(f + c(\theta)x_0).$$

Again applying Lemmas 2 and 3 yields

$$\lim_{h\to 0} \tilde{K}_{x,\alpha}(x_0, \alpha_0, \theta) = -2c(\theta).$$

Hence,

$$\lim_{h\to 0} \tilde{K}_{x,x}(x_0, \alpha_0, \theta)\tilde{K}_{\alpha,\alpha}(x_0, \alpha_0, \theta) - (\tilde{K}_{x,\alpha}(x_0, \alpha_0, \theta))^2 = 0\cdot m - 4(c(\theta))^2 < 0,$$

which establishes the claim.

PROOF OF PROPOSITION 2:

The first best is achieved under the American system when $m = 0$. Hence, set $m = 0$ and $\alpha = \bar{\alpha}(x, \theta)$ in (7), and solve for the cost-minimizing lot size.

PROOF OF PROPOSITION 3:

Applying Proposition 1, the cost-minimizing lot sizes under the Japanese and American systems are implicitly defined by

$$x_J = \sqrt{\frac{2qf}{r[c(\theta) + 2(c(\theta) - c(\underline{\theta}))(\gamma(x_J, 0) + \gamma_x(x_J, 0)x_J)]}} \tag{A2}$$

and

(A3) $$x_A = \sqrt{\frac{2q(f + \bar{\alpha}(x_A, \theta)m)}{r[c(\theta) + 2\phi(x_A)\bar{\alpha}_x(x_A, \theta)m]}}.$$

Because

$$\gamma(x_J, 0) = \frac{2 + hx_J}{2 - hx_J} > 0$$

and

$$\gamma_x(x_J, 0) = \frac{4h}{(2 - hx_J)^2} > 0,$$

it follows from direct inspection of (10) and (A2) that $x_J < x_{FB}$. Next substitute $\alpha = \bar{\alpha}(x, \theta)$ into (7) and differentiate with respect to x to get

$$\tilde{K}_x(x_A, \bar{\alpha}(x_A, \theta), \theta) = \phi'(x_A)(f + \bar{\alpha}(x_A, \theta)m) + (c(\theta)/2) + \phi(x_A)\bar{\alpha}_x(x_A, \theta)m = 0.$$

From this it is evident that $x_A > x_{FB}$ if, and only if, $m > 0$ and

$$\frac{\partial x_A}{\partial m} \equiv -\frac{\tilde{K}_{x,m}(x_A, \bar{\alpha}(x_A, \theta), \theta)}{\tilde{K}_{x,x}(x_A, \bar{\alpha}(x_A, \theta), \theta)} > 0.$$

Note that

$$\lim_{x \to 0} \tilde{K}_x(x, \bar{\alpha}(x, \theta), \theta) = -\infty$$

and

$$\lim_{x \to \infty} \tilde{K}_x(x, \bar{\alpha}(x, \theta), \theta) = c(\theta)/2.$$

Together, these imply $\tilde{K}_{x,x}(x_A, \bar{\alpha}(x_A, \theta), \theta) > 0$. Hence, it must be shown that when h is sufficiently small $\tilde{K}_{x,m}(x_A, \bar{\alpha}(x_A, \theta), \theta) < 0$. This, however, is equivalent to

$$\phi'(x_A)\bar{\alpha}(x_A, \theta) + \phi(x_A)\bar{\alpha}_x(x_A, \theta) < 0,$$

and some straightforward algebra shows this to be equivalent to $h < 2(c(\theta)/f)$.

PROOF OF PROPOSITION 5:

The same methods of proof as in Lemmas 2 and 3 can be used to establish

(A4) $$\lim_{h \to 0} \sqrt{h}x_J = \sqrt{\frac{2f}{3c(\theta) - 2c(\underline{\theta})}}$$

and

(A5) $$\lim_{h \to 0} \sqrt{h}x_A = \sqrt{\frac{2[f + (1 - c(\underline{\theta})/c(\theta))m]}{c(\theta)}}.$$

Now, note that

$$\tilde{K}(x_J, 0, \theta) < \tilde{K}(x_A, \bar{\alpha}(x_A, \theta), \theta)$$

if, and only if,

$$\sqrt{h}[\phi(x_J)f + (c(\theta)x_J/2) + (c(\theta) - c(\underline{\theta}))\gamma(x_J, 0)x_J]$$

$$< \sqrt{h}[\phi(x_A)(f + \bar{\alpha}(x_A, \theta)m) + (c(\theta)x_A/2)].$$

Expressions (A4) and (A5) indicate that this is true in the limit as h approaches zero if, and only if,

$$\sqrt{f(3c(\theta) - 2c(\underline{\theta}))} < \sqrt{fc(\theta) + m(c(\theta) - c(\underline{\theta}))}.$$

This, in turn, is equivalent to

$$2f(c(\theta) - c(\underline{\theta})) < m(c(\theta) - c(\underline{\theta})),$$

from which the result follows directly.

REFERENCES

Ansari, Ahsanuddin. "Survey Identifies Critical Factors in Successful Implementation of Just in Time Purchasing Techniques." *Industrial Engineering*, October 1986, *18*(10), pp. 44–50.

Aoki, Masahiko. "Toward an Economic Model of the Japanese Firm." *Journal of Economic Literature*, March 1990, *28*(1), pp. 1–27.

Asanuma, Banri. "Japanese Manufacturer-Supplier Relationships in International Perspective: The Automobile Case." Department of Economics Working Paper No. 8, Kyoto University, 1988.

______. "Manufacturer-Supplier Relationships in Japan and the Concept of Relation-Specific Skill." *Journal of the Japanese and International Economies*, March 1992, *3*(1), pp. 1–30.

Asoudegi, Ehsan. "Quantitative Automated Inspection of Standard Parts Using Machine Vision." *Computers and Industrial Engineering*, January 1992, *23*(1), pp. 361–64.

Athey, Susan and Schmutzler, Armin. "Product and Process Flexibility in an Innovative Environment." *Rand Journal of Economics*, Winter 1995, *26*(4), pp. 557–75.

Baumol, William J. "The Transactions Demand for Cash: An Inventory Theoretic Approach." *Quarterly Journal of Economics*, November 1952, *66*(4), pp. 545–56.

Chapman, Kenneth W.; Johnson, W. Carroll and McClean, Thomas J. "A High Speed Statistical Process Control Application ofMachine Vision to Electronics Manufacturing." *Computers and Industrial Engineering*, January 1990, *19*(1), pp. 234–38.

Coase, Ronald H. "The Nature of the Firm." *Economica*, November 1937, *4*(16), pp. 386–405.

Dyer, Jeffrey H. and Ouchi, William G. "Japanese-Style Partnerships: Giving Companies a Competitive Edge." *Sloan Management Review*, Fall 1993, *35*(1), pp. 51–63.

Eaton, B. Curtis and Schmitt, Nicolas. "Flexible Manufacturing and Market Structure." *American Economic Review*, September 1994, *84*(4), pp. 875–88.

Gilbert, Faye W.; Young, Joyce A. and O'Neal, Charles R. "Buyer-Seller Relationships in Just In Time Purchasing Environments." *Journal of Business Research*, February 1994, *29*(2), pp. 111–20.

Goldberg, Victor. "Regulation and Administered Contracts." *Bell Journal of Economics*, Autumn 1976, *7*(2), pp. 426–48.

Hald, Anders. *Statistical theory of sampling inspection by attributes*. San Francisco: Academic Press, 1981.

Holmstrom, Bengt and Milgrom, Paul. "The Firm as an Incentive System." *American Economic Review*, September 1994, *84*(4), pp. 972–91.

Johnston, Russell and Lawrence, Paul R. "Beyond Vertical Integration: The Rise of the Value Adding Partnership." *Harvard Business Review*, July–August 1988, *66*(4), pp. 94–101.

Judd, Kenneth. *Numerical methods in economics*. Cambridge, MA: MIT Press, 1997.

Kawasaki, Seiichi and McMillan, John. "The Design of Contracts: Evidence from Japanese Subcontracting." *Journal of the Japanese and International Economies*, September 1987, *1*(3), pp. 327–49.

Klein, Benjamin and Leffler, Keith B. "The Role of Market Forces in Assuring Contractual Performance." *Journal of Political Economy*, August 1981, *89*(4), pp. 615–41.

Masten, Scott E. "The Organization of Production: Evidence from the Aerospace Industry." *Journal of Law and Economics*, October 1984, *27*(2), pp. 403–17.

McMillan, John. "Reorganizing Vertical Supply Relationships," in Horst Siebert, ed., *Trends in business organization: Do participation and cooperation increase competitiveness*? Tübingen, Germany: Mohr, 1994, pp. 203–22.

Milgrom, Paul and Roberts, John. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *American Economic Review*, June 1990, *80*(3), pp. 511–28.

______. "Complementarities and Fit: Strategy, Structure, and Organizational Change in Manufacturing." *Journal of Accounting and Economics*, March–May 1995, *19*(2), pp. 179–208.

Milgrom, Paul and Shannon, Christina. "Monotone Comparative Statics." *Econometrica*, January 1994, *62*(1), pp. 157–80.

Monteverde, Kirk and Teece, David J. "Supplier Switching Costs and Vertical Integration in the Automobile Industry." *Bell Journal of Economics*, Spring 1982, *13*(1), pp. 206–13.

Richardson, James and Roumasset, James. "Sole Sourcing, Competitive Sourcing, Parallel Sourcing: Mechanisms for Supplier Performance." *Managerial and Decision Economics*, January–February 1995, *16*(1), pp. 71–84.

Riordan, Michael H. and Sappington, David E. M. "Second Sourcing." *Rand Journal of Economics*, Spring 1989, *20*(1), pp. 41–58.

Rogerson, William P. "Profit Regulation of Defense Contractors and Prizes for Innovation." *Journal of Political Economy*, December 1989, *97*(6), pp. 1284–1305.

Sako, Mari. *Prices, quality, and trust: Interfirm relations in Britain and Japan*. Cambridge: Cambridge University Press, 1992.

Schilling, Edward G. *Acceptance sampling and quality control*. New York: Dekker, 1982.

Schonberger, Richard. *Japanese manufacturing techniques*. New York: Free Press, 1982.

Shapiro, Carl. "Consumer Information, Product Quality, and Seller Reputation." *Bell Journal of Economics*, Spring 1982, *13*(1), pp. 20–35.

______. "Premiums for High Quality Products as Returns to Reputation." *Quarterly Journal of Economics*, November 1983, *98*(4), pp. 659–79.

Taylor, Curtis R. and Wiggins, Steven N. "Japanese-Style Partnerships or Business as Usual: Two Ways of Playing the Subcontracting/Procurement Game." Private Enterprise Research Center Working Paper No. 9415, Texas A&M University, July 1994.

Tobin, James. "The Interest-Elasticity of Transactions Demand for Cash." *Review of Economics and Statistics*, August 1956, *38*(3), pp. 241–47.

Topkis, Donald M. "Minimizing a Submodular Function on a Lattice." *Operations Research*, April–March 1978, *26*(2), pp. 305–21.

Ventura, Jose A.; Chang, Alec C. and Klein, Cerry M. "Automated Inspection of Circular Parts." *Computers and Industrial Engineering*, January 1988, *15*(1), pp. 349–54.

Ventura, Jose A. and Yeralan, Sencer. "Optimization Techniques for Automated Inspection of Rectangular and Square Parts." *Computers and Industrial Engineering*, January 1990, *19*(1), pp. 239–43.

Wald, Abraham. *Sequential analysis*. New York: Wiley, 1947.

Williamson, Oliver E. "Transaction Cost Economics: The Governance of Contractual Relations." *Journal of Law and Economics*, October 1979, *22*(2), pp. 233–61.

______. *The economic institutions of capitalism*. New York: Free Press, 1985.