

Tradeoffs between Specific Investment and Optimal Resource Allocation: A Comparison of Different Transfer Pricing Policies

Savita A Sahay¹

¹ Assistant Professor, Rutgers Business School, USA

Correspondence: Savita A Sahay, Assistant Professor, Rutgers Business School, USA

Received: June 22, 2018

Accepted: July 11, 2018

Online Published: July 17, 2018

doi:10.5430/afr.v7n3p221

URL: <https://doi.org/10.5430/afr.v7n3p221>

Abstract

This paper uses a principal-agent framework to analyze the tension between incentives for specific investment by the agent, and resource allocation that is optimal from the principal's perspective. The analysis considers a decentralized firm in which central management can institute different transfer pricing policies to motivate divisional managers to undertake investment and production decisions. Some well-known properties of the methods are identified: a transfer price that uses a markup over and above actual costs can provide investment incentives, but leads to sub-optimal resource allocation; negotiated transfer pricing suffers from the problem of under-investment even though its *ex post* performance is optimal; and standard cost-based transfer pricing entails over-reporting standards, which results in inefficient levels of trade as well as low investment.

The paper establishes a clear ranking amongst the three methods studied. It is shown that the overall performance of actual cost-based transfer pricing is superior if the buying division's investments are important, while negotiated transfer pricing dominates if those of the selling division are important. The overall performance of standard cost-based method is inferior to that of the actual cost-based method, even though the latter has weaker investment incentives.

Keywords: specific investment, transfer pricing, principal-agent analysis, performance evaluation

1. Introduction

Widespread use of transfer pricing (TP) policies in large corporations has been documented in various studies (Note 1). Given the considerable variety of TP methods in use, it is natural to ask what drives a company to choose one method over another, or to identify characteristics of a decentralized firm's environment that favor particular TP policies.

This paper theoretically analyzes the performance of three commonly used TP policies – actual cost-based TP, negotiated TP and standard cost-based TP. (Note 2) The researcher studies these methods using the incomplete contracting framework in which divisional managers have the opportunity to make relationship-specific investments that enhance the value of the intra-firm transfer. These investments can take different forms, e.g. research and development (R&D), machinery and equipment, costly external partnerships or specialized training for personnel.

Investments require an up-front fixed cost, but reduce the *ex post* variable cost, or, in the case of investments made by the downstream division, increase the revenue obtained for each unit of the good. The transfer price must not only create incentives for these investments, but also guide the resource allocation for intra-company trade. Thus, the performance of a TP policy is measured by overall firm profit, which depends on the levels of investment chosen as well as on the level of production.

The current paper is an attempt to unite several comparisons of different TP schemes in a common framework. For each of the three TP policies studied in this paper, the initial analysis focuses on a quantitative evaluation of its problems and characteristics. Then, It evaluates the performance of these methods relative to one another. By applying a uniform framework to the different policies, the researcher is able to identify conditions under which one method is demonstrably superior to the others.

The paper's first result is to extend the finding of Sahay (2003) who shows the optimality of additive markups among

the family of actual cost based TP methods. It shows that additive markup is better even when both divisions have investment opportunities, even though such a markup distorts the level of operation by raising the effective cost to the buying division. This result reflects the conventional wisdom of the accounting profession that judicious use of transfer pricing in a decentralized firm must trade off profit maximization against fairness in evaluating divisional performance. Unless each party can earn a reasonable margin on the transaction, it will not be motivated to further the long-term interests of the firm as a whole.

Next, the analysis characterizes the optimal markup that the transfer price should provide. The desired markup trades off the beneficial effect of inducing investment from the seller against the detrimental effect of reducing both intra-firm trade and investment from the buyer. The optimal markup depends, naturally enough, on various model parameters such as the relative costs of investment from the two divisions.

The second policy studied is negotiated transfer pricing. By adapting the surplus sharing model of Edlin and Reichelstein (1995), the paper shows that although trade levels are optimal under this policy, investments are 'held up' because managers anticipate the failure of ex post negotiations to take the prior 'sunk' costs into consideration. The first comparative result is that the performance of actual cost-based transfer pricing is superior to that of negotiated transfer pricing, except for firms in which the seller's investment is more important than the buyer's. In particular, if investment opportunities are only available to the buyer, actual cost-based TP dominates, and if they are only available to the seller, negotiated transfer pricing dominates. In the general case, with investment opportunities available bilaterally, the outcome of the comparison depends on the relative costs and benefits to the firm of investments from the two divisions.

The third policy in this research study is standard cost-based transfer pricing. It reaffirms the observation of Baldenius, Rechelstein and Sahay (1999) that due to the over-reporting of standards (Note 3), the buying division faces a higher cost and orders a sub-optimal production quantity. On the other hand, this method has investment incentives 'built in' since regardless of other considerations, the seller would like to keep actual costs as low as possible.

This paper demonstrates the dominance of the actual cost based method by showing that the overall performance of standard cost-based TP is inferior to that of actual cost-based TP. This result holds even though the seller's investment is higher in the former. A useful interpretation of the result is to view the seller's overstatement of cost in the standard cost-based method as an 'endogenous' markup that the transfer pricing policy leaves to his discretion. In contrast, the actual cost-based method keeps a measure of control with upper management by letting it specify the markup exogenously. This added control is sufficient to keep trade levels from getting excessively distorted.

The earliest literature on comparative transfer pricing was based on surveys. A seminal work by Eccles and White (1988) who study thirteen firms in depth to examine the tensions and tradeoffs resulting from each of the three common forms of transfer pricing (negotiated, cost-based and market-based) conclude that choice of transfer pricing policy is dictated largely by the firm's strategy of vertical integration. Shelanski (1993) investigates the use of negotiated and administered transfer pricing policies by a large high-technology firm and finds that transaction costs affect the choice of transfer pricing policy.

More recently, several researchers have studied this question in a limited theoretical setting by analyzing two TP policies and studying how one of them performs relative to the other in the presence of specific economic factors. Using a model of incomplete contracting with specific investments, Baldenius, Reichelstein and Sahay (1999) compare the performance of negotiated TP with that of a standard cost-based policy. Dikolli and Vaysman (2006) also compare negotiated and standard cost-based TP in a model that studies the impact of information technology. Lengsfeld, Pfeiffer and Schiller (2006) have also attempted a comparison between actual cost-based TP with two standard cost-based schemes. However, the focus of their study is the cost of information gathering as the main deciding factor. Pfeiffer, Schiller and Wagner (2011) also compare the performance of some cost-based transfer pricing policies and show that the ranking of methods depends upon the degree of *ex ante* cost uncertainty. Matsui (2012) compares the performance of two different cost based transfer pricing policies and demonstrates the superiority of full cost based method over Variable cost method. However, this paper does not consider bilateral investments and Per unit markup in the paper.

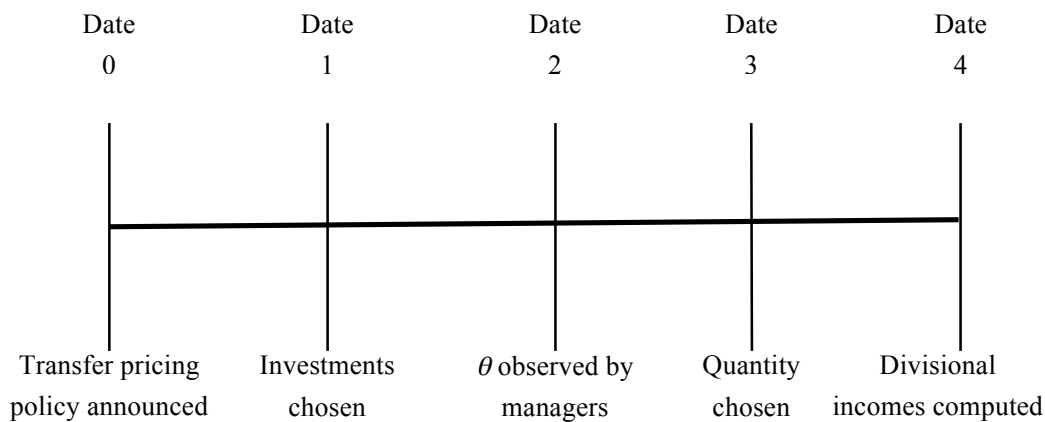
The remaining paper is organized as follows: Section 2 describes the model. Section 3 analyzes actual cost-based transfer pricing and demonstrates its superiority when only the buyer invests; Section 4 analyzes negotiated transfer pricing and Section 5 compares the two policies when only the seller invests. Section 6 compares actual cost-based TP with standard cost-based TP and Section 7 discusses some directions for future research. All proofs are given in the appendix to this paper.

2. The Model

Consider a firm comprising two divisions, the selling division (Division 1) and the buying division (Division 2). Each division operates under the control of a manager employed by the firm’s central management (“headquarters” or HQ.) Division 1 produces a specialized intermediate good and transfers it to Division 2. Division 2 uses the intermediate good as one component of a final product (or “system”), which it sells in an external market.

Managers may undertake relationship-specific investments that enhance the value of the internal transfer. For example, Division 1 might acquire specialized machinery, or hire a consultant or invest in R&D to meet some unique requirements of the intermediate good. Similarly, Division 2 might need to train some employees or buy special equipment to process the intermediate good. Hennart (1988) gives an example from the aluminum industry where refining equipment had to be tailored to the characteristics of the ore from a particular mine, making it unusable at other mines.

Such investments entail up-front fixed costs, and are made in an uncertain environment in which full costs and revenues are not known. Uncertainty is represented in my model by a (possibly multi-dimensional) random variable θ whose value is realized after investments have been chosen. This is indicated in the following timeline:



At Date 0, HQ announces the transfer pricing policy that will be used by the firm. This policy specifies a formula or procedure that will be used to determine the transfer price for the intermediate good. At Date 1, the divisional managers choose their specific investments. Division j 's investment, denoted I_j is chosen from the interval $[0, \bar{I}_j]$ and requires a fixed cost of $V_j(I_j)$, where $V_j(\cdot)$ is a continuous and differentiable function satisfying $V_j(0) = 0$. (Note 4) At Date 2, the state variable θ is realized and jointly observed by both managers. The state θ , in conjunction with investments, determines the actual costs and revenues as described below. Prior to Date 2, all parties hold common beliefs about the distribution of θ .

At Date 3, when all uncertainty about cost and revenue has been resolved, Division 2 chooses q , the number of units of the intermediate good that it will require from Division 1. (Note 5) Division 1 produces and transfers the requested quantity of the intermediate good, incurring a dollar cost of $c(\theta, I_1) \cdot q$, where $c(\theta, I_1)$ is the variable cost of production in state θ , given a specific investment level of I_1 . The function $c(\cdot)$ is differentiable and decreasing in I_1 , i.e. the seller’s specific investment reduces the subsequent variable cost of production. Division 2 uses the intermediate good to put together the final product and sells it in an external market for a revenue of $R(q, \theta, I_2)$ where $R(\cdot)$ is strictly concave in q .

At Date 4, the transfer pricing policy is used by headquarters to arrive at a price T for the internal transaction. The transfer price appears as a charge in the divisional income statement of Division 2 (the buyer), and as a revenue in that of Division 1 (the seller) just as if the intermediate good were being “sold” by one division to the other. Thus, the investments made at Date 1 and the transfer at Date 3 induce the following changes in divisional income:

$$\begin{aligned} \Pi_1 &= T - c(\theta, I_1) \cdot q - V_1(I_1) \\ \Pi_2 &= R(q, \theta, I_2) - T - V_2(I_2) \end{aligned} \tag{2.1}$$

The incentive properties of a transfer pricing policy derive from the fact that divisional income is used for performance evaluation and compensation of managers. In this model, this incentive provision is implicit since managers take all decisions to maximize expected divisional income (Note 6).

The transfer price T may depend on the unit variable cost of production of the intermediate good, but on no other details of the transaction. In large firms with well-diversified divisions, inferring the variable cost for an intermediate product is non-trivial. However, since many firms do employ cost-based methods, it is reasonable to assume that divisional statements reveal at least an approximation of the variable cost to upper management. It should also be noted that while headquarters is assumed to observe the unit variable cost $c(\theta, I_1)$ *ex post*, it does not have the information to disentangle the effects of θ and I_1 and so infer I_1 . Thus, it cannot use the transfer price to directly reimburse the seller for his investment.

As a benchmark, consider the optimal solution from headquarters' point of view. At Date 3, I and θ are fixed so that the efficient (or *first-best*) level of transfer is that which maximizes the firm's contribution margin:

$$M(q, \theta, I) \equiv R(q, \theta, I_2) - c(\theta, I_1)q \quad (2.2)$$

By concavity of $R(\cdot)$ this has a unique maximizer $q^*(\theta, I)$. I assume that $q^*(\cdot)$ is in the interior of the feasible range of q for all values of I and θ . It will be notationally convenient to write $M^*(\theta, I) \equiv M(q^*(\cdot), \theta, I)$.

Since the investments must be chosen under uncertainty, the optimal levels of investment are those that maximize *expected* firm profit:

$$\Pi^*(I) \equiv \Pi^*(I_1, I_2) \equiv E_\theta [M^*(\theta, I)] - (V_1(I_1) + V_2(I_2)) \quad (2.3)$$

where $E_\theta[\cdot]$ denotes the expectation of $[\cdot]$ with respect to the probability distribution of θ . I assume that the profit function (2.3) has a unique maximizer denoted $I^* = (I_1^*, I_2^*)$.

3. Actual Cost-Based Transfer Pricing

To begin with, suppose that headquarters institutes a policy that sets the transfer price equal to unit variable cost, i.e. $T = c(\theta, I) \cdot q$. Given the divisional income computation formulae (2.1), it is clear that no matter what happens at Dates 2 and 3, the contribution to Division 1's income will be:

$$\Pi_1 = T - c(\theta, I_1) \cdot q - V_1(I_1) = -V_1(I_1) \quad (3.1)$$

Thus, Division 1 will not choose any investment at Date 1. To mitigate this problem, headquarters may consider marking up the transfer price (Note 7).

From a pure economic efficiency standpoint, such markups are not necessarily a good idea. Since the effective marginal cost for the buying division is higher when the transfer price is above cost, the scale of operation for the firm will be below the profit-maximizing level. For exactly the same reason, the buyer's investment choice will also be smaller than the profit-maximizing level. Thus, headquarters would find it beneficial to induce investment from the seller, only if it can do so with a sufficiently small markup. This observation also highlights the importance of the form of the markup formula.

Consider first the investment incentives created by a transfer price formula based on a 'fixed percentage' markup, viz. $T = [(1+m)c(\theta, I_1)] \cdot q$ with m being a positive percentage chosen by headquarters at Date 0. (This is the form of markup that most textbooks use to illustrate the cost-based transfer pricing.) Since the positive contribution to the selling division's income is proportional to both $c(\cdot)$ and q , it would like to maximize the total variable cost $C = qc(\cdot)$. Higher investment decreases unit variable cost, causing a decrease in C . But a decreased unit cost has the indirect effect of increasing levels of trade, which increases C . Thus a markup that is proportional to cost has ambiguous investment incentives, with details of the firm's technology likely to determine if the incentives are strong enough to justify the markup.

The actual cost-based policy proposed in this paper does not suffer from this ambiguity. This policy sets the transfer price using the formula:

$$T = (c(\theta, I_1) + m)q \quad (3.2)$$

so that the positive contribution to the seller's divisional income is mq . Since q increases with decreased unit cost — the indirect effect of investment identified earlier — this policy makes specific investment attractive to the seller.

Under this policy the Date 3 quantity would be chosen by the buying division to maximize its contribution to divisional income:

$$R(q, \theta, I_2) - (c(\theta, I_1) + m)q \quad (3.3)$$

Let $\hat{q}(\cdot) \equiv \hat{q}(\theta, I, m)$ denote the maximizer of (3.3). Observe that $\hat{q} < q^*$ whenever $m > 0$ showing that markups distort *ex post* levels of trade.

At Date 1, the two managers find themselves in a game in which each manager wants to choose investment to maximize his expected divisional income. Since divisional incomes depend on the quantity to be decided at Date 3, which in turn depends on both investment levels, each manager must make a conjecture about the other's investment choice when making his own. The model is restricted so that there is a unique Nash equilibrium in this game.

Given m and given a conjecture I_1 for Division 1's investment, Division 2's expected income is given by:

$$\hat{\Pi}_2(I_2, I_1, m) = E_\theta \left[R(\hat{q}(\cdot), \theta, I_2) - (c(\theta, I_1) + m)\hat{q}(\cdot) \right] - V_2(I_2) \quad (3.4)$$

Similarly, Division 1's expected income is given by:

$$\hat{\Pi}_1(I_1, I_2, m) = E_\theta \left[T - c(\theta, I_1)\hat{q}(\cdot) \right] - V_1(I_1) = mE_\theta \left[\hat{q}(\theta, I, m) \right] - V_1(I_1) \quad (3.5)$$

To ensure that the reaction curves are well-defined, the following technical assumptions are necessary

(A1) For I_1 sufficiently close to zero, the firm's expected profit function $\Pi^*(I_1, I_2)$ given by (2.3) has a unique maximizer $I_2^*(I_1)$.

(A2) For m sufficiently close to zero, the seller's objective function $\hat{\Pi}_1(I_1, I_2, m)$ has a unique maximizer $\hat{I}_1(I_2, m)$ whenever $0 \leq I_2 \leq I_2^*$.

Some technical points deserve mention in this context. First of all, observe that unique maximization is only required for small mark ups. This is because we are only interested in ascertaining if a positive markup is beneficial for the firm. As a consequence, the analysis can be restricted to a range of markups close to zero. (Note 8) Secondly, note that while (A2) is a condition on the seller's objective function, (A1) is a requirement of first-best profit, not the buyer's divisional income. The reason (A1) suffices is that for small markups, the transfer price (3.2) allocates the entire profit to Division 2, aligning its interests with that of the firm. Thirdly, the assumptions in themselves ensure only that the reaction curves are well-defined and not that they have a unique intersection. That is, neither existence nor uniqueness of a Nash equilibrium is guaranteed *a priori*. However, from the shape of the seller's reaction curve, it can be shown that there must be a unique equilibrium at small markups.

One final assumption is required for the main result of this section:

(A3) The seller's investment cost function satisfies $V_1'(0) = 0$.

The significance of this assumption is that the seller is willing to invest positive amounts even for small markups (and even if the buyer does not invest at all.) I can now state my first result:

Proposition 1. Suppose that (A1)-(A3) hold. Then it is advantageous for the firm to impose a positive markup over actual cost.

Proof: All proofs are given in the Appendix. ■

This result provides a theoretical justification for the widely observed use of markups in practice. It identifies specific investment opportunities for the seller as the key to understanding the prevalence of such "cost-plus" methods. If firm efficiency is improved by the seller's investment in specific assets, central management should motivate the seller by offering positive markups, despite the distortions of *ex post* trade levels and buyer's investment that these entail.

This analysis also suggests a particular (and particularly simple) form for the markup, given by (3.2). It should be remarked that this form of the markup formula is crucial for the result. In particular, if the markup was chosen in textbook fashion as a fixed percentage of the total variable cost, then the markup would have to exceed a certain threshold before the seller would find it optimal to choose positive investment. The reason is that while such investments lower variable cost, they also lower the seller's contribution margin given the percentage markup. (Note 9) Thus, firm profit would be decreasing in markup until the threshold, since the seller's investment continues to be zero while the quantity decreases. This precludes a guarantee that positive markups are desirable.

3.1 The Optimal Markup

While Proposition 1 establishes the desirability of a positive markup, it does not prescribe a value for the markup. In order for actual cost-based transfer pricing to be implementable, we would like to provide a formula for computing the optimal markup. This value is determined by an interplay among three distinct effects that markups

have on firm profit: higher markups lead to increased investment from the seller, but they lower the investment as well as the quantity chosen by the buyer. Naturally, the magnitudes of these effects depend on specifics of the model's functional parameters.

Accordingly, the optimality question is studied in a setting characterized as follows:

(A4) *The firm's revenue, cost, and investment functions are given by:*

$$\begin{aligned}
 R(q, \theta, I_2) &= (a(\theta) - bq + I_2)q \\
 c(\theta, I_1) &= c(\theta) - I_1 \\
 V_1(I_1) &= \frac{1}{2}v_1I_1^2 \\
 V_2(I_2) &= \frac{1}{2}v_2I_2^2
 \end{aligned}$$

(A5) $2bv_j > 1$ for $j = 1, 2$.

The simple linear and quadratic forms in (A4) allows to show the tradeoffs that headquarters faces when choosing a markup and how the optimal markup achieves a balance between these tradeoffs. Given these forms, (A5) implies that firm profit is strictly concave in either division's investment. It will be convenient to define the shorthand $\gamma_j = 2bv_j$ for $j = 1, 2$. (Note 10)

Derivation of the optimal markup requires one further assumption. In order to motivate this assumption, consider the quantity \hat{q} that the buyer will order at Date 3:

$$\hat{q}(\theta, I, m) \equiv \operatorname{argmax}_q \left\{ (a(\theta) - bq + I_2)q - (c(\theta) - I_1 + m)q \right\} \tag{3.6}$$

The solution to this is given by:

$$\hat{q}(\theta, I, m) = \begin{cases} \frac{a(\theta) - c(\theta) + I_1 + I_2 - m}{2b} & \text{if } a(\theta) - c(\theta) \geq m - (I_1 + I_2) \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

Note that when m is close to zero, \hat{q} is close to the first-best and therefore, positive. When m gets very large, \hat{q} approaches zero. For intermediate values of m , the quantity will be positive only if investments are high enough, and only in 'good' states, i.e. those in which the random component of the firm's unit contribution margin, $a(\theta) - c(\theta)$, is high.

Definition 3.1. *Define the random variable $\delta(\theta) \equiv a(\theta) - c(\theta)$, and let $f(\cdot)$ and $F(\cdot)$ denote its induced probability density and cumulative distribution functions. Let δ_{\min} and δ_{\max} denote the lowest and highest values attained by $\delta(\theta)$ and let Δ denote its expected value.*

From (3.7) it follows that whenever $m \leq \delta_{\min}$, the Date 3 quantity would be in the interior, irrespective of the realization of θ at Date 2 and even if no investments were made at Date 1.

It can be shown that as long as the quantity is in the interior in all states, the buyer and seller will have the following reaction curves when choosing an investment at Date 1:

$$\begin{aligned}
 \hat{I}_1(I_2, m) &= \frac{m}{\gamma_1} \\
 \hat{I}_2(I_1, m) &= \frac{\Delta + I_1 - m}{\gamma_2 - 1}
 \end{aligned} \tag{3.8}$$

The most striking feature of these curves is that they are largely unaffected by the distribution of θ — the seller's choice is completely independent of θ while the buyer's choice depends on it only via Δ . Thus, the interiority of the *ex post* quantity leads to a much more robust formula for the optimal markup that is unaffected by slight errors in the specification of θ . The following assumption guarantees interiority by ensuring that the firm need not consider any markup beyond δ_{\min} . ((Note 11)

(A6) *The model parameters satisfy the joint condition $\delta_{\min} \geq \gamma_1 \bar{I}_1$.*

To see that this allows exclusion of markups larger than δ_{\min} consider why the firm wants to choose any markup at all. The only reason for having a markup is to induce investment from the seller. So the firm need not consider any markup that would result in a higher investment than \bar{I}_1 , the largest conceivable investment from the seller. (Note 12) (A6) states that this 'largest conceivable markup' is smaller than δ_{\min} . The optimal markup formula can now be stated:

Proposition 2. *Suppose that A4-A6 hold. Then the optimal markup, and the investments it induces are given by:*

$$\hat{m} = \min \left\{ \frac{\Delta}{\gamma_1 + 1 - \gamma_1^{-1} - \gamma_2^{-1}}, \gamma_1 \bar{I}_1 \right\}$$

$$\hat{I}_1 = \frac{\hat{m}}{\gamma_1}$$

$$\hat{I}_2 = \frac{\Delta + \hat{I}_1 - \hat{m}}{\gamma_2 - 1}$$

The explicit form of the markup allows for some interesting comparative statics. The first thing to observe is that the optimal markup is not affected by small changes in the distribution of θ . This is a direct result of the fact that the *ex post* quantity is interior in all states.

Secondly, the optimal markup is rising in Δ . The reason for this is that if $\delta(\theta)$ is large, the distortion created by a markup is a smaller percentage of the quantity. Thus a larger absolute distortion is acceptable to the firm. Similarly, if Δ is high, the buying division's investment is sufficiently high to start with and can withstand the negative impact of higher markups.

Thirdly, the optimal markup is decreasing in the investment cost measures γ_1 and γ_2 , with the dependence on γ_1 being much stronger. To understand this, it is simplest to view the optimal markup as a proxy for the optimal investment to be induced from the seller. If the marginal cost of the seller's investment is higher, the optimal investment (hence the optimal markup) will be lower. Similarly, if the buying division's investment is lower, the negative effects of markup, build up more quickly, breaking even with the advantage of seller investment at a smaller value.

Finally, it is interesting to study the limiting cases of the formula in Proposition 2 as the investment costs for one of the divisions becomes extremely high, making it optimal for the firm to reduce its investment to zero. This corresponds to a unilateral investment regime in which specific investment may not be available to one of the divisions for exogenous reasons. (Note 13) Consider first the case in which only the buyer needs to invest (Note 14). This is achieved by letting the seller's cost of investment $\gamma_1 \rightarrow \infty$. It is readily verified that the optimal markup approaches zero, bearing out the intuition that the only reason for having markups in actual cost-based transfer pricing is to induce the seller to invest. In fact, the following proposition holds in a completely general setting:

Proposition 3. *If only the buyer invests, then transfer pricing at actual cost achieves the first-best.*

The proof is omitted because it is completely straightforward. The performance of pure actual cost-based transfer pricing is optimal because with no markup, the buyer sees the fair cost of production. This motivates efficient decisions on its part. There are no decisions for the seller to make, and hence no need for the headquarters to coordinate his behavior using a markup.

At the other extreme is a unilateral investment regime in which only the seller invests. The optimal markup in this scheme is given by letting $\gamma_2 \rightarrow \infty$ in Proposition 2:

$$\hat{m} = \frac{\Delta}{\gamma_1 + 1 - \gamma_1^{-1}} \quad (3.9)$$

Comparison with Proposition 2 shows surprisingly, that the optimal markup is lower in the unilateral setting than in the bilateral one. Since no markup is required if only the buyer invests, one might expect the optimal markup to keep increasing as we move across the investment spectrum to the other end, where only the seller invests. (Note 15) After all, there are two negative effects of markup in the bilateral regime — *ex post* distortion and reduced investment from the buyer — and only one in the unilateral one. However, it is precisely the presence of buyer's investment that makes the higher distortion associated with higher markups affordable.

4. Negotiated Transfer Pricing

To analyze negotiated transfer pricing, this paper uses the relative bargaining model of Edlin and Reichelstein (1995) and Baldenius, Reichelstein and Sahay (1999). In this model, the managers are assumed to have unequal bargaining strengths when negotiating the transfer price. Specifically, their negotiation is characterized by a parameter, $\alpha \in [0,1]$, which represents the proportion of the trading surplus that accrues to the selling division. Since a high value of the transfer price T favors the seller and a low value favors the buyer, higher values of α correspond to higher transfer prices.

The first thing to observe is that negotiated transfer pricing leads to *ex post* efficiency, since a larger total surplus or contribution margin would be bilaterally favored regardless of how it was to be shared. In other words, the gains-from-trade at Date 3 is exactly $M^*(\theta, I)$, the first-best contribution margin following an investment of I at Date 1 and a state realization of θ at Date 2. α -sharing of the surplus implies that the negotiated transfer price satisfies:

$$T - c(\theta, I)q^*(\theta, I) = \alpha M^*(\theta, I); R(q^*(\theta, I), \theta) - T = (1 - \alpha)M^*(\theta, I) \tag{4.1}$$

Note the absence of specific investment costs from the gains-from-trade computation; as described in the introduction, these costs have been ‘sunk’ when the parties negotiate at Date 3, and must be borne even if there were no trade.

The hold-up result also holds in this model. To see this, it suffices to note that at Date 1, the selling division maximizes its expected contribution to divisional income:

$$N_1(I) = \alpha E_\theta [M^*(\theta, I)] - V(I) \tag{4.2}$$

whereas the firm would prefer that it maximize total firm profit:

$$\Pi^*(I) = E_\theta [M^*(\theta, I)] - V(I) \tag{4.3}$$

5. Comparison of Actual Cost-Based and Negotiated Transfer Pricing

As a starting point in the comparative analysis, it should be noted that the outcome of comparisons is likely to depend on whether the buyer’s or the seller’s investment is more important for the firm. In particular, if the seller’s investment is relatively unimportant, Proposition 3 may be invoked to infer that actual cost-based transfer pricing will dominate. In view of this, the paper concentrates on a setting in which only the seller invests. Extensions of my comparisons to the bilateral regime will follow as a matter of course.

The notation may also be simplified if the buyer’s investment is always zero: the variables I_2 and γ_2 may be dropped, as may the subscripts on I_1 and γ_1 . Accordingly, we write the firm’s revenue function as $R(q, \theta)$ and its cost function as $c(q, \theta, I) = (c(\theta) - I)q$. In this section I assume that:

(A7) *The firm’s revenue function satisfies $R''(\cdot, \theta) \geq 0$ and $R'(q, \theta) \rightarrow \infty$ as $q \rightarrow 0$.*

The latter assumption ensures that the quantity chosen under actual cost-based transfer pricing is positive no matter how high the markup, and the former ensures that it satisfies the following technical inequalities:

Lemma 5.1.

- (a) $\frac{m}{(-R''(\hat{q}(\theta, I, m)))} \leq q^*(\theta, I) - \hat{q}(\theta, I, m) \leq \frac{m}{(-R''(q^*(\theta, I)))}$
- (b) $\frac{\partial}{\partial I} \hat{q}(\theta, I, m) > 0, \frac{\partial}{\partial m} \hat{q}(\theta, I, m) < 0$
- (c) $\frac{\partial^2}{\partial I^2} \hat{q}(\theta, I, m) \geq 0, \frac{\partial^2}{\partial I \partial m} \hat{q}(\theta, I, m) \leq 0$

Let $\Pi(m, I)$ denote the expected profit for the firm if headquarters chooses a markup m and the seller chooses an investment I . Note that I is endogenously chosen in response to m so the profit to the firm from a markup of m may be written as:

$$\hat{\Pi}(m) = \Pi(m, \hat{I}(m)) \tag{5.1}$$

with $\hat{I}(m)$ being the investment chosen by the seller when the markup is m .

For comparative purposes it is better to consider expected firm profit as a function of I . This is the profit that the firm would get if it were to choose the markup in a way that makes the investment I incentive-compatible for the seller. Thus, we would like to define:

$$\hat{\Pi}(I) = \Pi(\hat{m}(I), I) \tag{5.2}$$

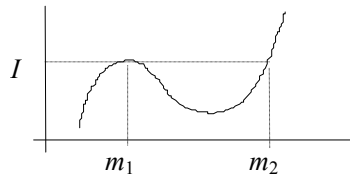
where $\hat{m}(I)$ is the markup required to induce I . However, there is a problem with this definition in that the same investment may be induced by more than one markup. To rectify this problem, this research first shows that among all markups that induce a given level of investment, the firm would prefer the smaller one.

Lemma 5.2 $m_1 < m_2 \Rightarrow \Pi(m_1, I) \geq \Pi(m_2, I)$.

From this lemma, it is clear that $\hat{m}(I)$ should be chosen as the *smallest* markup that induces I .

Definition 5.1. $\hat{m}(I) \equiv \min\{m : \hat{I}(m) = I\}$

Under this definition the firm profit (5.2) is well-defined and the optimal investment under actual cost-based transfer pricing is the maximizer \hat{I} of $\hat{\Pi}(I)$. It should be noted that $\hat{\Pi}(I)$ need not be continuous everywhere. This is because $\hat{m}(I)$ need not be continuous as depicted in the following graph of $\hat{I}(m)$:



By definition, however, $\hat{m}'(I) > 0$ except where it is discontinuous. Moreover, for each discontinuity I of $\hat{m}(I)$ (or $\hat{\Pi}(I)$) there exist two markups m_1 and m_2 such that:

- (i) $\hat{I}(m_1) = \hat{I}(m_2) = I$
- (ii) $\hat{m}(I) = m_1 < m_2$
- (iii) $m \in (m_1, m_2) \Rightarrow \hat{I}(m) < I$

This allows us to show that while $\hat{\Pi}(I)$ may be discontinuous, it only “jumps downward” at its discontinuities.

Lemma 5.3: Let I be any point of discontinuity of $\hat{m}(\cdot)$. Then:

$$\lim_{\substack{h \rightarrow 0 \\ h > 0}} \hat{\Pi}(I + h) \leq \hat{\Pi}(I)$$

Finally, we note that $\hat{\Pi}(I)$ is differentiable everywhere except at its discontinuities. The following lemma characterizes the incremental change in firm profit for a small increase in the seller’s investment (including the cost of the extra markup required.)

Lemma 5.4: $\hat{\Pi}'(I) = E[\hat{q}(\theta, I, \hat{m}(I))] - \hat{m}'(I)v'(I)$

The stage is now set for comparing the performance of negotiated transfer pricing with that of actual cost-based transfer pricing. The fundamental difference between the two policies should be clear by now. Negotiated transfer pricing yields *ex post* efficient quantities, but suffers from underinvestment. Actual cost-based transfer pricing, on the other hand, results in *ex post* trade distortions, but may provide better investment incentives with a suitable markup. The comparison will be facilitated by the following fundamental observation, which is an immediate consequence of *ex post* efficiency under negotiated transfer pricing:

Observation. A necessary condition for any transfer pricing policy to dominate negotiated transfer pricing is that it induce a higher level of investment from the seller.

This observation, together with Lemma 5.4, is used to establish the following comparative result:

Proposition 4. Suppose that only the seller invests, (A7) holds, and $\alpha \geq \frac{1}{2}$. Then negotiated transfer pricing dominates actual cost-based transfer pricing.

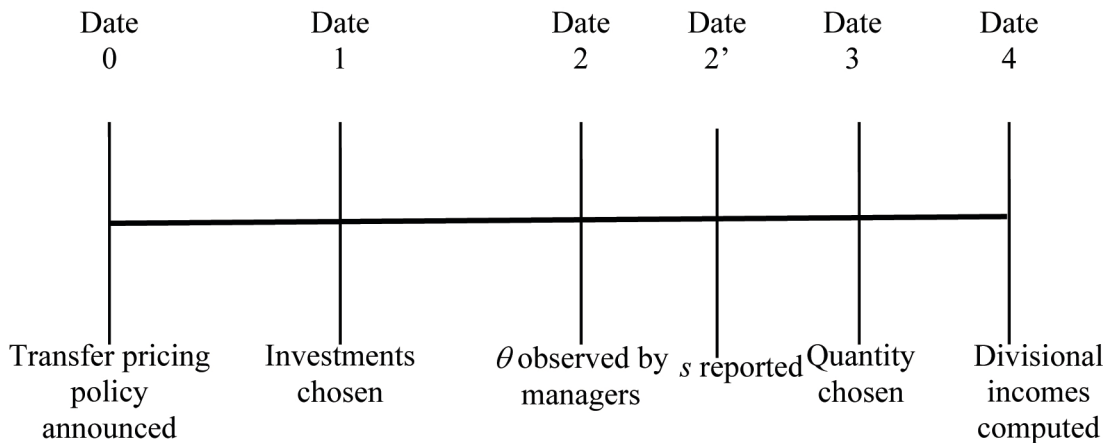
Proposition 4 should be contrasted with the following corollary of Proposition 3, whose proof is immediate, given the underinvestment under negotiated transfer pricing:

Proposition 5. *Suppose that only the buyer invests and $\alpha < 1$. Then actual cost-based transfer pricing dominates negotiated transfer pricing.*

This pair of propositions shows that the relative importance of the investments of the firm’s divisions determines which of the two methods is superior. If the seller’s investment is relatively unimportant, the firm is best off choosing actual cost-based transfer pricing. The opposite is true if the buyer’s investment is unimportant. This ambivalence in ranking carries over to the bilateral investment scenario as well. With the markup chosen as in Proposition 2, the buyer invests more than he would under negotiated transfer pricing while the seller invests less. Whether overall firm profit is higher or lower is determined by the relative costs of investment.

6. Comparison of Actual Cost-Based and Standard Cost-Based Transfer Pricing

To incorporate standard cost-based transfer pricing into my model, the timeline is slightly modified to include a standard cost report by the seller before the buyer chooses the quantity:



The transfer pricing policy announced at Date 0 stipulates that the per unit transfer price will be set equal to the standard cost s stated by the seller at Date 2'. Thus, this policy gives the seller monopoly power in setting the transfer price. This might appear to be an extreme assumption but there is substantial evidence from practice that production managers often have such discretion in the setting of standard costs. (Note 16)

It should also be noted that in this time-line, the seller can delay reporting the standard until all uncertainty has been resolved about actual production costs. In practice, while the production division might have considerable influence in the setting of standards, actual costs will usually be subject to random fluctuations even after standards have been set; the seller will just have to ‘live with’ the standards he sets. This feature could be incorporated into the model by decomposing θ into two components, one of which is revealed only after the standard has been quoted. Again, the results would not be qualitatively affected by this refinement.

To see the effect of the seller’s monopoly power in this model, consider the buyer’s problem at Date 3. Since the transfer price is simply sq , the quantity chosen will be: (Note 17)

$$q_M(\theta, s) = \arg \max_q \{R(q, \theta) - sq\} \tag{6.1}$$

Thus, the buyer’s demand for the intermediate good, as a function of standard cost, is given by the inverse of the marginal net revenue for the final product, denoted $MR^{-1}(s, \theta)$. This is a decreasing function, so that in stating a higher standard cost at Date 2', the seller trades off the higher transfer price received against loss in trade. This is exactly the position of a monopolist who must choose a price when faced with the downward-sloping demand curve MR^{-1} .

Using this demand curve, the direct form of the seller’s Date 2' reporting problem:

$$\max_s q_M(\theta, s)(s - c(\theta, I)) \tag{6.2}$$

may be replaced with an indirect form which treats the choice of s as a vehicle for inducing a desired quantity response. Since the standard cost that should be reported to elicit an order of q_0 units is precisely $s_0 = R'(q_0)$, the seller really solves:

$$\max_q q(a(\theta) - 2bq - c(\theta) + I) \quad (6.3)$$

This yields:

$$q_M(\theta, I) = \frac{a(\theta) - c(\theta) + I}{4b} \quad (6.4)$$

which is exactly half the efficiency level. This is a manifestation of the ‘deadweight’ loss of potential trade associated with monopolistic price-setting. Thus, the model captures precisely the main problems with standard cost-based transfer pricing: the seller takes advantage of the unverifiability of standard cost by overstating it, and trade is inefficient because the transfer price is set too high compared to the actual cost.

Inefficiencies in *ex post* trade translate into inefficiencies in investments as well. This is because returns from the investment are increasing in quantity traded. (Note 18) Formally, the investment chosen at Date 1 maximizes:

$$S_1(I) = E_\theta \left[q_M(\cdot) (a(\theta) - bq_M(\cdot) - c(\theta) + I) \right] - \frac{1}{2} vI^2 \quad (6.5)$$

Using the Envelope Theorem, the first-order condition may be written as:

$$vI = E_\theta \left[q_M(\theta, I) \right] = \frac{\Delta + I}{4b} \quad (6.6)$$

Comparing this with the first-order condition for investment under negotiated transfer pricing (4.2), we can infer that investments are equal under these two policies. It follows from this and the fundamental observation of Section 4.1 that negotiated transfer pricing dominates the standard cost-based transfer pricing. (Note 19)

As shown by equation (6.4), standard cost-based transfer pricing is not *ex post* efficient. Thus, while (6.6) shows that it has stronger investment incentives than actual cost-based transfer pricing, the firm might yet prefer the latter. This is because the optimal markup under actual cost-based transfer pricing accounts for the ill effects of trade distortions on total firm profit, whereas the seller’s standard-setting behavior is driven by divisional interests alone. The next result shows that in fact, the control that management retains over *ex post* distortions in the choice of a markup is enough to ensure its superiority over the standard cost-based method.

Proposition 6. *Suppose that only the seller invests, and that A4, A5 and A7 hold. Suppose also that the model parameters satisfy the joint condition $\delta_{\min} \geq \Delta/(\gamma + 1 - \gamma^{-1})$. Then actual cost-based transfer pricing dominates the standard cost-based transfer pricing.*

The key to understanding this result is to focus on trade distortions alone. It is clear that the only way for actual cost-based transfer pricing to outperform the standard cost-based method is if it results in smaller distortions in trade. Under the actual cost-based scheme, trade distortions are under the explicit control of the headquarters. One way to isolate the effect of these distortions is to use a sub-optimal markup $m_S = \gamma I_S$ that exactly replicates the investment chosen by the seller under standard cost-based transfer pricing. For this choice of markup, comparison of firm profit under the two methods is completely determined by the quantities chosen at Date 3. It can be shown that these are equal in expectation:

$$E_\theta \left[q_M(\theta, I_S) \right] = E_\theta \left[\hat{q}(\theta, I_S, m_S) \right] \quad (6.7)$$

Since the performance of actual cost-based transfer pricing can only improve if the markup is chosen optimally, the result follows.

The technical condition that appears in Proposition 5 is due to the fact that in general, \hat{q} under an optimal markup need not be interior in all states. In such situations, the relationship $m = I\gamma$ (3.8) that quantifies the markup required to induce a given level of investment does not hold, rendering replication of arbitrary investment levels infeasible. The technical condition in effect, allows us to use replication and invoke (6.7).

Of course, the actual cost-based transfer pricing also dominates in the unilateral regime where only the buyer invests. It is no surprise then, that its superiority continues into the bilateral regime. Unlike against negotiated transfer pricing, the superiority here is unconditional and does not depend on the relative costs of investment.

The proof of dominance under a bilateral scenario is based on the same replication idea leading to Proposition 6. By choosing a suitable markup, the firm can replicate the seller’s investment under the standard cost-based method. It can be shown that at this replicating level of markup, the buyer’s investment is higher under the actual cost-based

scheme. (Note 20) The distortion effect then is even more unfavorable for standard cost-based transfer pricing than in the seller-only regime, and the actual cost-based scheme continues to dominate.

The overall ranking of the three policies is extremely interesting. Baldenius, Reichelstein and Sahay (1999) have compared negotiated and standard cost-based transfer pricing and show that the former is demonstrably superior under quite general conditions. The results of this section and Section 5 place actual cost-based transfer pricing at the head of the trio if only the buyer invests and in between the other two if only the seller invests. If both invest, the best policy for the firm depends on whose investment is more important.

7. Conclusion

This paper analyzes the performance of a transfer pricing policy based on actual variable cost. The analysis first establishes the desirability of a markup over variable cost, even though such markups lead to suboptimal levels of transfer. The reason a positive markup is desirable is that it motivates the seller to make an investment up front. The value of the investment induced more than offsets the 'cost' of providing incentives via a distorting markup. When functional forms are assumed for the model parameters, an explicit formula can be given for the optimal markup. The paper also finds that the optimal markup is larger if both parties invest than if only the seller does.

Secondly, the performance of this policy is compared against the negotiated transfer pricing method. The approach of this comparative analysis is different from that of classical mechanism design: rather than search for optimal mechanisms, this paper concentrates on designing mechanisms that reflect features of transfer pricing policies used in practice, and on comparing the extent to which they achieve congruence with overall firm goals. The analysis finds that actual cost-based transfer pricing also dominates negotiated transfer pricing if only the buyer invests, while the opposite is true if only the seller invests.

It would be desirable to extend the analysis presented here in several directions. The first generalization is to consider full cost-based transfer pricing. In my model, fixed cost was not considered observable. In reality, many firms use full cost-based methods (much to the perplexity of theoreticians.) But the accounting definition of fixed cost in most firms is usually based on some discretionary rule for allocating overhead across goods. (Note 21) Thus, one direction of extension is to a firm with more than one good involved in the determination of fixed cost. In such cases, the selling division usually has some scope for 'shifting' overheads and would therefore have indirect control of the transfer price. Transfer pricing based on actual cost may then become significantly less attractive.

Fixed costs also introduce the notion of depreciation since they are incurred in one period and accounted for in future periods. Allocating the total fixed cost over more than one period using a depreciation policy might create different investment incentives than suggested by the single-period analysis of this paper. Thus, there are many practical problems that occur when actual costs are used in transfer pricing, and the results presented in this paper may stimulate further research on analyzing these problems.

The results also lend themselves to empirical testing. For example, Proposition 1 may be viewed as asserting that a cost-based transfer price is likely to incorporate a markup if specific investment from the seller is important. The latter is usually associated with a specialized intermediate product, so that an empirical investigation could seek to determine if there is a correlation between the use of cost-plus policies and the degree of specialization of the transferred good. Shelanski (1994) has attempted to test similar hypotheses in the context of administered versus negotiated transfer pricing, concluding that administered transfer pricing is favored for transfer of specialized goods.

References

- Ancil, R. & Dutta, S. (1999). Negotiated Transfer Pricing and Divisional versus Firm-Wide Performance Evaluation. *The Accounting Review* 72, 87–104.
- Arrow, K.J. (1962). Economic Welfare and the Allocation of Resources for Inventions. In Nelson, R. (ed.), *The Rate and Direction of Inventive Activity* (pp. 609-626). Princeton University Press.
- Baldenius, T., Reichelstein, S. & Sahay, S. (1999). Negotiated vs. Administered Transfer Pricing. *Review of Accounting Studies*, Vol. 4, 67-91.
- Christensen, J. & Demski, J. (1990). *Transfer Pricing in a Limited Communication Setting*. Working paper, Yale University, New Haven, CT.
- Dikolli, S. & Vaysman, I. (2006). Information technology, organizational design, and transfer pricing. *Journal of Accounting and Economics*, Volume 41, Issues 1–2, 201-234.
- Eccles, R. & White, H. (1988). Price and Authority in Inter-Profit Center Transactions. *American Journal of*

Sociology, Vol. 94 Supplement, S17-S48.

- Eccles, R. (1985). *The Transfer Pricing Problem: A Theory for Practice*. Lexington Books, Lexington.
- Edlin, A. & Reichelstein, S. (1995). Specific Investment under Negotiated Transfer Pricing: An Efficiency Result. *The Accounting Review, Vol. 70, No. 2*, 88-94.
- Ernst & Young (2003). *Increased importance of transfer pricing documentation*. Global survey.
- Ernst & Young (2007). *Global Transfer Pricing Trends, Practices and Analysis*. Global surveys.
- Ernst & Young (2013). *Global Transfer pricing survey*.
- Hennart, K.S. (1988). Upstream Vertical Integration in the Aluminum and Tin Industries: A Comparative Study of the Choice between Market and Intra-firm Co-ordination. *Journal of Economic Behavior and Organization, 9*, 281-300.
- Hirshleifer, J. (1956). On the Economics of Transfer Pricing. *Journal of Business*, July 1956.
- Holmstrom, B. & Tirole, J. (1991). Transfer Pricing and Organizational Form. *Journal of Law, Economics and Organization, 7*, 201-228.
- Lengsfeld, S., Pfeiffer, T. & Schiller, U. (2006). *Centralized versus decentralized transfer pricing and cost-system choice*. Available at SSRN 925170.
- Matsui, K. (2012). Cost-based transfer pricing under R&D risk aversion in an integrated supply chain. *International Journal of Production Economics, Volume 139, Issue 1, September 2012*, 69-79.
- Pfeiffer, T., Schiller, U. & Wagner, J. (2011). Cost-based transfer pricing. *Review of Accounting Studies 16*, 219-246.
- Price Waterhouse (1984). *Transfer Pricing Practices of American Industry*. New York.
- Sahay, S. (2003). Transfer Pricing based on Actual Cost, *Journal of Management Accounting Research, Vol. 15*, 177-192.
- Shelanski, H. (1994). *Transfer Pricing and the Organization of Internal Exchange*. Unpublished dissertation, University of California, Berkeley.
- Tang, R. (1992). Transfer Pricing in the 1990s. *Management Accounting, 73, No.8*, 22-26.
- Vancil, R. (1978). *Decentralization: Managerial Ambiguity by Design: A Financial Executives Research Foundation Study*. Homewood, Illinois, Dow Jones-Irwin.
- Vaysman, I. (1996). A Model of Cost-based Transfer Pricing. *Review of Accounting Studies, 1*, 73-108.
- Vaysman, I. (1998). A model of negotiated transfer pricing. *Journal of Accounting and Economics, 25*, 349-384.

Notes

Note 1. Early examples of such studies are Vancil (1978), Price-Waterhouse (1984), Eccles (1985), Eccles and White (1988), while some more recent research is presented in Tang (2003) and Ernst and Young (2003, 2008, and 2013).

Note 2. While market-based TP is a popular choice for certain kinds of products and firms, it cannot be used when there is no easily identified external market price for the transferred intermediate good. This could be either because the intermediate good is specialized, or because the internal product is different from the products available externally in terms of quality and customer service. This study addresses the large majority of cases where no market price is available for the transferred good. (See, for example, Shelanski (1994) who notes that for about 90% of the products in his study, there was no ready market.)

Note 3. Eccles and White (1985) provide empirical support for the “soft” standards that result from production managers’ participation in their determination. They quote a manager as saying: “all the problems [in standard cost-based transfer pricing] result from the setting of standards. They are never strict enough....”

Note 4. Note that I is an abstract measure of investment, and not necessarily its dollar cost. For example, if specific investment is in specialized machinery, I could denote the reduction in variable cost; if specific investment is in training of personnel I could denote the number of hours of training.

Note 5. Some transfer pricing policies might allow q to be decided bilaterally by a negotiation between the two divisions. See Section 4.

Note 6. I suppress moral hazard problems that may generate a conflict between the manager's and the firm's profit objective. For principal-agent models using transfer pricing, see Vaysman (1996, 1998), Christensen and Demski (1999), and Anctil and Dutta (1999).

Note 7. Allowing a profit above the division's cost is common in practice. Even "arms length" formulae allowed to calculate transfer prices use "Royalty based" methods, distributing the total gains from the transaction between buyer and seller.

Note 8. In Section 3.1 I analyze the value of the optimal markup in a specialized setting. For that setting, I will present conditions that ensure uniqueness for every m .

Note 9. Recall the ambiguous investment incentives of percentage-of-cost markups identified in the build-up to (3.2).

Note 10. My analysis extends readily to forms more general than (A4) except that the first-order condition characterizing the optimal markup would be more complicated and, in general, difficult to solve explicitly. In particular, $V_j(\cdot)$ could be any function that is sufficiently convex.

Note 11. (A6) also ensures the existence of a unique Nash equilibrium in the investment game at Date 1.

Note 12. This is the reason that no analogue of (A6) is required for the buyer's investment. There is no need to exogenously rule out high markups from the standpoint of the buyer's investment since higher markups *decrease* his investment and would therefore be endogenously precluded from consideration.

Note 13. Specific investment may not be required if the intermediate good is not specialized, or if similar goods are traded in the divisions' other lines of business.

Note 14. This setting captures situations in which the seller supplies a standard intermediate product, like an engine for a car but the buyer has to hire an expert to adapt the final good to the local market.

Note 15. Intuitively, one feels that markups should be smaller when both divisions invest since there is less "urgency" to induce investment from the seller.

Note 16. For example, in a detailed case analysis of the standard-setting process, Eccles and White (1988) quote a senior executive: "It is here that games can really be played. [Standard costs] should have been reviewed by the corporate controller and corporate vice-president for manufacturing. As it was, they were determined by the general manager of [the selling division]" (p. S28.)

Note 17. The subscript is a mnemonic for 'monopoly.'

Note 18. Arrow (1962) recognizes a similar 'output effect' in an industrial organization context.

Note 19. This is a special case of the more general result derived in Baldenius, Reichelstein and Sahay (1999).

Note 20. Note that in general a single markup will not be able to replicate investments from both divisions.

Note 21. My model avoids allocation issues by (implicitly) assuming that the divisions' other lines of business do not impinge on the accounting for this transfer.

APPENDIX

Proof of Proposition 1: Since (3.3) is concave in q the optimal quantity $\hat{q}(\cdot)$ is unique. As $m \rightarrow 0$, $\hat{q}(\cdot)$ approaches the first-best level, so there must exist a range of markups $[0, m_0]$ in which $\hat{q}(\cdot)$ is interior. In this range, the sign of $\partial \hat{q} / \partial I_1$ is the same as that of that of:

$$\frac{\partial^2}{\partial q \partial I_1} \{R(q, \theta, I_2) - (c(\theta, I_1) + m) \hat{q}(\cdot)\} = -\frac{\partial}{\partial I_1} c(\theta, I_1) > 0 \quad (1)$$

Lemma 1.1. Under the conditions of Proposition 1, there exists a positive value m_1 such that there is a unique equilibrium at Date 1 as long as $0 < m \leq m_1$.

Proof: By (A2) the seller's response function $\hat{I}_1(I_2, m)$ is well-defined for sufficiently small m . By (A3) it is also interior since,

$$\frac{\partial}{\partial I_1} \Pi_1(0, I_2, m) = mE_\theta \left[\frac{\partial}{\partial I_1} \hat{q}(\theta, 0, I_2, m) \right] > 0 \tag{2}$$

the inequality following from (1) above. So \hat{I}_1 satisfies the first- and second- order conditions for maximization of (3.5). Differentiating the first-order condition with respect to I_2 , we get:

$$\frac{\partial}{\partial I_2} \hat{I}_1(I_2, m) = \frac{\frac{\partial^2}{\partial I_1 \partial I_2} \Pi_1(\hat{I}_1, I_2, m)}{\left(-\frac{\partial^2}{\partial I_1^2} \Pi_1(\hat{I}_1, I_2, m) \right)} \tag{3}$$

The denominator is positive by the second-order condition, so the slope of the seller’s reaction function has the same sign as the numerator. As $m \rightarrow 0$,

$$\frac{\partial^2}{\partial I_1 \partial I_2} \Pi_1(\hat{I}_1, I_2, m) = mE_\theta \left[\frac{\partial^2}{\partial I_1 \partial I_2} \hat{q}(\theta, I, m) \right] \rightarrow 0 \tag{4}$$

Thus, the reaction curve approaches the straight line $I_1 = 0$ as m approaches zero. Therefore, it suffices to show that the buyer’s reaction function is well-defined at $I_1 = 0$, because any well-defined function will intersect the line $I_1 = 0$ exactly once. As $m \rightarrow 0$, we find that the quantity $\hat{q} \rightarrow q^*$ and the seller’s investment $I_1 \rightarrow 0$ so that the buyer’s expected divisional income function in (3.4) approaches the first-best level:

$$\Pi^*(0, I_2) = E_\theta \left[M^*(\theta, 0, I_2) \right] - V_2(I_2) \tag{5}$$

But then it must have a unique maximizer by (A1). Thus, $\hat{I}_2(I_1, m)$ is well-defined. The lemma holds if we define m_1 to be the minimum of m_0 and the values implied by (A1)-(A2). ■

Let $\Pi(m, I_1, I_2)$ denote the expected profit to the firm if markup is m and if the Date 1 investments are I_1 and I_2 :

$$\Pi(m, I_1, I_2) \equiv \Pi(m, I) = E_\theta \left[R(\hat{q}(\cdot), \theta, I_2) - \hat{q}(\cdot) \cdot c(\theta, I_1) \right] - V_1(I_1) - V_2(I_2) \tag{6}$$

Let $\hat{I} \equiv \hat{I}(m) \equiv (\hat{I}_1(m), \hat{I}_2(m))$ denote the unique equilibrium of Lemma 1.1. Then, for $0 < m \leq m_1$, firm profit as a function of the markup is given by:

$$\hat{\Pi}(m) \equiv \Pi(m, \hat{I}(m)) \tag{7}$$

Note that $\hat{\Pi}(\cdot)$ is differentiable when $0 < m \leq m_1$ (since each of its components is.) Its derivative is given by:

$$\begin{aligned} \hat{\Pi}'(m) &\equiv \frac{\partial}{\partial m} \Pi(m, \hat{I}(\cdot)) + \frac{\partial}{\partial I_1} \Pi(m, \hat{I}(\cdot)) \hat{I}'_1(\cdot) + \frac{\partial}{\partial I_2} \Pi(m, \hat{I}(\cdot)) \hat{I}'_2(\cdot) \\ &= \frac{\partial}{\partial q} E_\theta \left[R(\hat{q}(\cdot), \theta, \hat{I}_2(\cdot)) - \hat{q}(\cdot) \cdot c(\theta, \hat{I}_1(\cdot)) \right] \frac{d}{dm} \hat{q}(\cdot) \\ &\quad - \left\{ E_\theta \left[\hat{q}(\cdot) \cdot \frac{\partial c}{\partial I_1}(\theta, \hat{I}_1(\cdot)) \right] - V'_1(\hat{I}_1(\cdot)) \right\} \hat{I}'_1(\cdot) \\ &\quad + \left\{ E_\theta \left[\frac{\partial}{\partial I_2} R(\hat{q}(\cdot), \theta, \hat{I}_2(\cdot)) - V'_2(\hat{I}_2(\cdot)) \right] \right\} \hat{I}'_2(\cdot) \end{aligned} \tag{8}$$

The last term in the final expression of (8) is exactly the marginal benefit of investment to the seller (the derivative of (3.4) in the main text), which is zero by the Envelope Theorem. Using the first-order condition for the quantity decision, $\partial R / \partial q - c(\theta, I_1) = m$, to simplify the first term of this expression, (8) may be simplified to:

$$\hat{\Pi}'(m) = mE_{\theta} \left[\frac{d}{dm} \hat{q}(\cdot) \right] - \left(E_{\theta} \left[\hat{q}(\cdot) \frac{\partial c}{\partial I_1} \hat{I}'_1(\cdot) \right] - V'_1(\hat{I}_1(\cdot)) \right) \hat{I}'_1(\cdot) \tag{9}$$

As $m \rightarrow 0$ the first term in (9) approaches zero. Also, $\hat{I}_1 \rightarrow 0$ so by (A3) the second term approaches:

$$-E_{\theta} \left[\hat{q}(\cdot) \frac{\partial c}{\partial I_1} \right] \hat{I}'_1(0_+) \tag{10}$$

The expectation in (10) is negative, so the sign of $\hat{\Pi}'$ is the same as that of \hat{I}'_1 . It suffices to show that $\hat{I}'_1 > 0$ in the limit for then $\hat{\Pi}(\varepsilon) = \hat{\Pi}(0) + \int_0^{\varepsilon} \hat{\Pi}'(m) dm > \hat{\Pi}(0)$ for small enough ε . Since the seller's investment is positive, the sign of \hat{I}'_1 is the same as that of:

$$\frac{\partial^2}{\partial I_1 \partial m} \Pi_1(\hat{I}_1(\cdot), \hat{I}_2(\cdot), m) = E_{\theta} \left[\frac{\partial}{\partial I_1} \hat{q}(\theta, \hat{I}(\cdot), m) \right] + mE_{\theta} \left[\frac{\partial^2}{\partial I_1 \partial m} \hat{q}(\theta, \hat{I}(\cdot), m) \right] \tag{11}$$

In the limit, the right-hand-side of (11) approaches:

$$E_{\theta} \left[\frac{\partial}{\partial I_1} \hat{q}(\theta, \hat{I}(\cdot), m) \right] \tag{12}$$

which is positive by (1). Thus, $\hat{I}'_1(0_+) > 0$ and we are done. ■

Proof of Proposition 2: For any markup $m \leq \delta_{\min}$, we have:

$$\hat{q}(\theta, I, m) = (\delta(\theta) + I_1 + I_2 - m) / (2b) \tag{1}$$

Thus, the investments are chosen to maximize:

$$\begin{aligned} \hat{\Pi}_2(I_2, I_1, m) &= E_{\theta} [R(\hat{q}(\cdot), \theta, I_2) - \hat{q}(\cdot)(c(\theta, I_1) + m)] - V_2(I_2) \\ &= \frac{1}{4b} E_{\theta} [(\delta(\theta) + I_1 + I_2 - m)^2] - \frac{1}{2} v_2 I_2^2 \end{aligned} \tag{2}$$

and:

$$\hat{\Pi}_1(I_1, I_2, m) = mE_{\theta} \left[\frac{\delta(\theta) + I_1 + I_2 - m}{2b} \right] - \frac{1}{2} v_1 I_1^2 \tag{3}$$

By (A5) each of these functions is concave in the variable of choice and the unique maximizers are given by the following reaction curves:

$$\hat{I}_1(I_2, m) = \frac{m}{\gamma_1}; \hat{I}_2(I_1, m) = \frac{\Delta + I_1 - m}{\gamma_2 - 1} \tag{4}$$

This gives a unique equilibrium, $\hat{I}(m) \equiv (\hat{I}_1(m), \hat{I}_2(m)) = \left(\frac{m}{\gamma_1}, \frac{\Delta + m/\gamma_1 - m}{\gamma_2 - 1} \right)$. Substituting these values into

the profit function yields a function that is concave in m . The unique maximizer can be found using the first-order condition. This turns out to be:

$$\hat{m} = \frac{\Delta}{\gamma_1 + 1 - \gamma_1^{-1} - \gamma_2^{-1}}$$

Proof of Lemma 5.1: Given a markup m and investment I , the seller's choice of quantity satisfies the first-order condition:

$$R'(\hat{q}(\theta, I, m), \theta) - (c(\theta) - I + m) = 0 \tag{1}$$

Now:

$$\int_{\hat{q}(\theta, I, m)}^{q^*(\theta, I)} (-R''(t, \theta)) dt = R'(\hat{q}(\theta, I, m), \theta) - R'(q^*(\theta, I), \theta) = (c(\theta) - I + m) - (c(\theta) - I) = m \tag{2}$$

Since the integrand is positive, $q^* > \hat{q}$ and since $R''' \geq 0$, we have:

$$\begin{aligned} (q^*(\theta, I) - \hat{q}(\theta, I, m))(-R''(q^*(\theta, I))) &\leq \int_{\hat{q}(\theta, I, m)}^{q^*(\theta, I)} (-R''(t, \theta)) dt = m \\ (q^*(\theta, I) - \hat{q}(\theta, I, m))(-R''(\hat{q}(\theta, I, m))) &\geq \int_{\hat{q}(\theta, I, m)}^{q^*(\theta, I)} (-R''(t, \theta)) dt = m \end{aligned} \tag{3}$$

This establishes (a). Differentiating (1) with respect to I and m twice we have:

$$\begin{aligned} R''(\hat{q}(\theta, I, m), \theta) \frac{\partial}{\partial I} \hat{q}(\theta, I, m) + 1 &= 0 \\ R''(\hat{q}(\theta, I, m), \theta) \frac{\partial}{\partial m} \hat{q}(\theta, I, m) - 1 &= 0 \\ R'''(\hat{q}(\theta, I, m), \theta) \left(\frac{\partial}{\partial I} \hat{q}(\theta, I, m) \right)^2 + R''(\hat{q}(\theta, I, m), \theta) \frac{\partial^2}{\partial I^2} \hat{q}(\theta, I, m) &= 0 \\ R'''(\hat{q}(\theta, I, m), \theta) \frac{\partial}{\partial m} \hat{q}(\theta, I, m) \frac{\partial}{\partial I} \hat{q}(\theta, I, m) + R''(\hat{q}(\theta, I, m), \theta) \frac{\partial^2}{\partial m \partial I} \hat{q}(\theta, I, m) &= 0 \end{aligned} \tag{4}$$

which establish (b) and (c) ■

Proof of Lemma 5.2: We have:

$$\begin{aligned} \Pi(m_2, I) - \Pi(m_1, I) &= \int_{m_1}^{m_2} \frac{\partial}{\partial m} \Pi(m, I) dm \\ &= \int_{m_1}^{m_2} E \left[(R'(\hat{q}(\theta, I, m), \theta) - (c(\theta) - I)) \frac{\partial}{\partial m} \hat{q}(\theta, I, m) \right] dm \\ &= \int_{m_1}^{m_2} E \left[m \frac{\partial}{\partial m} \hat{q}(\theta, I, m) \right] dm \end{aligned} \tag{1}$$

which is negative from Lemma 5.1(b) ■

Proof of Lemma 5.3: For any discontinuity I there exist m_1 and m_2 such that $\hat{m}(I) = m_1$ and $\lim_{h \rightarrow 0} \hat{m}(I+h) = m_2 > m_1$. But then:

$$\begin{aligned} \lim_{h \rightarrow 0} \hat{\Pi}(I+h) &= \lim_{h \rightarrow 0} \Pi(\hat{m}(I+h), I+h) \\ &= \Pi(m_2, I) \\ &\leq \Pi(m_1, I) \\ &= \hat{\Pi}(I) \end{aligned} \tag{1}$$

where the inequality follows from Lemma 5.3. ■

Proof of Lemma 5.4: Differentiating (5.2) of the main text:

$$\begin{aligned}
 \hat{\Pi}'(I) &= E \left[(R'(\hat{q}(\theta, I, \hat{m}(I)), \theta) - (c(\theta) - I)) \left(\frac{\partial}{\partial I} \hat{q}(\theta, I, \hat{m}(I)) + \hat{m}'(I) \frac{\partial}{\partial m} \hat{q}(\theta, I, \hat{m}(I)) \right) \right] \\
 &\quad + E[\hat{q}(\theta, I, \hat{m}(I))] - v'(I) \\
 &= E \left[\hat{m}(I) \left(\frac{\partial}{\partial I} \hat{q}(\theta, I, \hat{m}(I)) + \hat{m}'(I) \frac{\partial}{\partial m} \hat{q}(\theta, I, \hat{m}(I)) \right) + \hat{q}(\theta, I, \hat{m}(I)) \right] - v'(I) \\
 &= E \left[\hat{m}(I) \hat{m}'(I) \frac{\partial}{\partial m} \hat{q}(\theta, I, \hat{m}(I)) + \hat{q}(\theta, I, \hat{m}(I)) \right] \\
 &= E \left[\hat{q}(\theta, I, \hat{m}(I)) - \hat{m}(I) \hat{m}'(I) \frac{\partial}{\partial I} \hat{q}(\theta, I, \hat{m}(I)) \right] \\
 &= E[\hat{q}(\theta, I, \hat{m}(I))] - v'(I) \hat{m}'(I)
 \end{aligned} \tag{1}$$

as required. ■

Proof of Proposition 4: It suffices to show that the investment is higher under negotiated transfer pricing. Let I_N denote the investment chosen under negotiated transfer pricing. From (3.2) we have:

$$I_N = \arg \max_I \frac{1}{2} E[q^*(\theta, I)] - v(I) \tag{1}$$

Let \hat{I} denote the investment chosen under actual cost-based transfer pricing. Thus,

$$\begin{aligned}
 \Pi_N(I_N) &\geq \Pi_N(\hat{I}) \\
 \hat{\Pi}(\hat{I}) &\geq \hat{\Pi}(I_N)
 \end{aligned} \tag{2}$$

Adding $1/\alpha$ times the first inequality to the second and rearranging:

$$1/\alpha \Pi_N(I_N) - 1/\alpha \Pi_N(\hat{I}) \geq \hat{\Pi}(I_N) - \hat{\Pi}(\hat{I}) \tag{3}$$

from which:

$$1/\alpha \int_{\hat{I}}^{I_N} \Pi'_N(I) dI \geq \int_{\hat{I}}^{I_N} \hat{\Pi}'(I) dI \tag{4}$$

Hence, it suffices to show that:

$$\begin{aligned}
 &1/\alpha \Pi'_N(I) \geq \hat{\Pi}'(I) \\
 \Leftrightarrow &E[q^*(\theta, I)] - 1/\alpha v'(I) \geq E[\hat{q}(\theta, I, \hat{m}(I))] - \hat{m}'(I) v'(I) \\
 \Leftrightarrow &E \left[\frac{q^*(\theta, I) - \hat{q}(\theta, I, \hat{m}(I))}{v'(I)} \right] + \hat{m}'(I) \geq 1/\alpha
 \end{aligned} \tag{5}$$

Claim 1. Each of the two terms in the left-hand side of the last inequality in (5) is at least 1.

Proof. From Lemma 5.1(a) we know that:

$$q^*(\theta, I) - \hat{q}(\theta, I, \hat{m}(I)) \geq \frac{\hat{m}(I)}{-R''(\hat{q}(\theta, I, \hat{m}(I)))} \tag{6}$$

and from the first-order condition for \hat{I} we have:

$$\begin{aligned}
 &\hat{m}(I) E \left[\frac{\partial}{\partial I} \hat{q}(\theta, I, \hat{m}(I)) \right] - v'(I) = 0 \\
 \Leftrightarrow &\hat{m}(I) E \left[\frac{1}{(-R''(\hat{q}(\theta, I, \hat{m}(I)), \theta))} \right] = v'(I)
 \end{aligned} \tag{7}$$

This establishes the claim for the first term. As for the second, note that:

$$\hat{I}'(m) = \frac{E \left[\frac{\partial}{\partial m} \hat{q}(\theta, \hat{I}(m), m) + m \frac{\partial^2}{\partial I \partial m} \hat{q}(\theta, \hat{I}(m), m) \right]}{-E \left[m \frac{\partial^2}{\partial I^2} \hat{q}(\theta, \hat{I}(m), m) - v''(\hat{I}(m)) \right]} \tag{8}$$

Simplifying:

$$\begin{aligned} \hat{I}'(m) &= \frac{E \left[\frac{\partial}{\partial m} \hat{q}(\theta, \hat{I}(m), m) + m \frac{R'''(\hat{q}(\theta, \hat{I}(m), m), \theta)}{(R''(\hat{q}(\theta, \hat{I}(m), m), \theta))^3} \right]}{-E \left[m \frac{R'''(\hat{q}(\theta, \hat{I}(m), m), \theta)}{(R''(\hat{q}(\theta, \hat{I}(m), m), \theta))^3} - v''(\hat{I}(m)) \right]} \\ &= \frac{E \left[\frac{\partial}{\partial m} \hat{q}(\theta, \hat{I}(m), m) \right] - mE \left[\frac{R'''(\hat{q}(\theta, \hat{I}(m), m), \theta)}{(R''(\hat{q}(\theta, \hat{I}(m), m), \theta))^3} \right]}{v''(\hat{I}(m)) - mE \left[\frac{R'''(\hat{q}(\theta, \hat{I}(m), m), \theta)}{(R''(\hat{q}(\theta, \hat{I}(m), m), \theta))^3} \right]} \end{aligned} \tag{9}$$

Since the common term in the numerator and denominator of (9) is positive, it suffices that

$E \left[\frac{\partial}{\partial I} \hat{q}(\theta, \hat{I}(m), m) \right] \leq v'(\hat{I}(m))$. From concavity of the first-best profit function, we know:

$$E \left[\frac{\partial}{\partial I} q^*(\theta, I) \right] - v''(I) \leq 0 \tag{10}$$

and from (A7) we know that for any I :

$$\frac{\partial}{\partial I} q^*(\theta, I) = \frac{1}{(-R''(q^*(\theta, I), \theta))} \geq \frac{\partial}{\partial I} \hat{q}(\theta, I, \hat{m}(I)) = \frac{1}{(-R''(\hat{q}(\theta, I, \hat{m}(I)), \theta))} \tag{11}$$

This completes the proof, since $\alpha \geq 1/2$ ■

Proof of Proposition 6: The investment induced by standard cost-based transfer pricing is:

$$I_s = \frac{\Delta}{2\gamma - 1} \tag{1}$$

The expected firm profit under standard cost-based transfer pricing is given by:

$$\begin{aligned} \hat{\Pi}_2(I_2, I_1, m) &= E_\theta \left[R(q_M(\theta, I_s), \theta) - q_M(\theta, I_s)(c(\theta) - I) \right] - V(I_s) \\ &= \frac{3}{16b} E_\theta \left[(\delta(\theta) + I_s)^2 \right] - \frac{1}{2} v I_s^2 \end{aligned} \tag{2}$$

Comparing this with the value of the ACTP profit function at $I = I_s$:

$$\begin{aligned} \hat{\Pi}(I_s) - \Pi_s &= \frac{1}{4b} E_\theta \left[(\delta(\theta) + I_s)^2 - (\gamma I_s)^2 \right] - \frac{3}{16b} E_\theta \left[(\delta(\theta) + I_s)^2 \right] \\ &\geq \frac{1}{16b} (\Delta + I_s)^2 - \frac{\gamma^2}{4b} I_s^2 \\ &= \frac{1}{4b} \left(\frac{\Delta + I_s}{2} + \gamma I_s \right) \left(\frac{\Delta + I_s}{2} - \gamma I_s \right) \\ &= 0 \end{aligned} \tag{3}$$

the last equality following from (1). This completes the proof. ■