A COOPERATIVE GAME THEORY MODEL OF QUANTITY DISCOUNTS*

RAJEEV KOHLI AND HEUNGSOO PARK

Joseph M. Katz Graduate School of Business, University of Pittsburgh,
Pittsburgh, Pennsylvania 15260

College of Business and Economics, Yonsei University, Seoul, Korea

Quantity discounts offered by a monopolist are considered in the context of a bargaining problem in which the buyer and the seller negotiate over the order quantity and the average unit price. All-units and incremental quantity discounts that permit transaction at a negotiated outcome are described. The effects of risk sensitivity and bargaining power on quantity discounts are discussed for alternative bargaining models.

(MARKETING—DISTRIBUTION; MARKETING—PRICING; BARGAINING; INVENTORY/PRODUCTION—EQQ)

1. Introduction

Quantity discounts are often used in pricing policies. Rachman (1975) and Howell, Kuzdrall, Britney and Wilcox (1986) note their widespread use and their importance to marketing strategy. Rao (1980) suggests that quantity discounts should take into account inflation, mergers among buyers, and increased sophistication in buyers' ordering policies. Heirritz and Farrel (1971) examine quantity discounts from the purchaser's viewpoint, urging buyers to take advantage of discounts because the price differentials can be significant.

Researchers in marketing and economics suggest three main reasons for quantity discounts: (i) perfect discrimination against a single buyer or a homogeneous group of buyers (Buchanan 1953, Gabor 1965, Moorothy 1984), (ii) partial discrimination against a heterogeneous group of buyers (Oi 1971, Leland and Meyer 1973, Faulhaber and Panzar 1977, Murphy 1977, Spence 1977, Oren, Smith and Wilson 1982, 1983), and (iii) improved efficiency of transactions between a seller and a buyer (Crowther 1964, Foraker 1961, Dolan 1978, Lal and Staelin 1984, Monahan 1984, Lee and Rosenblatt 1986, Dada and Srikanth 1987). Dolan (1987) provides an excellent review of the literature dealing with each motivation for quantity discounts. Here, we briefly discuss the transactions-efficiency rationale, which is of direct relevance to this paper.

The relationship between quantity discounts and the efficiency of buyer-seller transactions was first discussed by Buchanan (1953) in economics, and by Foraker (1961) and Crowther (1964) in marketing. These researchers note that while the inventory holding cost jointly incurred by the buyer and the seller can sometimes increase when the buyer orders more than his economic order quantity, the ordering and order-processing costs can simultaneously decrease. Therefore under certain conditions, larger purchases by the buyer than his economic order quantity can reduce the total transaction cost for the buyer-seller system. To provide an incentive for making these larger purchases, the seller can pass on a part of the transaction-cost savings to the buyer via a quantity discount.

Dolan (1978) formalizes these ideas in a model in which incremental discounts are offered by a single seller to a single buyer (or a homogeneous group of buyers) with fixed demand. Building on Dolan's work, Lal and Staelin (1984) describe a model to identify

* Accepted by Jehoshua Eliashberg; received March 26, 1987. This paper has been with the authors 3 months for 2 revisions.
"optimal" incremental quantity discounts for a seller offering a product to one group of homogeneous buyers or to heterogeneous groups of buyers. The discount schedules identified are "optimal" in the sense that they maximize the seller's profit without increasing any buyer's cost. Dada and Srikanth (1987) generalize these results by relaxing a number of conditions in Lal and Staelin's model resulting from the assumption that quantity discounts lie on an average cost curve with a specific functional form. Dada and Srikanth show that the only necessary condition for quantity discounts to increase transactions efficiency is that the inventory holding cost be higher for the buyer than for the seller.

This paper extends previous work on the transactions-efficiency motivation for quantity discounts in two ways. First, the focus of previous research is on maximizing profit for a seller. As Lal and Staelin observe, the normative results of this approach are appropriate if the process underlying discounting is that the seller presents a discount schedule to the buyer, who then decides whether or not to avail of a quantity discount. However, this is only one possible description of the process by which quantity discounts are obtained by a buyer. Buffa (1984) and Banerjee (1986) observe that questions of pricing and lot sizing are traditionally settled through negotiations between the buyer and the seller. This is so particularly when the seller supplies a single (or one major) buyer, when buyers negotiate prices through a cooperative, or when buying policies are negotiated on a customer-by-customer basis. What will be the outcome of such negotiation? Will the resulting quantity discount necessarily maximize the joint cost saving for the buyer and the seller? To examine these issues, we analyze quantity discount in the context of a two-party bargaining problem. Specifically, we consider bargaining models due to Nash (1950), Kalai and Smorodinsky (1975) and Eliahsberg (1986), which have received significant attention in the bargaining literature. Because each model uses Pareto efficiency either as an axiom or as a prediction of the bargaining solution (Roth's 1979 re-characterization of Nash's model), the set of Pareto efficient discounts for a monopolistic seller transacting with a single buyer (or a homogeneous group of buyers) are first identified. These discounts are shown to correspond exactly to the set of discounts that maximize the joint cost saving (efficiency gain) between the seller and the buyer(s) (Lal and Staelin 1984, Dada and Srikanth 1987). The Nash and Kalai and Smorodinsky bargaining solutions are described and illustrated, the effect of the buyer's (seller's) risk sensitivity on the bargaining solution is discussed, and the role of bargaining power on the negotiated outcome is illustrated in the context of Eliahsberg's model.

Second, previous research has restricted itself to incremental quantity discounts. For many products (e.g., computers, computer diskettes, stud bolts, envelopes, cassette tapes, printing materials), all-units quantity discounts are more commonly offered (Howell, Kuzdral, Britney and Wilcox 1986). Is there a reason why the type of discount offered differs among products? Should sellers (buyers) of certain products prefer one type of discount to the other? The present analysis shows that there is no efficiency rationale for selecting an incremental or all-units discounts, and no bargaining rationale that makes one type of discount better than the other for either a buyer or a seller. Put another way, from a transactions-efficiency perspective, the choice between incremental and all-units discounts is a matter of firm or industry practice, not a result of their desirability for the buyer, the seller, or the buyer-seller system.

2. The Model

Three assumptions made by Dolan (1978), Lal and Staelin (1984), and Dada and Srikanth (1987) are retained in the analysis. The first assumption is that the inventory policies of seller(s) and buyer(s) can be described by a simple EOQ model based on

---

1 A definition of each type of discount is provided in § 2.
deterministic demand, no stock outs and deterministic lead times. As in the preceding models, the principal reason for this assumption is that it simplifies analysis, is used by many firms for determining their inventory levels, and provides a good approximation to more complex inventory policies.

The second assumption is that the buyer’s annual demand does not increase in response to a quantity discount. This implies that the analysis is most appropriate when the seller’s product is not a major component in the buyer’s final product, when the buyer is the ultimate consumer and has a fixed requirement for the product, or when the buyer and the seller are independent divisions of a parent company and one division supplies components used in a process or product by the other division. As Lal and Staelin observe, assuming fixed demand has two important consequences. First, the seller’s optimal production schedule, which is based on demand and production setup costs, is not altered if the buyer changes his ordering policy. Second, it is possible to restrict attention to a single period model.

The third assumption is that both buyer(s) and seller(s) know their own and each other’s holding and ordering costs. This is a simplification of reality, but a necessary first step for understanding motivations for quantity discounts. Assessing the situation of complete information should suggest how to subsequently analyze the more realistic, but also more difficult, situation in which sellers and buyers have incomplete information about their own and each other’s inventory cost parameters.

We begin by defining incremental and all-units quantity discounts. An incremental quantity discount has the following structure: the seller charges price \( P_0 \) for the first \( Q_1 \) units, price \( P_1 \) for any additional units between \( Q_1 \) and \( Q_2 \), price \( P_2 \) for any additional units between \( Q_2 \) and \( Q_3 \), units, and so on down to \( P_m \). In contrast, an all-units quantity discount has the following structure: the seller charges unit price \( P_0 \) up to \( Q_1 \) units, price \( P_1 \) for all units ordered provided the order is greater than or equal to \( Q_1 \) but less than \( Q_2 \), price \( P_2 \) for all units ordered provided the order is greater than or equal to \( Q_2 \) but less than \( Q_3 \), and so on down to \( P_m \), where \( P_0 > P_1 > \cdots > P_m \) (Johnson and Montgomery 1974).

Consider all possible incremental and all-units quantity discounts for which an order quantity of size \( Q_x \) has a discounted average unit price \( P_x \) (i.e., \( P_x \) is the revenue earned by the seller from an order of \( Q_x \) units, divided by the number of units \( Q_x \)). The following analysis refers to \( x = (Q_x, P_x) \) as discount \( x \), with the implicit understanding that it corresponds to not one but all quantity discounts for which the average unit price of \( Q_x \) units is \( P_x \).

Assume that no discount is initially offered. Let “\( D \)” denote the buyer’s annual demand for the product.\(^2\) Let “\( P_0 \)” denote the undiscounted unit price (= average unit price) of the product. Let “\( H \)” denote the buyer’s inventory holding cost per unit, and let “\( A \)” denote the buyer’s cost of placing an order. Then the buyer’s annual cost from ordering lots of \( Q_0 \) (\( \leq D \)) units is (Lal and Staelin 1984, p. 1529)

\[
C(0) = DP_0 + A(D/Q_0) + H(Q_0/2)
\]

where \( DP_0 \) is the purchase price of \( D \) units of the product; \( H(Q_0/2) \) is the cost of holding an average inventory of \( (Q_0/2) \) units throughout the year; and \( A(D/Q_0) \) is the cost of placing \( (D/Q_0) \) orders in a year.\(^3\)

\(^2\) The model is developed for a single buyer. Like Lal and Staelin’s (1984) and Dada and Srikant’s (1987) models, it is equally applicable to a homogeneous group of buyers.

\(^3\) Like Dolan and Lal and Staelin, we do not require the number of orders to be an integer. This assumption is made to simplify the analysis, and is standard in the literature on inventory models (e.g., Johnson and Montgomery 1974).
Because \( C(0) \) is a convex function of \( Q_0 \), the buyer minimizes total cost by purchasing lots of size

\[
Q_0 = (2AD/H)^{1/2}
\]

which is the expression for the economic order quantity for a buyer with fixed demand and constant inventory holding and ordering costs (Johnson and Montgomery 1974, p. 32).

Assume that the seller offers quantity discount \( x = (Q_x, P_x) \) so that the average unit price for an order of \( Q_x \) units is \( P_x \). Then the annual cost to the buyer is given by the following expression analogous to equation (1):

\[
C(x) = DP_x + A(D/Q_x) + H(Q_x/2).
\]

Subtracting equation (3) from equation (1) and setting \( C(0) - C(x) = \Delta C(x) \) yields

\[
P_x = P_0 - \Delta C(x)/D - A(1/Q_x - 1/Q_0) - (H/2D)(Q_x - Q_0).
\]

In equation (4), \( \Delta C(x) \) is the decrease \((\Delta C(x) > 0)\) or increase \((\Delta C(x) < 0)\) in the buyer’s annual cost resulting from quantity discount \( x \). For example, let \( D = 1000 \) units, \( Q_0 = 100 \) units, \( P_0 = $2/\text{unit} \), \( H = $3/\text{unit} \) and \( A = $15/\text{order} \). Then \( C(0) = $2000 + $150 + $150 = $2300 \). If the seller offers an all-units quantity discount, reducing the price to \$1.60/unit for lots of 300 or more units, the buyer’s annual cost from ordering lots of 300 units is \( C(x) = $1600 + $50 + $450 = $2100 \). Thus, the buyer saves \$200 by changing his ordering policy to 300 units (i.e., \( \Delta C(x) = $2300 - $2100 = $200 \) at \( x = (300, 1.60) \)).

Equation (4) describes \( P_x \) as a concave function of \( Q_x \) (i.e., \( d^2P_x/dQ_x^2 = -2A/Q_x^3 < 0 \)). Given a value of \( \Delta C(x) \), equation (4) specifies an iso-cost curve in the \((Q, P)\) space (Figure 1). All points on this curve correspond to combinations of lot size orders \( Q_x \) and average unit prices \( P_x \) for which the buyer incurs an annual cost of \( C(x) = C(0) - \Delta C(x) \). When no confusion results, we will say that discount \( x \) lies on the iso-cost curve \( C(x) \).

Referring to Figure 1, when \( \Delta C(x) = 0 \), the iso-cost curve \( C(0) \) passes through \((Q_0, P_0)\), which corresponds to the buyer purchasing \( Q_0 \) units at an undiscounted unit price \( P_0 \). For all values of \( \Delta C(x) > 0 \ (< 0) \), the buyer is on an iso-cost curve that is closer to (further away from) the origin in Figure 1. Therefore the buyer:

![Figure 1. Feasible and Undominated Quantity Discounts: Single Seller-Single Buyer Case.](image)
1. is indifferent between purchasing $Q_0$ units at an undiscounted unit price $P_0$ and availing of discounts which lie on the iso-cost curve $C(0)$;
2. prefers all discounts which lie in the region below the iso-cost curve $C(0)$ to no discount; and
3. prefers purchasing $Q_0$ units at an undiscounted unit price $P_0$ to all discounts which lie in the region above the iso-cost curve $C(0)$.

Turning to the seller, let the cost of processing an order be denoted "$a$", and let the cost of capital for holding one unit in inventory for a year be denoted "$h$". If the buyer purchases $Q_0$ units at an undiscounted price $P_0$, the seller incurs an annual inventory cost of (Lal and Staelin 1984, p. 1527)

$$a(D/Q_0) - h(Q_0/2)$$

where $a(D/Q_0)$ is the cost of processing the $(D/Q_0)$ orders placed by the buyer, and $h(Q_0/2)$ is the decrease in the seller's cost of capital due to the buyer purchasing the product from the seller. Therefore, the seller's annual profit is

$$\Pi(0) = D(P_0 - v) - a(D/Q_0) + h(Q_0/2) - FC$$

where $v$ is the unit variable cost of the product (assumed to be constant), $D(P_0 - v)$ is the seller's margin on the sale of $D$ units at an undiscounted unit price $P_0$, and $FC$ is the seller's annual fixed cost.

Let $\Pi(x)$ denote the annual profit to the seller when the buyer avails of an offered discount $x = (Q_x, P_x)$. Then

$$\Pi(x) = D(P_x - v) - a(D/Q_x) + h(Q_x/2) - FC.$$  

(7)

Subtracting equation (6) from equation (7) yields

$$\Pi(x) = DP_x - DP_0 - aD(1/Q_x - 1/Q_0) + (h/2)(Q_x - Q_0) + \Pi(0).$$

(8)

Let $\Pi(x) = \Pi(0) + \Delta \Pi(x)$, where $\Delta \Pi(x)$ is the increase ($\Delta \Pi(x) > 0$) or decrease ($\Delta \Pi(x) < 0$) in the seller's profit resulting from quantity discount $x$. Then the seller offers the discount only if $\Pi(x) > \Pi(0)$; i.e., if $\Delta \Pi(x) > 0$. For example, let $D = 1000$ units, $P_0 = $2/unit, $h = $1/unit and $a = 75/unit. Let the seller offer an all-units quantity discount, reducing the price to $1.60/unit for lots of 300 or more units. Assume that the buyer changes his ordering policies to 300 units. Then the seller's profit is $\Pi(0) = $2000 - $750 + $50 = $1300 before the quantity discount, and is $\Pi(x) = $1600 - $250 + $150 = $1500 after the quantity discount. Thus $\Delta \Pi = $200 at $x = (300, 1.60)$.

Substitute $\Pi(x) = \Pi(0) + \Delta \Pi(x)$ in equation (8) and re-arrange terms to obtain the following expression for the discounted average unit price $P_x$:

$$P_x = P_0 + \Delta \Pi(x)/D + a(1/Q_x - 1/Q_0) - (h/2D)(Q_x - Q_0).$$

(9)

Equation (9) describes $P_x$ as a convex function of $Q_x$ (i.e., $d^2P_x/dQ_x^2 = 2a/Q_x^3 > 0$). Given a value of $\Delta \Pi(x)$, equation (9) specifies an iso-profit curve corresponding to profit $\Pi(x) = \Pi(0) + \Delta \Pi(x)$ in the $(Q, P)$ space (Figure 1). All points on this curve correspond to combinations of lot size orders $Q_x$ and discounted average unit prices $P_x$, for which the seller obtains an annual profit of $\Pi(x)$. When no confusion results, we will say that discount $x$ lies on the iso-profit curve $\Pi(x)$.

For $\Delta \Pi(x) = 0$, the iso-profit curve $\Pi(0)$ passes through $(Q_0, P_0)$, which corresponds to selling $Q_0$ units at an undiscounted unit price $P_0$. For all values of $\Delta \Pi(x) > 0 (<0)$, the seller is on an iso-profit curve further away from (closer to) the origin in Figure 1. Therefore the seller:

1. is indifferent between selling $Q_0$ units at an undiscounted unit price $P_0$ and offering any discount which lies on the iso-profit curve $\Pi(0)$;
2. prefers all discounts which lie in the region above the iso-profit curve \( \Pi(0) \) to no discount; and

3. prefers selling \( Q_0 \) units at an undiscounted unit price \( P_0 \) to all discounts which lie in the region below the iso-profit curve \( \Pi(0) \).

Both the concave iso-cost curve, \( C(0) \), and the convex iso-profit curve, \( \Pi(0) \), pass through \((Q_0, P_0)\). These two curves cannot be tangent to each other at \((Q_0, P_0)\), because the slope \( (A/Q_0^2 - H/(2D)) \) of the iso-cost curve \( C(0) \) is zero at \( Q_0 = (2AD/H)^{1/2} \), and the slope \( -(a/Q_0^2 - h/(2D)) \) of the iso-profit curve \( \Pi(0) \) is strictly negative at \( Q_0 = (2AD/H)^{1/2} \). Therefore, the two curves intersect at \((Q_0, P_0)\), and there always exist values of \( Q \) and \( P \) for which \( \Delta \Pi(x) \geq 0 \) and \( \Delta C(x) \geq 0 \) simultaneously. Referring to Figure 1:

1. \( \Delta \Pi(x) = \Delta C(x) = 0 \) at points \( A \) and \( C \) at which the iso-profit curve \( \Pi(0) \) intersects the iso-cost curve \( C(0) \);
2. \( \Delta \Pi(x) = 0 \) and \( \Delta C(x) > 0 \) for all points (excluding \( A \) and \( C \)) on the arc \( A-B-C \) of the iso-profit curve \( \Pi(0) \);
3. \( \Delta \Pi(x) > 0 \) and \( \Delta C(x) = 0 \) for all points (excluding \( A \) and \( C \)) on the arc \( A-D-C \) of the iso-cost curve \( C(0) \); and
4. \( \Delta \Pi(x) > 0 \) and \( \Delta C(x) > 0 \) for all points which lie entirely within the bounded region \( A-B-C-D-A \).

A discount for which the conditions \( \Delta \Pi(x) \geq 0 \) and \( \Delta C(x) \geq 0 \) are simultaneously satisfied (i.e., which lie on or within the boundary \( A-B-C-D-A \) in Figure 1) will be called feasible, and the set

\[
X = \{(Q_x, P_x) | \Delta \Pi(x) \geq 0, \Delta C(x) \geq 0, Q_x \neq Q_0, P_x \neq P_0 \}
\]

will be called the feasible set of discounts. Note that in equation \( (10) \), \( Q_x \neq Q_0 \) and \( P_x \neq P_0 \) because \((Q_0, P_0)\) is not a discount (i.e., because \((Q_0, P_0)\) refers to the undiscounted transaction of \( Q_0 \) units at price \( P_0 \)).

Assume that the seller and the buyer engage in bargaining for a discount in \( X \). Which discount should one observe as an outcome of bargaining? We consider bargaining solutions predicted by Nash’s (1959), Kalai and Smorodinsky’s (1975) and Elishberg’s (1986) models. Because all three models assume Pareto efficiency as a condition of the bargaining solution, we first identify Pareto efficient discounts among the feasible quantity discounts in \( X \), then discuss the solutions predicted by these models.

Let \( x \) and \( y \) denote any two discounts in \( X \). Then \( x \) dominates \( (Pareto \ dominates) \ y \) if either

\[
C(x) < C(y) \quad \text{and} \quad \Pi(x) \geq \Pi(y) \quad \text{or} \quad (11)
\]

\[
C(x) \leq C(y) \quad \text{and} \quad \Pi(x) > \Pi(y). \quad (12)
\]

That is, \( x \) dominates \( y \) if either (a) the buyer incurs a lower cost and the seller earns no less profit at \( x \) than at \( y \), or (b) the buyer incurs no more cost and the seller earns more profit at \( x \) than at \( y \).

Let \( y \in X \) be a discount for which neither condition \( (11) \) nor condition \( (12) \) is satisfied. Then \( y \) is Pareto efficient, because there is no other quantity discount in \( X \) at which the buyer incurs a lower cost and the seller makes no less profit, or the buyer incurs no more cost and the seller makes more profit.

**Proposition 1.** Let \( Y \) denote the subset of discounts in \( X \) which lie on the locus of tangency between the seller’s iso-profit curves and the buyer’s iso-cost curves. Then \( Y \) describes the subset of Pareto efficient discounts in \( X \).

\[4\] Roth (1979) reformulates Nash’s model by replacing Pareto optimality with individual rationality as an axiom. Under the revised set of axioms, he shows that Pareto optimality is a prediction of Nash’s bargaining solution.
Proposition 1 says that Pareto efficient discounts exist, and that they are described by the locus of tangency between the seller’s iso-profit curves and the buyer’s iso-cost curves in the region \(A-B-C-D-A\) in Figure 1. For ease of exposition, the proofs of this and all subsequent propositions are presented in the appendix.

Pareto efficient discounts have three important properties. First, all of them are associated with a single lot size order and differ only in terms of their discounted average unit prices. Referring to Figure 1, Pareto efficient discounts are described by the vertical line segment \(B-D\) at \(Q^*\). Second, all Pareto efficient discounts, but none other, maximize the sum of the seller’s profit increase \((\Delta \Pi(x))\) and the buyer’s cost saving \((\Delta C(x))\). We call the maximum value of \(\Delta \Pi(x) + \Delta C(x)\) the efficiency gain \((G)\), because it is the joint economic gain attained by the buyer and the seller by moving from a Pareto dominated, undispersed pricing policy to a Pareto efficient, discounted pricing policy. Third, every possible partitioning of the efficiency gain negotiated by the seller and the buyer is achievable via some Pareto efficient discount. Propositions 2–4 describe these results more specifically.

**Proposition 2.** The collection of Pareto efficient discounts is described by the set

\[
Y = \{(Q^*, P_y) | P_{\text{min}} \leq P_y \leq P_{\text{max}}\}
\]

where

\[
Q^* = [2(A + a)D/(H-h)]^{1/2},
\]

\[
P_{\text{max}} = P_0 - A(1/Q^* - 1/Q_0) - (H/2D)(Q^* - Q_0)
\]

and

\[
P_{\text{min}} = P_0 + a(1/Q^* - 1/Q_0) - (h/2D)(Q^* - Q_0).
\]

Proposition 2 says that all Pareto efficient discounts are associated with a single lot size \(Q^*\) and with average unit prices that lie between \(P_{\text{max}}\) and \(P_{\text{min}}\). Referring to Figure 1, \(P_{\text{max}}\) (\(P_{\text{min}}\)) is the average unit price at which an iso-profit (iso-cost) curve is tangent to the iso-cost (iso-profit) curve \(C(0) (\Pi(0))\). The locus of tangency lies on a vertical line segment at \(Q^*\) because the expressions for both iso-cost and iso-profit curves contain only linear price terms, so that vertically shifting an iso-cost (iso-profit) curve yields another iso-cost (iso-profit) curve. As the tangent point also shifts vertically when a pair of (tangent) iso-profit and iso-cost curves shift vertically, the locus of tangency traces a vertical line segment in the present case. Note that it is the special shape of iso-cost and iso-profit curves that is responsible for this result in the present instance, which is not generally witnessed in such Edgeworth-box type analyses.

**Proposition 3.**

\[
G = \max \{ \Delta \Pi(x) + \Delta C(x) | x \in X\} = D(P_{\text{max}} - P_{\text{min}}) = \Delta \Pi(y) + \Delta C(y)
\]

for all \(y \in Y\).

Proposition 3 says that, among all discounts in \(X\), the sum of the decrease in the buyer’s cost and the increase in the seller’s profit is maximized (i.e., the efficiency gain is attained) for every Pareto efficient quantity discount, but not for any other quantity discount. Note that this proposition implies that the difference between the seller’s profit \((\Pi(y) = \Pi(0) + \Delta \Pi(y))\) and the buyer’s cost \((C(y) = C(0) - \Delta C(y))\) is maximized only if \(y\) is a Pareto efficient discount. Equivalently, the set of Pareto efficient discounts corresponds to exactly the set of discounts that, in Dada and Srikant’s (1987) and Lal and Staelin’s (1984) analysis, maximize the joint cost saving for the buyer-seller system.

Proposition 3 suggests an intuitive explanation for why the joint cost saving is maximized at a single lot size \(Q^*\). Given any discount \(x \in X\), the joint cost saving to the buyer and the seller is proportional to the height of vertical line segment that passes through \(x\), touching the iso-cost curve \(C(0)\) at its upper end, and touching the iso-profit curve \(\Pi(0)\) at its lower end (Figure 1). The length of this line segment is greatest at \(Q^*\).
which therefore is the quantity that maximizes the joint cost saving for the buyer and the seller.

**Proposition 4.** Any Pareto efficient transaction \((Q^*, P^*)\), \(P^* \in [P_{\min}, P_{\max}]\), provides a cost saving of

\[
\Delta C(y) = D(P_{\max} - P^*) \tag{18}
\]

to the buyer, and a profit increase of

\[
\Delta \Pi(y) = D(P^* - P_{\min}) \tag{19}
\]

to the seller.

Proposition 4 says that at any \(P^* \in [P_{\min}, P_{\max}]\), the buyer’s share of the efficiency gain is proportional to \((P_{\max} - P^*)\), and the seller’s share of the efficiency gain is proportional to \((P^* - P_{\min})\). Hence if \(P^* = P_{\max}\), the entire efficiency gain is reflected as an increase in the seller’s profit, and if \(P^* = P_{\min}\), the entire efficiency gain is reflected as a reduction in the buyer’s cost.

What do the preceding results imply about bargaining between the buyer and the seller over the set of feasible discount? The property that all Pareto efficient transactions occur at \(Q^*\) implies that, according to both Nash’s and Kalai and Smordinsky’s models, the buyer and the seller will agree to change the lot size order to \(Q^*\) from \(Q_0\). However, they will negotiate over discounts with average unit prices that range between \(P_{\max}\) and \(P_{\min}\) at \(Q^*\), a price closer to \(P_{\max}\) being preferred by the seller, and a price closer to \(P_{\min}\) being preferred by the buyer. Nash’s bargaining model predicts that the average unit price upon which the seller and the buyer agree maximizes the product of individual marginal utilities over the disagreement outcome of no discount. Kalai and Smordinsky’s model predicts that the negotiated average unit price corresponds to the point at which the Pareto frontier in the utility space intersects the line connecting the disagreement point (i.e., the point associated with \((Q_0, F_0)\)) with the “ideal-point” (i.e., the point corresponding to the infeasible outcome at which both the seller and the buyer obtain the entire efficiency gain). We illustrate the bargaining solutions below using three examples, then investigate the effect of negotiating power of agents by examining the bargaining solution for a model due to Elishberg (1986).

**Example 1.** Consider a buyer and a seller who are both risk neutral; i.e., the utility function for the buyer and the seller are

\[
u_b(x) = \Delta C(x) \quad \text{and} \quad \tag{20} \\
u_s(x) = \Delta \Pi(x), \quad \tag{21}
\]

respectively. Both the buyer and the seller obtain a maximum utility of \(D(P_{\max} - P_{\min})\), the buyer when the bargaining outcome is \((Q^*, P_{\min})\), and the seller when the bargaining outcome is \((Q^*, P_{\max})\). The Pareto frontier is described by the equation

\[
u_b = D(P_{\max} - P_{\min}) - u_s(x). \tag{22}
\]

Each point on the (linear) Pareto frontier corresponds to a Pareto efficient discount. Nash’s model predicts a bargaining solution at a point on the Pareto frontier that maximizes

\[
u_b(x)u_s(x) = \Delta C(x)\Delta \Pi(x) = D^2(P_{\max} - P^*)(P^* - P_{\min}), \tag{23}
\]

where \(P^*\) is the average unit price of the negotiated discount at \(Q^*\). Expression (23) has a maximum at

\[
P^* = (P_{\max} + P_{\min})/2, \tag{24}
\]
the mid point between $P_{\text{max}}$ and $P_{\text{min}}$. Thus, when both the seller and the buyer are risk neutral, Nash’s model predicts that the quantity discount upon which the buyer and the seller agree will be \((Q^*, (P_{\text{max}} + P_{\text{min}})/2)\), and that they will equally split the efficiency gain, i.e.,

\[
\Delta C(Q^*, P^*) = \Delta \Pi(Q^*, P^*) = D(P_{\text{max}} - P_{\text{min}})/2
\]

at \(P^* = (P_{\text{max}} + P_{\text{min}})/2\). \(25\)

The Kalai and Smordinsky model also predicts that a risk-neutral buyer and a risk-neutral seller agree upon a discount at which they equally split the efficiency gain. To see this, observe that the line joining the ideal point \((D(P_{\text{max}} - P_{\text{min}}), D(P_{\text{max}} - P_{\text{min}}))\) to the disagreement point \((0, 0)\) intersect the Pareto frontier at the point \((D(P_{\text{max}} - P_{\text{min}})/2, D(P_{\text{max}} - P_{\text{min}})/2)\). Thus when both the buyer and the seller are risk neutral, both models predict an equal split of the efficiency gain between the buyer and the seller.

**Example 2.** Now consider a risk-neutral buyer and a risk-averse seller with utility functions

\[
u_b(x) = \Delta C(x) \quad \text{and} \quad \nu_s(x) = \Delta \Pi(x)^{1/2},
\]

respectively. The buyer obtains a maximum utility of \(D(P_{\text{max}} - P_{\text{min}})\) when the bargaining outcome is the quantity discount \((Q^*, P_{\text{min}})\). The seller obtains a maximum utility \(\{D(P_{\text{max}} - P_{\text{min}})\}^{1/2}\) when the bargaining outcome is the quantity discount \((Q^*, P_{\text{max}})\). The Pareto frontier is described by the equation

\[
u_b(x) = D(P_{\text{max}} - P_{\text{min}}) - \nu_s(x)^2,
\]

a transfer of \(g\) dollars of the efficiency gain from the buyer to the seller resulting in a decrease of \(g\) units in the buyer’s utility and an increase of \(g^{1/2}\) units in the seller’s utility.

Nash’s model predicts that the buyer and the seller will agree at a quantity discount with an average unit price \(P^*\) that maximizes

\[
D(P_{\text{max}} - P^*) \{D(P^* - P_{\text{min}})\}^{1/2}.
\]

Expression (29) has a maxima at

\[
P^* = (P_{\text{max}} + 2P_{\text{min}})/3
\]

which is below the mid point between \(P_{\text{max}}\) and \(P_{\text{min}}\) (Figure 1). Equivalently, the risk-neutral buyer obtains two thirds of the efficiency gain, and the risk-averse seller obtains one third of the efficiency gain; i.e.,

\[
\Delta C(Q^*, (P_{\text{max}} + 2P_{\text{min}})/3) = 2D(P_{\text{max}} - P_{\text{min}})/3
\]

and

\[
\Delta \Pi(Q^*, (P_{\text{max}} + 2P_{\text{min}})/3) = D(P_{\text{max}} - P_{\text{min}})/3.
\]

The Kalai and Smordinsky model predicts that the buyer and the seller will agree at a discount at \(Q^*\) with average unit price \(P^*\) that lies at the intersection of the Pareto frontier and the straight line joining the ideal point \((D(P_{\text{max}} - P_{\text{min}}), \{D(P_{\text{max}} - P_{\text{min}})\}^{1/2})\) with the disagreement point \((0, 0)\). That is, it predicts a bargaining outcome at a point where the Pareto frontier (28) intersects the line

\[
u_b(x) = \{D(P_{\text{max}} - P_{\text{min}})\}^{1/2} \nu_s(x),
\]

which occurs at

\[(u_b(x), u_s(x)) = (0.618D(P_{\text{max}} - P_{\text{min}}), 0.618 \{D(P_{\text{max}} - P_{\text{min}})\}^{1/2}).\]

Because \(u_b(x) = \Delta C(x)\) and \(u_s(x) = \Delta \Pi(x)^{1/2}\), the buyer obtains 61.8% of the efficiency
gain, and the seller obtains 38.2% of the efficiency gain from the quantity discount on which the buyer and the seller agree. Equivalently, the average unit price of the quantity discount at which they agree is

$$P^* = 0.384 P_{\text{max}} + 0.618 P_{\text{min}},$$

(35)

which is below the midpoint between $P_{\text{max}}$ and $P_{\text{min}}$ (Figure 1). Thus, the Kalai and Smorodinsky model also predicts a larger share of the efficiency gain for a risk-neutral buyer than for a risk-averse seller.

**Example 3.** Finally, consider a buyer and a seller who are both risk averse, and whose utility functions are

$$u_b(x) = \Delta C(x)^{1/2}$$

and

$$u_s(x) = \Delta \Pi(x)^{1/2},$$

(36)

(37)

respectively. Both the buyer and the seller obtain a maximum utility of $\{D(P_{\text{max}} - P_{\text{min}})\}^{1/2}$, the buyer when the bargaining outcome is $(Q^*, P_{\text{min}})$, and the seller when the bargaining outcome is $(Q^*, P_{\text{max}})$. The Pareto frontier is described by the equation

$$u_b(x) = \{D(P_{\text{max}} - P_{\text{min}}) - u_s(x)^2\}^{1/2}.$$

(38)

Nash’s model predicts a bargaining solution at a point on the Pareto frontier that maximizes

$$u_b(x)u_s(x) = \Delta C(x)^{1/2}\Delta \Pi(x)^{1/2} = D\{(P_{\text{max}} - P^*)(P^* - P_{\text{min}})\}^{1/2},$$

(39)

where $P^*$ is the average unit price of the negotiated discount at $Q^*$. Expression (39) has a maximum at

$$P^* = (P_{\text{max}} + P_{\text{min}})/2,$$

(40)

the mid point between $P_{\text{max}}$ and $P_{\text{min}}$. Thus, when both the seller and the buyer are equally risk averse, Nash’s model predicts that, as in the risk-neutral case (Example 1), the transaction upon which the buyer and the seller agree will be $(Q^*, (P_{\text{max}} + P_{\text{min}})/2)$, and that they will equally split the efficiency gain. The Kalai and Smorodinsky solution is identified by the point at which the straight line

$$u_b(x) = u_s(x)$$

(41)

intersects the Pareto frontier (39). The solution once again corresponds to an equal division of the efficiency gain between the buyer and the seller.

Observe that the Nash and Kalai and Smorodinsky models do not incorporate the power of negotiating agents in predicting the bargaining outcome. An arbitration model due to Elbashberg (1986) suggests one way to reflect the effect of differences in buyer and seller power on the distribution of efficiency gain. Briefly, his model predicts a solution at the point at which the Pareto frontier intersects an allocation trajectory in the utilities space. The slope of the allocation trajectory is inferred from the aggregation weights of a group utility function. Aggregation weights are in turn estimated by maximizing a group utility function that reflects the joint preferences of the buyer and the seller. In a bargaining context, the relative magnitude of the aggregation weights reflects the relative power of the buyer and the seller. To illustrate, let $\lambda_b$ and $\lambda_s$ denote the aggregation weights for the buyer and seller, and let $\lambda_b/\lambda_s = 4$ reflect that the buyer has four times the bargaining power of the seller. Consider the risk-averse buyer and seller of Example 3. Elbashberg (p. 968, Example 2) shows that in this case the allocation trajectory in the utilities space is

$$u_b = \left([1 - \alpha]/\alpha\right)^{1/2} u_s,$$

where

$$u_b = \left([1 - \alpha]/\alpha\right)^{1/2} u_s,$$  

(42)
\[ \alpha = (1 + (\lambda_b/\lambda_s)^2)^{-1} \]  \hspace{1cm} (43)

For \( \lambda_b/\lambda_s = 4 \), \( \alpha = 1/17 \), and the allocation trajectory is \( u_b = 4u_s \). Equivalently, the buyer’s share of the efficiency gain is \( 16 D(P_{\text{max}} - P_{\text{min}})/17 \), the seller’s share is \( D(P_{\text{max}} - P_{\text{min}})/17 \), and the negotiated average unit price is \( (P_{\text{max}} + 16P_{\text{min}})/17 \).

The preceding examples illustrate important properties of the transaction negotiated by the buyer and the seller. First, if the buyer and the seller are risk neutral (Example 1), or are equally risk averse (e.g., Example 3), both the Nash and the Kalai and Smorodinsky models predict that the parties equally share the efficiency gain. Second, the less risk-averse agent (e.g., the buyer in Example 2) obtains a larger share of the efficiency gain according to both models (Roth 1988). Third, both models predict that the more risk averse the opponent, the higher the share of the efficiency gain obtained by an agent. For example, a risk-neutral buyer obtains a higher share of the efficiency gain by bargaining against a risk-averse seller (Example 2) than against a risk neutral seller (Example 1) (Kihlstrom, Roth and Schmeidler 1981). Finally, Eliahsberg’s model suggests that the higher the bargaining power of an agent, the larger his/her share of the efficiency gain.

Consider an average unit price \( P^* \) negotiated by the buyer and the seller. A quantity discount is useful because it is a self-enforcing mechanism for achieving transactions at the negotiated outcome \( (Q^*, P^*) \). Certain contractual agreements themselves are equivalent to a quantity discount. For example, consider a buyer and a seller who agree that the unit price will be reduced to \( P^* \) if the buyer orders lots of \( Q^* \) units, but not otherwise. The agreement is equivalent to the seller providing an all-units quantity discount with a single price reduction of \( P_0 - P^* \) at \( Q^* \). The important aspect of the discount from our perspective is that it is also self-enforcing. That is, it is in the economic interest of the buyer not to deviate from the negotiated outcome \( (Q^*, P^*) \), because he incurs a higher annual cost from pursuing any other ordering policy.\(^5\)

Any other all-units quantity discount is also reasonable for enforcing transaction at \( (Q^*, P^*) \), as long as the discount schedule touches the iso-cost curve \( C^* = C(Q^*, P^*) \) only at \( (Q^*, P^*) \), and lies above the iso-cost curve \( C^* \) everywhere else (Figure 2a). The reason for the condition is that all ordering policies \( (Q, P) \) that lie below the iso-cost \( C^* \) have a lower annual cost than \( C^* \), and hence the buyer has an incentive to deviate from \( (Q^*, P^*) \). Of course, there is no reason for the seller to offer a discount schedule with a break at quantities other than \( Q^* \), because additional breaks provide no benefit to the seller. The purpose of the discount schedule is to enforce transaction at \( (Q^*, P^*) \), and a single break at \( Q^* \) suffices for the purpose.

Incremental discounts considered by Dolan (1978), Lal and Staelin (1984) and Dada and Srikanth (1987) are more complex pricing policies that can also serve as self-enforcing mechanisms for affecting transaction at \( (Q^*, P^*) \). The only relevant constraint on the discount is that the corresponding average unit cost curve should touch the iso-cost curve \( C^* \) at \( (Q^*, P^*) \) and should be above the iso-cost curve \( C^* \) everywhere else (Figure 2b). If the latter condition is not satisfied, then a transaction at a point below the iso-cost curve \( C^* \) is feasible, and the buyer again has an incentive to unilaterally deviate from \( (Q^*, P^*) \).

Thus from the standpoint of enforcing a negotiated outcome, there is no advantage that an all-units or incremental quantity discount has over the other. Also, the efficiency gain is maximized by every Pareto efficient transaction, and each such transaction is achieved by both types of discounts. Hence there also is no efficiency rationale for preferring one discount to the other. Even a certain discount structure is not preferred to

\(^5\) Observe that reducing the unit price to \( P^* \) for all quantities does not ensure transaction at \( (Q^*, P^*) \), because the buyer then has an incentive to order lots of \( Q_0 \) units, his economic order quantity for any fixed unit price.
another: the simple policy of offering a single break of $P_0 - P^*$ at $Q^*$ is as good as any more complicated discount policy that can be implemented.

Two-part tariffs can also be used to enforce transaction at $(Q^*, P^*)$. In this case, the buyer pays the seller a fixed fee $F$ up front, and an order-processing fee $f$ for each order. It can be verified that a cost-minimizing buyer will have no incentive to deviate from ordering lots of $Q^*$ units if $f = (A_h + aH)/(H - h)$ and $F = D[P^* - (f/Q^*)]$. Thus as far as self-enforcing mechanisms for transactions at $(Q^*, P^*)$ are concerned, quantity discounts represent a pricing method that is commonly observed in practice (Howell et al. 1986), but not the only possible method.⁶

3. Conclusion

A transactions-efficiency rationale for quantity discounts is analyzed in the context of a bargaining problem in which a buyer and a seller negotiate over lot size orders and average unit prices. The outcome of the negotiation is shown to maximize the joint efficiency gain between the buyer and the seller, and hence to belong to the set of quantity discounts previously described by Lal and Staelin (1984) and Dada and Srikant (1987). The effect of risk sensitivity and bargaining power on the negotiated discount is examined.⁷

⁶ The authors thank Sridhar Moorthy and a reviewer for bringing this alternative mechanism to their attention.
⁷ The authors thank the reviewers for helpful suggestions on earlier versions of the article.

Appendix: Proofs of Propositions

PROOF OF PROPOSITION 1. Referring to Figure 1, consider the point of tangency, $y$, between the buyer's iso-cost curve $C(y)$ and the seller's iso-profit curve $II(y)$. Although the buyer's cost is the same for all points
on $C(y)$, the seller’s profit is maximum at point $y$ on this iso-cost curve. Hence $y$ is not dominated by any other point on $C(y)$. Similarly, although the seller’s profit is the same for all points on $\Pi(y)$, the buyer’s cost is minimum at point $y$ on this iso-profit curve. Consequently $y$ is also not dominated by any other point on $\Pi(y)$. Now consider any point $x$ in the region $A-B-C-D-A$ that does not lie on either of the curves $C(y)$ and $\Pi(y)$. Then $x$ lies on either a higher iso-cost curve $C(x)$ (i.e., $C(x) > C(y)$), in which case the buyer prefers $y$ to $x$, or it lies on a lower iso-profit curve $\Pi(y)$ (i.e., $\Pi(y) < \Pi(y)$), in which case the seller prefers $y$ to $x$; in either case, $y$ is not dominated by $x$. As $x \neq y$ is any arbitrary point in $X$, $y$ is not dominated by any point in $X$. Therefore $y \in Y$, and the set of Pareto efficient discounts is not empty. Also, as $C(y)$ and $\Pi(y)$ are any pair of iso-cost and iso-profit curves with a common tangent point in $X$, the locus of tangencies between iso-cost and iso-profit curves in $X$ describes Pareto efficient discounts. Let $Z \subset X$ be the set of discounts described by this locus of tangencies. We show next that all discounts not in $Z$ are dominated by some discount in $Z$, so that only discounts in $Z$ are Pareto efficient (i.e., $Y = Z$).

Consider any discount $x \in X - Z$ which lies on the iso-cost curve $C(x)$. Let $z \in Z$ be the discount corresponding to the point at which the iso-cost curve $C(x)$ is tangent to some iso-profit curve $\Pi(z)$ (Figure 1). Then the seller prefers $z$ to $x$, the buyer is indifferent between $z$ and $x$, and therefore $z$ dominates $x$. Therefore every discount $x \in X - Z$ is dominated by some discount $z \in Z$. It follows that the only Pareto efficient discounts are those given by the locus of tangencies between the iso-profit and iso-cost curves in $X$; i.e., $Z = Y$.

**Proof of Proposition 2.** The iso-cost curves for the buyer and the iso-profit curves for the seller are described by equations (4) and (9), respectively. The first order derivatives of equations (4) and (9) w.r.t. $Q$, are

$$dP_b/dQ_s = A/Q_s^2 - H/(2D), \tag{A.1}$$
$$dP_s/dQ_s = -a/Q_s^2 - h/(2D). \tag{A.2}$$

respectively. Because Pareto efficient discounts are described by the points of tangency between the seller’s iso-profit and buyer’s iso-cost curves, they occur when (A.1) and (A.2) are equal. Equating (A.1) and (A.2) and solving for $Q_s$ leads to the desired expression for the lot size order associated with the Pareto efficient discounts:

$$Q_s^* = (2A + aD)(H - h)^{1/2}. \tag{A.3}$$

Because $Q_s^*$ is independent of $P$, the values of $P_{max}$ and $P_{min}$ are obtained by identifying the points at which the vertical line $Q = Q_s^*$ intersects the buyer’s iso-cost curve $C(0)$, and the seller’s iso-profit curve $\Pi(0)$, respectively. Setting $Q_s = Q_s^*$ and $\Delta C(x) = 0$ in equation (4) yields expression (15) for $P_{max}$. Similarly, setting $Q_s = Q_s^*$ and $\Delta \Pi(x) = 0$ in equation (9) yields expression (16) for $P_{min}$.

**Proof of Proposition 3.** Let $x \in X$ be a feasible discount for which the seller’s profit is $\Pi(x)$. Then

$$\Delta \Pi(x) = \Pi(x) - \Pi(0) \tag{A.4}$$

where $\Pi(0)$ and $\Pi(x)$ are given by equations (6) and (7), respectively. Thus

$$\Delta \Pi(x) = \{D(P_b - v) - aD/Q_s^0 + hQ_s/2 - FC\} - \{D(P_0 - v) - aD/Q_b^0 + hQ_b/2 - FC\}$$

$$= DP_b - D\{P_0 + a(1/Q_s^0 - 1/Q_b^0) - (h/2D)(Q_s - Q_b)\}$$

$$= DP_b - DP_0 \tag{A.5}$$

where

$$P_t = P_0 + a(1/Q_s^0 - 1/Q_b^0) + (h/2D)(Q_s - Q_b). \tag{A.6}$$

Similarly, the buyer’s cost at $x$ is $C(x)$, and therefore

$$\Delta C(x) = C(0) - C(x) \tag{A.7}$$

where $C(0)$ and $C(x)$ are given by equation (1) and (3), respectively. Thus

$$\Delta C(x) = (DP_b + AD/Q_b^0 + HQ_b/2) - (DP_s + AD/Q_s^0 + HQ_s/2)$$

$$= -DP_s + D\{P_0 - A(1/Q_s - 1/Q_b^0) - (H/2D)(Q_s - Q_b)\}$$

$$= -DP_s + DP_0 \tag{A.8}$$

where

$$P_t = P_0 - A(1/Q_s - 1/Q_b) - (H/2D)(Q_s - Q_b). \tag{A.9}$$

Also, by definition

$$G(x) = \Delta \Pi(x) + \Delta C(x). \tag{A.10}$$

Substituting for $\Delta \Pi(x)$ from (A.4) and $\Delta C(x)$ from (A.5) into equation (A.7) yields

$$G(x) = D(P_t - P_t) \tag{A.11}$$

$$= D\{(A + a)/Q_b + Q_b(H - h)/(2D) - (A + a)/Q_s - Q_s(H - h)/(2D)\}. \tag{A.12}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Differentiating the expression for $G(x)$ in (A.9) with respect to $Q$, setting it equal to zero and solving for $Q$, yields $Q = \left\{ (A + a)D/(H - a) \right\}^{1/2}$ which is the expression for $Q^\ast$. Because $d^2G(x)/dQ^2 = -(A + a)/Q^2 < 0$, $Q^\ast$ maximizes $G(x)$. Finally, it can be verified that $P_t = P_{max}$ and $P_s = P_{max}$ at $Q^\ast$. It therefore follows from (A.8) that $G = D(P_{max} - P_{min}) = \Delta II(y) + \Delta C(y)$ for $y \in Y$, which is the desired result.

**Proof of Proposition 4.** Recall that the buyer's annual cost and the seller's profit associated with $(Q^\ast, P^\ast)$ are denoted $C(y)$ and $II(y)$, respectively. Then

$$\Delta C(y) = C(0) - C(y)$$

$$= \left( DP_0 + AD/Q_0 + HQ_0/2 \right) - \left( DP^\ast + AD/Q^\ast + HQ^\ast/2 \right)$$

$$= D \left[ P_0 - A(1/Q^\ast - 1/Q_0) - (H/2D)(Q^\ast - Q_0) - P^\ast \right]$$

$$= D(P_{max} - P_{min})$$

and

$$\Delta II(y) = II(y) - II(0)$$

$$= \left\{ D(P^\ast - v) - aD/Q^\ast + hQ^\ast/2 - FC \right\} - \left\{ D(P_0 - v) - aD/Q_0 + hQ_0/2 - FC \right\}$$

$$= DP^\ast - D \left\{ P_0 + a(1/Q^\ast - 1/Q_0) - (h/2D)(Q^\ast - Q_0) \right\}$$

$$= D(P^\ast - P_{min})$$

which are the desired expressions for $\Delta C(y)$ and $\Delta II(y)$, respectively.

**References**


