MORAL HAZARD, MONITORING COST, AND THE CHOICE OF CONTRACTS*

Yoram Barzel
University of Washington

and

Wing Suen
University of Hong Kong

Revised
June 1992

* We wish to thank Gary Becker, Fahad Khalid, Jacques Lawarrée, Dean Lueck, Sherwin Rosen, the referee, and workshop participants at the University of Chicago and at the Hong Kong Baptist College for their comments.

MORAL HAZARD, MONITORING COST, AND THE CHOICE OF CONTRACTS

by

Yoram Barzel
University of Washington

Wing Suen
University of Hong Kong

ABSTRACT. When individuals' contributions to team production are costly to measure, the incentives for different team members to supply and monitor inputs will depend on the assignment of residual claims to team output. In an optimal employment contract, individuals with relatively low marginal product will become employees because the efficiency loss will be relatively small as they shirk. In a share contract, each member of the team receives a partial residual claim and the optimal amount of monitoring is reduced. Like employment contracts, but not as extremely, an optimal share contract will allocate a relatively small marginal share of team output to individuals with relatively low marginal product.

KEYWORDS. Employment contract, monitoring cost, residual claimancy, share contract, shirking, team production.

JEL CLASSIFICATION NUMBER. D23

Productive activities often involves the cooperation of more than one input owner: clinics are run by doctors and nurses, and law firms consist of attorneys and secretaries. In neoclassical economics, it does not matter whether doctors hire nurses or nurses hire doctors. Yet we often observe that it is the more productive workers (doctors, attorneys) who are hiring the less productive ones (nurses, secretaries). What determines the kind of contracts governing the relationship between cooperating input owners?

Contracts allocate rewards and penalties to cooperating input owners. Those input owners whose rewards are fixed with respect to output can gain by supplying less than the optimal amount of inputs. Residual claimants, on the other hand, not only lack such an incentive to shirk, but will also monitor cooperating input owners to prevent them from shirking. Given such a reward structure, the amount of inputs to be stipulated in a contract will depend on the relative costs of monitoring and of shirking. The structure of rewards is not fixed, however. Reversing the status of those input owners who are employees and of those who are residual claimants will reverse their pattern of behavior as well. The choice of contracts, therefore, also involves a global comparison of the (maximum) net gains under different reward structures. While individual resource owners will maximize their personal gains subject to contractual terms, market competition ensures that the contractual terms will maximize the joint value of the transaction net of the costs resulting from monitoring and shirking. As will be shown below, contracts are structured in such a way that the owners of resources that are relatively more productive and more difficult to monitor will become residual claimants. This may help explain, for example, why doctors hire nurses instead of the reverse.

Residual claimancy need not be complete; in particular, every resource owner may receive a share of the output such that each becomes a partial residual claimant. Since output is then shared, individuals do not receive their full contribution to production as their reward, and each will shirk to a certain extent in supplying input. The contract will specify shares that balance the marginal cost of shirking by each input owner. This paper demonstrates that, under a share contract, the higher the marginal product of an input, the greater the marginal share of output the input owner will receive. Share contracts and employment contracts are also compared in order to illustrate the trade-offs between constraining moral hazard through residual claimancy and constraining it through monitoring.

The theme of this article builds on the important works by Alchian and Demsetz (1972) and by Holmstrom (1982). Both papers emphasize the importance of moral hazard in team production. Alchian and Demsetz are concerned with the relationship between monitoring and residual claimancy, whereas Holmstrom concentrates on the principal's role in enforcing contracts. Neither paper, however, explicitly discusses the *cost* of monitoring. By introducing monitoring cost into the model, it is possible to give a much clearer picture of the trade-offs involved in the choice of contracts.

This paper is also related to the literature on sharecropping.[1] Some of those who have written about the subject (e.g., Stiglitz 1974; Hallagan 1978) have assumed that the inputs of laborers are difficult to measure, while those of landowners are not. The asymmetry in this assumption is unattractive. "Land" is a resource possessing many attributes, the amounts of which are variable and costly to measure. The productivity of land depends, in part, on drainage and on the amount of nutrients in the soil. Because a parcel of land is seldom uniform in its various characteristics the accurate measurement of land is very costly. Moreover, landowners often provide other inputs (such as farming equipment and marketing service) which are costly to monitor. Farming activities should therefore be regarded as team production, and it is appropriate to analyze them using the theory of contract choice presented here. Our analysis of the choice of contracts is similar to that in Eswaran and Kotwal (1985). Eswaran and Kotwal's model contains a cost of non-specializing which arises because difficult-to-measure inputs will not be exchanged under

[1] See especially Cheung (1969), Stiglitz (1974), Hallagan (1978), Eswaran and Kotwal (1985). Newbery and Stiglitz (1979) provide a brief and useful overview of the theories of sharecropping.

a fixed-rent or fixed-wage contract. The cost of non-specializing in their model plays a role equivalent to that of monitoring cost in the present model.[2] Unlike the numerical simulations of their model, however, the results of this paper are derived analytically.

The analysis of contracts presented here does not assume risk aversion. Cheung (1969) argues that the share contract requires a higher monitoring cost than does the wage or rental contract, but that it has the advantage of sharing risks. Here we argue that the share contract *saves* on monitoring cost. Under a share contract, instead of constraining moral hazard through monitoring, each team member is given a partial residual claim, which reduces the need for monitoring. This emphasis on monitoring cost also extends principal-agent theory, in which the status of principal and agent is determined outside the model.[3] Once it is recognized that the use of monitoring and residual claimancy are determined endogenously to constrain moral hazard, the implications of differences in productivity and monitoring costs for the choice of contracts can be readily obtained. The results we obtain complement Grossman and Hart (1986) and Hart and Moore (1990). In those models the allocation of residual rights of control affects prior investments in specific assets, and relative inefficiencies in investments determine ownership rights. In our paper the allocation of residual claimancy affects current input supplies, and traditional welfare loss comparisons determine contractual forms.

I. Moral Hazard in Team Production

We use the term "team production" to refer to production activities in which inputs are provided by more than one person. The value of team output is a function of the level of inputs provided by each team member. In this and the following three sections it is assumed that the team consists of two persons, A and B. Individual A provides an input

[2] Both monitoring cost and the cost of non-specializing are likely to affect the choice of contracts. We use the former because its interpretation is more straightforward.

[3] See, for example, Stiglitz (1974), Holmstrom (1979), and Grossman and Hart (1983).

a at a cost $c_a(a)$ and B provides input b at a cost $c_b(b)$, with $c_i' > 0, c_i'' > 0$. For simplicity the value of team output, x, is modeled as a linear function of the inputs:[4]

$$x = \alpha a + \beta b - F + e. \tag{1}$$

The advantage of team production arises from the savings in fixed cost, F. The term e represents the random element in production. We assume individuals are risk neutral, and in what follows only the expected values matter.

The first-best optimal level of inputs, a^* and b^*, is obtained by maximizing the expected value of $x - c_a(a) - c_b(b)$, which gives

$$\begin{aligned} \alpha &= c_a'(a^*), \\ \beta &= c_b'(b^*). \end{aligned} \tag{2}$$

Both individuals know their own contributions, but because of the random element in production neither of them can infer the contribution of the other. If it were costless to measure or monitor the level of inputs, the first-best outcome could be achieved by a price mechanism, or by directly instructing the input owners to supply a^* and b^*. When monitoring is not free, it will be applied only to the point where the marginal gain from reduced shirking equals the marginal cost of monitoring. The net value of the team will be less than the first-best optimum because resources are spent on monitoring and because some shirking remains. Alternatively, the team members may forgo the measurement of inputs altogether and structure their rewards on the basis of output (which is assumed to be costless to measure). Since the value of output has to be shared between team members, however, each will shirk partially on input supply, and it is again impossible to achieve a first-best outcome (Holmstrom 1982). The choice of contracts, then, must be a choice among second-best outcomes.[5]

[4] This production function will be generalized in Section V.

[5] Holmstrom proposes a solution to the moral hazard problem by introducing a "princi-

II. The Nature of Monitoring Cost

Other things being equal, an employed worker is expected to exert more effort as the employer spends more on monitoring. If we write the worker's input supply level, b, as a function of monitoring cost, m, then $b = b(m)$ and $b(\cdot)$ will be an increasing function. Inverting the function $b(\cdot)$ gives the monitoring cost function, $m = m(b)$, which is increasing in b. While the detailed specification of monitoring cost will depend on the monitoring technology and is expected to differ across transactions, we believe a reasonable restriction is that monitoring cost increases with the stipulated input level.[6] Since the cost of supplying inputs is assumed to be increasing and convex, both the total and the marginal incentive to shirk increases as the employer stipulates that the worker supplies more inputs. To combat this increased tendency to shirk would require more expenses on monitoring the worker's behavior.

As an example, suppose employer A monitors worker B. A pays wb to B for b units of input. When B supplies less than b units, there is a probability $\pi(m)$ that this will be found out and B will then have to pay a penalty δ.[7] The probability of detection is assumed to be increasing in the monitoring expenditure, m. Given this situation,[8] the

pal" to break the budget constraint. However, as Arrow (1985) points out, there are many Nash equilibria in Holmstrom's scheme which are inefficient. The existence of multiple Nash equilibria provides opportunity for strategic manipulations by team members. A related difficulty of the proposed scheme is that the principal will have a strong incentive to reduce output, say, by colluding with one of the team members (see Eswaran and Kotwal 1984) or by falsifying the measured output. In what follows, the possibility of an outside principal is ruled out, and attention will be focused solely on contracts that satisfy the budget constraint.

[6] This is the only restriction on the monitoring cost function that we need in the rest of the paper.

[7] The penalty is constrained by wealth. If there were no wealth constraint, monitoring cost could be made arbitrarily small by increasing the punishment to infinity. The penalty to shirking may be viewed as a bond to guarantee performance. Since the provision of guaranteeing capital is costly, some positive level of monitoring will still take place. Also, if one party posted a large bond, the other might take the money and run.

[8] A more elaborate version of this model in which shirking is not all-or-nothing is described in Calvo and Wellisz (1978).

worker will supply b units of inputs if

$$wb - c_b(b) \geq wb - \pi(m)\delta - c_b(0); \tag{3}$$

that is,

$$\pi(m)\delta \geq c_b(b) - c_b(0). \tag{4}$$

As the stipulated input level b increases, the gains from shirking (i.e., the right-hand side of (4)) increases and m must rise to preserve the inequality. This will result in a monitoring cost function that satisfies our restriction.

As another example, suppose the employer uses a one-tailed test and imposes the penalty δ whenever output falls below some critical level. In a variety of situations, this strategy or variants thereof will be an optimal monitoring policy.[9] The probability of detection will then be a function of the deviation of the actual from the stipulated input level, $\pi = \pi(m, b-\tilde{b})$.[10] Figure 1 depicts the typical behavior of π. The curve corresponding to higher monitoring expenditure (m_2) gives a more powerful one-tailed test. The worker takes m as given. The employer is a Stackelberg leader in this monitoring game. We assume the employer can credibly precommit to the monitoring policy (through some reputation mechanism, for instance) and chooses $\tilde{b}$ to maximize

$$wb - \pi(m, b - \tilde{b})\delta - c_b(\tilde{b}). \tag{5}$$

If the employer wants to implement $\tilde{b} = b$, monitoring cost m must be chosen such that

$$\pi_2(m, 0)\delta = c_b'(b). \tag{6}$$

[9] See, for example, Dye (1986) and Rasmusen and Zenger (1990). In Dye's model, however, monitoring occurs only after output becomes known.

[10] We can also let the probability of detection be a function of the percentage deviation of the actual input level from the stipulated level. The results will be essentially unchanged.

When the stipulated input level b increases, the marginal gain from shirking (i.e., the right-hand side of (6)) increases and $\pi_2(m, 0)$ must be raised to preserve the equality. A higher $\pi_2(m, 0)$ requires a more powerful test (see Figure 1). Consequently, monitoring cost must increase with the stipulated input level.

The paper assumes that the cost of monitoring is independent of the productivity of the employer. This will be the case if monitoring ability and direct productive ability are uncorrelated. The employer's allocation of time between production and monitoring can be explicitly modeled as in Barzel (1987). Introducing this feature in our paper will increase its realism but not alter its conclusions. Another possibility is that monitoring can be acquired from the market (e.g., equipments for quality control, professional auditors). In that case monitoring cost will depend on the market price of these monitoring inputs and will be independent of the employer's own productivity.[11]

III. Employment Contracts and Residual Claimancy

In an employment contract, an employee whose pay is fixed (with respect to output) will gain by supplying less than the optimal amount of input. Monitoring which reduces the extent of shirking will be adopted if the gain exceeds the cost. The residual claimant, on the other hand, has no incentive to shirk, and it is thus unnecessary to monitor his or her input. Since the costs of monitoring and of shirking differ across team members, it *does* matter whether "capital hires labor" or "labor hires capital."[12] The choice of employment contracts, then, raises the question: Who will be the residual claimant?

Suppose, under contract 1, person A is the employer and person B the employee. To prevent B from shirking, A will spend resources to measure and monitor the level of input

[11] The use of third parties, however, raises the issue of the control of monitors (Alchian and Demsetz 1972; Eswaran and Kotwal 1984; Baiman et al., 1987). This problem will be moot if the employer undertakes supervision directly (Barzel 1987, Calvo and Wellisz 1978).

[12] The usual contention (e.g., Samuelson 1957) that in a competitive market it does not matter who hires whom implicitly assumes costless monitoring.

b. The cost of monitoring, $m_b(b)$, is an increasing function in the specified input level, as we argue in the previous section. The optimization problem for A is:

$$\max_{a,b,w_b} \quad \alpha a + \beta b - F - c_a(a) - m_b(b) - w_b b \qquad \text{subject to} \quad w_b b - c_b(b) \geq u_b, \tag{7}$$

where w_b and u_b refers to B's wage rate and reservation utility. The solution values of a and b, denoted a_1 and b_1, satisfy

$$\alpha = c_a'(a_1), \qquad \beta - m_b'(b_1) = c_b'(b_1). \tag{8}$$

Comparing (2) to (8), it is evident that $a_1 = a^*$ and $b_1 < b^*$. That is, the residual claimant will provide the optimal level of input, whereas the employee will shirk even though he or she is being monitored.[13] The total costs of transaction under this contract are the sum of the resource cost spent on monitoring, $m_b(b_1)$, and the deadweight cost of shirking, $(\beta b^* - c_b(b^*)) - (\beta b_1 - c_b(b_1))$. This definition of transaction costs is synonymous to Jensen and Meckling's (1976) concept of "agency costs."

The costs of transaction depend on the form of the contract governing the relationship between A and B. Instead of A hiring B, it is possible to choose another contract in which B hires A. Under the latter arrangement (contract 2), B's choice problem is

$$\max_{a,b,w_a} \quad \alpha a + \beta b - F - c_b(b) - m_a(a) - w_a a \qquad \text{subject to} \quad w_a a - c_a(a) \geq u_a. \tag{9}$$

The solution values of a and b, denoted a_2 and b_2, satisfy

$$\alpha - m_a'(a_2) = c_a'(a_2), \qquad \beta = c_b'(b_2). \tag{10}$$

[13] In this article "shirking" always means that less than the optimal level of input is supplied. Strictly speaking, shirking does not involve any cheating or contractual non-performance; shirking occurs because transactors rationally contract for a less-than-optimal level of input to economize on monitoring cost.

For this contract, $a_2 < a^*$ and $b_2 = b^*$. The transaction costs are $(\alpha a^* - c_a(a^*)) - (\alpha a_2 - c_a(a_2)) + m_a(a_2)$.

While the choice of input levels is constrained by the presence of monitoring cost, the choice of contractual forms is not. As long as individuals have full knowledge of the contractual terms before they accept or reject the contract, Coase's theorem applies: Contracts will be chosen so as to maximize the net value of the team or, equivalently, to minimize the costs of transactions. We let V_1 and V_2 be the net value of the team under contract 1 and contract 2 respectively where

$$V_1 = \alpha a_1 + \beta b_1 - F - c_a(a_1) - c_b(b_1) - m_b(b_1) \tag{11}$$

and

$$V_2 = \alpha a_2 + \beta b_2 - F - c_a(a_2) - c_b(b_2) - m_a(a_2). \tag{12}$$

Since V_1 and V_2 are functions of α and β, the relative advantage of contract 1 and contract 2 will differ according to differences in the marginal products of A and B. Specifically, consider the effects of changes in α. As α approaches zero, a^* also approaches zero so that transaction costs under contract 2 are negligible. Transaction costs under contract 1, on the other hand, are positive because monitoring and shirking will take place. Therefore, $V_1 < V_2$. As α increases, the costs of transaction under contract 1 (with A as the employer) remain unchanged but transaction costs under contract 2 (with A as the employee) will increase. Thus, $V_1 > V_2$ will exceed V_2 when α gets very large. Moreover, from the envelope theorem,

$$\begin{aligned} \frac{\partial V_1}{\partial \alpha} = a_1, \quad \frac{\partial^2 V_1}{\partial \alpha^2} = \frac{1}{c_a''}; \\ \frac{\partial V_2}{\partial \alpha} = a_2, \quad \frac{\partial^2 V_2}{\partial \alpha^2} = \frac{1}{c_a'' + m_a''}. \end{aligned} \tag{13}$$

Both V_1 and V_2 are increasing and convex in α, with the slope of V_1 being greater than that of V_2 (since $a_1 > a_2$). The relationship between V_1, V_2 and α is depicted in Figure 2. For $\alpha < \overline{\alpha}, V_2 > V_1$ and B willbe the employer. For $\alpha > \overline{\alpha}, V_1 > V_2$ and A will be the employer. The higher is A's marginal product relative to B's, the greater will be the cost imposed by shirking on the part of person A compared to shirking on the part of person B. As the marginal product of A increases beyond a critical value, transaction costs can be reduced by making A the residual claimant.[14]

A similar exercise can be performed for the effect of monitoring cost on the choice of contracts. Suppose there is a shift parameter which changes the magnitude of m'_a. Since contract 1 does not monitor A's input, V_1 is independent of m'_a. The value of V_2, on the other hand, will decrease (at a decreasing rate) as m'_a increases. Figure 3 shows a possible case where V_1 and V_2 intersect.[15] For this case, contract 1 will be chosen over contract 2 whenever $m'_a > \overline{m'_a}$. That is, the more costly it is to monitor A's input, the more likely it is that A will become the residual claimant.

We have, then, partly answered the question posed at the beginning of this section: The greater is one's ability to affect team output in costly-to-monitor ways, the more likely it is that one will be the residual claimant and will monitor the team. Note that this analysis is equally applicable to vertical production relationships. For example, it is expected that the upstream producer will acquire control over the downstream producer rather than the other way around if the former has a greater ability to affect the value of the final product in difficult-to-measure ways.[16]

Employer-employee relationships in the professions illustrate our model. In various professions such as law and medicine practitioners often operate in groups. As a rule

---

[14] This comparative statics result is based on the multiplicative shift parameter α. Changes in total product without corresponding changes in marginal product will not affect contract choice.

[15] The other possible case is that V_2 lies entirely above V_1.

[16] The distinction between forward and backward integration would be meaningless if monitoring cost were zero. Cf. Grossman and Hart (1986) and Hart (1988).

the more experienced members of such groups employ the less experienced ones. Since experience enhances productivity, it is the more productive team members who employ and monitor the less productive members as predicted by our model.

IV. Optimal Share Contracts

An employment contract gives a marginal share of zero percent of team output to the employee, and a marginal share of one hundred percent to the employer. It is also possible to allocate a positive marginal share to the employee by reducing the marginal share of the employer. The resulting contract is a share contract. Of course, under a share contract, the distinction between employer and employee is no longer meaningful. For our purposes, any contract in which payment to each team member depends on team output constitutes a share contract. Thus an employment contract with bonus payments is a share contract because employees receive a positive marginal share of team output (Hashimoto 1979). Another example of a share contract, although seldom recognized as such, is a piece-rate contract. When an owner of capital hires a worker and pays him or her on the basis of the number of units produced, the piece rate that the worker receives must be less than the price of the output, because capital which contributes to team output must also be compensated. Thus, piece-rate contracts involve value sharing, just as sharecropping involves output sharing.[17] The analysis that follows is applicable to both.

Under a share contract, since each worker is a partial residual claimant, each will voluntarily supply some input even if he or she is not being directly monitored. To focus on the role of the share contract as an alternative method to constrain moral hazard, we first assume there were no monitoring in a share contract. This assumption is relaxed at the end of this section.

Suppose, in team production, person A will receive $\lambda x + k$ and person B will receive

[17] To our knowledge, Hall (1984) is the first to clearly point this out. Such an interpretation of bonus and piece-rate payments, to be sure, presumes team production.

$(1-\lambda)x-k$, where λ and $1-\lambda$ are the marginal output shares and the constant k represents a side payment to satisfy competitive conditions. Since the parties do not monitor each other, the levels of inputs a and b will be chosen in a non-cooperative manner. The ~~design~~ decision problems for A and B are, respectively:

$$\begin{aligned} &\max_a \quad \lambda(\alpha a + \beta b - F) - c_a(a) + k, \\ &\max_b \quad (1-\lambda)(\alpha a + \beta b - F) - c_b(b) - k. \end{aligned} \tag{14}$$

The Nash equilibrium solution (a_3, b_3) is characterized by

$$\begin{aligned} \lambda\alpha &= c_a'(a_3), \\ (1-\lambda)\beta &= c_b'(b_3). \end{aligned} \tag{15}$$

Since A and B each receives only a fraction of his or her marginal product, each will shirk partially on input supply (i.e., $a_3 < a^*, b_3 < b^*$). Moreover, the extent of shirking depends on the marginal share that each receives. From (15),

$$\begin{aligned} \frac{\partial a_3}{\partial \lambda} &= \frac{\alpha}{c_a''} > 0, \\ \frac{\partial b_3}{\partial \lambda} &= \frac{-\beta}{c_b''} < 0. \end{aligned} \tag{16}$$

As λ increases, person A becomes more of a residual claimant, thereby reducing his or her incentive to shirk; at the same time, shirking on the part of person B becomes more severe.

The optimal share contract specifies a λ that maximizes the net value of the team. This problem can be written as

$$\max_\lambda \quad \alpha a_3(\lambda) + \beta b_3(\lambda) - F - c_a(a_3(\lambda)) - c_b(b_3(\lambda)). \tag{17}$$

The first-order condition is

$$(\alpha - c_a')\frac{\partial a_3}{\partial \lambda} + (\beta - c_b')\frac{\partial b_3}{\partial \lambda} = 0. \tag{18}$$

That is, the change in A's input supply, multiplied by A's net marginal value product, must be balanced against the change in B's input supply, multiplied by B's net marginal value product. Using (16), condition (18) can be rewritten as

$$\frac{\alpha(\alpha - c_a')}{c_a''} = \frac{\beta(\beta - c_b')}{c_b''}. \qquad (19)$$

If there is a shift parameter that reduces c_a'', the value of c_a' must be increased to satisfy the first-order condition. Since c_a is convex, c_a' is large when a_3 is large, which is achieved by choosing a high λ (and the appropriate side-payment k). When A shirks, the deadweight cost will be relatively high if A's input supply curve is very elastic. Transaction costs can be reduced by allocating a greater fraction of the residual claim to A so that A's incentive to shirk will be smaller.[18] Similarly, when α is large, the optimal λ will be high and A will become more of a residual claimant. This conclusion is similar to that obtained in the previous section—the greater one's ability to affect the output of the team, the greater the extent to which one's income will depend on team output.[19]

One of the results from the above analysis is that a contract in which λ equals one is never optimal. If $\lambda = 1$, then $a_3 = a^*$ and $b_3 = 0$.[20] An infinitesimal decrease in λ will decrease a_3 and increase b_3. By the envelope theorem, the cost due to the decrease in a_3 is of second order, whereas the benefit due to the increase in b_3 is of first order.[21] Thus, an employment contract which allocates a one-hundred percent share to one residual claimant

[18] The elasticity of the marginal cost curve of an input discussed here seems to correspond to Alchian and Woodward's (1987) notion of "plasticity." Resources with many alternative uses will have elastic input supply curves and will be vulnerable to moral hazard.

[19] It should be recognized that this is a conclusion about the relationship between *marginal* product and *marginal* share, not between average product and average share.

[20] The fact that $b_3 = 0$ does not imply person B will be excluded from the team, because b is just one attribute of B's input. B will still supply positive amounts of other attributes if they are not costly to measure.

[21] The fractions $1 - \lambda$ and λ may be regarded as the tax rates on input supplies a and b. Since the welfare cost of taxation increases with the square of the tax rate, taxing both inputs at a positive tax rate is preferred to taxing one input at a 100 percent tax rate.

and which does not monitor employees will be strictly dominated by a share contract.

Figure 4 compares share contracts to employment contracts. In this figure, curve V_E is just the upper envelope of curves V_1 and V_2 shown earlier. That is, $V_E = max\{V_1, V_2\}$ is the value of the employment contract with endogenously determined employer/employee status. Curve V_S shows the net value of team output under the optimal share contract:

$$V_S = \alpha a_3 + \beta b_3 - F - c_a(a_3) - c_b(b_3), \tag{20}$$

where a_3 and b_3 satisfy equation (19). Consider how V_S and V_E vary with α (the marginal product of person A).[22] If α is very large, the deadweight cost of having an inefficient level of a will be very large. Since $a_1 = a^*$ under the employment contract while $a_3 < a^*$ under the share contract, we have $V_E = V_1 > V_S$. If α is very small, on the other hand, the deadweight cost from an inefficient level of b will be the dominant consideration. As $b_2 = b^*$ and $b_3 < b^*$, we then have $V_E = V_2 > V_S$. Thus the share contract will dominate employment contracts only for inrmediate values of α, as shown in Figure 4.[23] When the marginal product of A is low, person A will be employed and be monitored by B (contract 2). For intermediate values of α, the team will adopt a share contract so that person A becomes a partial residual claimant, with a marginal share that increases with A's productivity. As A's marginal share increases, however, B's incentive to shirk also increases. At a certain point, transaction costs can be reduced by switching from the share contract to an employment contract (contract 1), with A being the full residual claimant and actively monitoring the employee B.[24]

[22] As drawn in the figure, V_S is increasing and convex in α. Proof: $\partial V_S/\partial\alpha = a_3 > 0, \partial^2 V_S/\partial\alpha^2 = \partial a_3/\partial\alpha > 0$.

[23] It is also possible that V_S will lie entirely below V_E. This occurs when monitoring is relatively cheap.

[24] The screening hypothesis (Hallagan 1978; Newbery and Stiglitz 1979) also suggests that, in a cross section, farmers working on rental contracts tend to be more productive than those working on share contracts, who in turn are more productive than fixed-wage farm laborers. The hypothesis, however, cannot explain changes in contract forms over

Our analysis has focused on the use of partial residual claimancy as a substitute for direct monitoring. The share contract can be improved if there is simultaneous application of monitoring and output-based compensation. For example, suppose the monitoring technology is as in Section II. When person B supplies less than b_0 units of input, there is a probability $\pi(m)$ that shirking will be detected and a penalty δ will be paid. If B's marginal share is $(1-\lambda)$ and b_λ is B's input supply in the absence of monitoring, person B will choose b_0 instead of b_λ if and only if

$$\pi(m)\delta \geq [(1-\lambda)\beta b_\lambda - c_b(b_\lambda)] - [(1-\lambda)\beta b_0 - c_b(b_0)]. \tag{21}$$

For small $b_0 - b_\lambda$, the term in the first brackets on the right-hand side of (21) will be of second order (the envelope theorem). Provided $\pi'(m) > 0$ at $m = 0$, the level of b then can be increased at a marginal cost which is second-order small. The increase in value of team output from a marginal increase in b, on the other hand, is of first-order magnitude because $b_\lambda < b^*$. Thus introducing monitoring to a share contract will strictly raise its value. In terms of the diagram in Figure 4, monitoring will shift the V_S curve up. This makes the share contract more attractive than the employment contract over a wider range of α. Our qualitative conclusions, however, remain the same.

V. Some Extensions

Up to this point, a two-person, two-input model has been used to analyze the forces affecting the choice of contracts. Since team production often involves more than two people, and since each person's contribution to the team has many different attributes, it will be useful to extend the discussion by considering a many-agent, many-attribute model.

Suppose a team consists of n people, each supplying an attribute a_i, which is costly to measure, and another attribute z_i which is costless to measure. The cost to individual i is

---

time because it is only concerned with the sorting of heterogeneous farm laborers by the use of different contracts.

$c^i(a_i, z_i)$ and team output is $x = f(a_1, \ldots, a_n, z_1, \ldots, z_n) + e$, where $c^i(\cdot)$ and $f(\cdot)$ satisfy the usual assumptions about cost and production functions. The analysis of employment contracts with one residual claimant and $n-1$ employees is the same as the two-person case, and therefore will be omitted. The discussion will focus on share contracts.

In a share contract, individual i with marginal share λ_i will choose a_i to maximize $\lambda_i x - c^i$. This implies

$$\lambda_i f_{a_i} - c^i_{a_i} = 0, \quad i = 1, \ldots, n. \tag{22}$$

On the other hand, the z's (which are costless to measure) and the λ's will be chosen to maximize the net value of the team, with

$$\sum_{i=1}^{n} \lambda_i = 1. \tag{23}$$

The optimal share contract solves

$$\max_{a,z,\lambda} \quad f(a,z) - \sum_{i=1}^{n} c^i(a_i, z_i) \tag{24}$$

subject to (22) and (23). Let $\theta_1, \ldots, \theta_n$ be the Lagrange multipliers for (22) and μ be the Lagrange multiplier for (23). The first-order conditions to this problem are:

$$f_{a_i} - c^i_{a_i} + \sum_j \theta_i(\lambda_j f_{a_j a_j} - c^j_{a_j a_i}) = 0, \tag{25}$$

$$f_{z_i} - c^i_{z_i} + \sum_j \theta_j(\lambda_j f_{a_j z_i} - c^j_{a_j z_i}) = 0, \tag{26}$$

$$\theta_i f_{a_i} - \mu = 0. \tag{27}$$

Using (27), equation (25) can be written as

$$\frac{(f_{a_i} - c^i_{a_i})}{\sum_j(\lambda_j f_{a_i a_i} - c^j_{a_j a_i})/f_{a_j}} = -\mu, \quad i = 1, \ldots, n. \tag{28}$$

Clearly, equation (28) is a generalization of equation (19). Comparative statics results for λ_i are straightforward. The greater is the net marginal value of a_i, and the more elastic are its input supply and input demand curves, the greater will be the marginal share for person i.

It should also be noted that the level of z_i under a share contract will be different from the first-best level, even though z_i is costless to measure. Specifically, the first-best level of z_i is given by

$$f_{z_i} - c^i_{z_i} = 0. \tag{29}$$

Comparing (26) to (29), the level of z_i under a share contract will be greater than the first-best level if $\sum_j \theta_j(\lambda_j f_{a_j z_i} - c^j_{a_j z_i})$ is positive. Thus, if z_i and the a_i's are complementary in the sense that on average an increase in z_i will increase the marginal product or reduce the marginal cost of the a_i's, the contract will specify an "excessive" use of z_i (or specify an "excessively" low price for it) in order to induce people to supply more of the costly-to-measure attributes.[25]

A final observation relates to the size of a team. As team size increases, each person's marginal share will decrease and each will shirk more. On the other hand, the extent of shirking under an employment contract depends only on monitoring cost, which is independent of team size. In a large team, a pure share contract in which everybody gets a positive marginal share and nobody is monitored will be relatively inefficient. It is expected that only those few people with the greatest ability to affect team output in difficult-to-measure ways will be residual claimants, while the majority of the team members will

[25] Braverman and Stiglitz's (1982) discussion of interlinked agrarian markets is based on this idea. See also Barzel (1977; 1981) for applications to slavery and tying arrangements.

become employees and be monitored.[26]

VI. Concluding Remarks

Moral hazard in teams can be constrained by monitoring the behavior of team members, or by aligning their rewards more closely with team output. However, monitoring is not without cost, and full residual claimancy cannot be given to every team member simultaneously. While a first-best outcome is not achievable, contracts will be chosen such that the value of team output net of the transaction costs resulting from moral hazard and from monitoring is maximized. The analysis performed in this paper demonstrates a general proposition: The greater is one's ability to affect team output in difficult-to-measure ways, the greater is the extent to which one's income will depend on team output, and the less likely that one will be directly monitored.

This article has focused on a team's cost of measuring inputs. It should be recognized that the "output" of a team is not just a simple dollar value. Team production generates outcomes with many different attributes, many of which are costly to measure. For example, the outputs of a farm include not only diverse farm products which differ in quality, but also soil conservation or deterioration, and human capital formation through on-the-job training. When the outputs of a team have many attributes, contracts can reward team members in more detailed ways so that costs and benefits are more closely aligned. At the same time, the costliness of measuring outputs introduces the additional moral hazard problem of output skimming or stealing (see Umbeck 1977; Hart 1988). A full analysis of the choice of contracts must take these considerations into account.

[26] To be sure this consideration has to be balanced against individuals' ability to provide capital and bear risks.

REFERENCES

Alchian, Armen A. and Demsetz, Harold, "Production, Information Costs, and Economic Organization," *American Economic Review*. December 1972, 62, 777-95.

Alchian, Armen A. and Woodward, Susan, "Reflections on the Theory of the Firm," *Journal of Institutional and Theoretical Economics*, March 1987, 143, 110-36.

Arrow, Kenneth J., "The Economics of Agency," in J. Pratt and R. Zeckhauser, eds., *Principals and Agents: The Structure of Business*. Boston: Harvard Business School Press, 1985.

Baiman, S., Evans, J.H. and Noel, J. "Optimal Contracts with a Utility-Maximizing Auditor," *Journal of Accounting Research*, Autumn 1987, 25, 217-44.

Barzel, Yoram, "An Economic Analysis of Slavery," *Journal of Law and Economics*, April 1977, 20, 87-110.

Barzel, Yoram, "Competitive Tying Arrangements: The Case of Medical Insurance," *Economic Inquiry*, October 1981, 19, 598-612.

Barzel, Yoram, "The Entrepreneur's Reward for Self-Policing," *Economic Inquiry*, January 1987, 25, 103-16.

Braverman, Avishay and Stiglitz, Joseph E., "Sharecropping and the Interlinking of Agrarian Markets," *American Economic Review*, September 1982, 72, 695-715.

Calvo, G.A., and Wellisz, S., "Supervision, Loss of Control, and the Optimal Size of the Firm," *Journal of Political Economy*, October 1978, 86, 943-52.

Cheung, Steven N.S., *The Theory of Share Tenancy*. Chicago: University of Chicago Press, 1969.

Dye, Ronald A. "Optimal Monitoring Policies in Agencies." *Rand Journal of Economics*, Autumn 1986, 17, 339-50.

Eswaran, Mukesh and Kotwal, Ashok, "The Moral Hazard of Budget Breaking," *Rand Journal of Economics*, Winter 1984, 15, 578-81.

Eswaran, Mukesh and Kotwal, Ashok, "A Theory of Contractual Structure in Agriculture," *American Economic Review*, June 1985, 75, 352-67.

Grossman, Sanford J. and Hart, Oliver D., "An Analysis of the Principle-Agent Problem," *Econometrica*, January 1983, 51, 7- 45.

Grossman, Sanford J. and Hart, Oliver D., "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, August 1986, 94, 691-719.

Hall, Christopher D., *Topics in Microeconomics.* Hong Kong: Academy Press, 1984.

Hallagan, William, "Self-Selection by Contractual Choice and the Theory of Sharecropping," *Bell Journal of Economics*, Autumn 1978, 9, 344- 54.

Hart, Oliver D., "Incomplete Contracts and the Theory of the Firm," *Journal of Law, Economics, and Organization*, Spring 1988, 4, 119-39.

Hart, Oliver D. and Moore, John, "Property Rights and the Nature of the Firm," *Journal of Political Economy*, December 1990, 98, 1119-58.

Hashimoto, Masanori, "Bonus Payments, On-The-Job Training, and Lifetime Employment in Japan," *Journal of Political Economy*, October 1979, 87, 1086-104.

Holmstrom, Bengt, "Moral Hazard and Observability," *Bell Journal of Economics*, Spring 1979, 10, 74-91.

Holmstrom, Bengt, "Moral Hazard in Teams," *Bell Journal of Economics*, Autumn 1982, 13, 324-40.

Jensen, Michael and Meckling, William, "Theory of the Firm: Managerial Behavior, Agency Costs, and Capital Structure," *Journal of Financial Economics*, 1976, 3, 305-60.

Newbery, David M.G. and Stiglitz, Joseph E., "Sharecropping, Risk Sharing and the Importance of Imperfect Information," in J. Roumasset *et al.*, eds., *Risk, Uncertainty and Agricultural Development.* New York: Agricultural Development Council, 1979.

Rasmusen, Eric and Zenger, Todd, "Diseconomies of Scale in Employment Contracts," *Journal of Law, Economics, and Organization*, Spring 1990, 6, 65-92.

Samuelson, Paul A., "Wage and Interest: A Modern Dissection of Marxian Economic Models," *American Economic Review*, December 1957, 47, 884-912.

Stiglitz, Joseph E., "Incentive and Risk Sharing in Sharecropping," *Review of Economic Studies*, April 1974, 41, 219-56.

Umbeck, John, "A Theory of Contract Choice and the California Gold Rush," *Journal of Law and Economics*, October 1977, 20, 421-37.

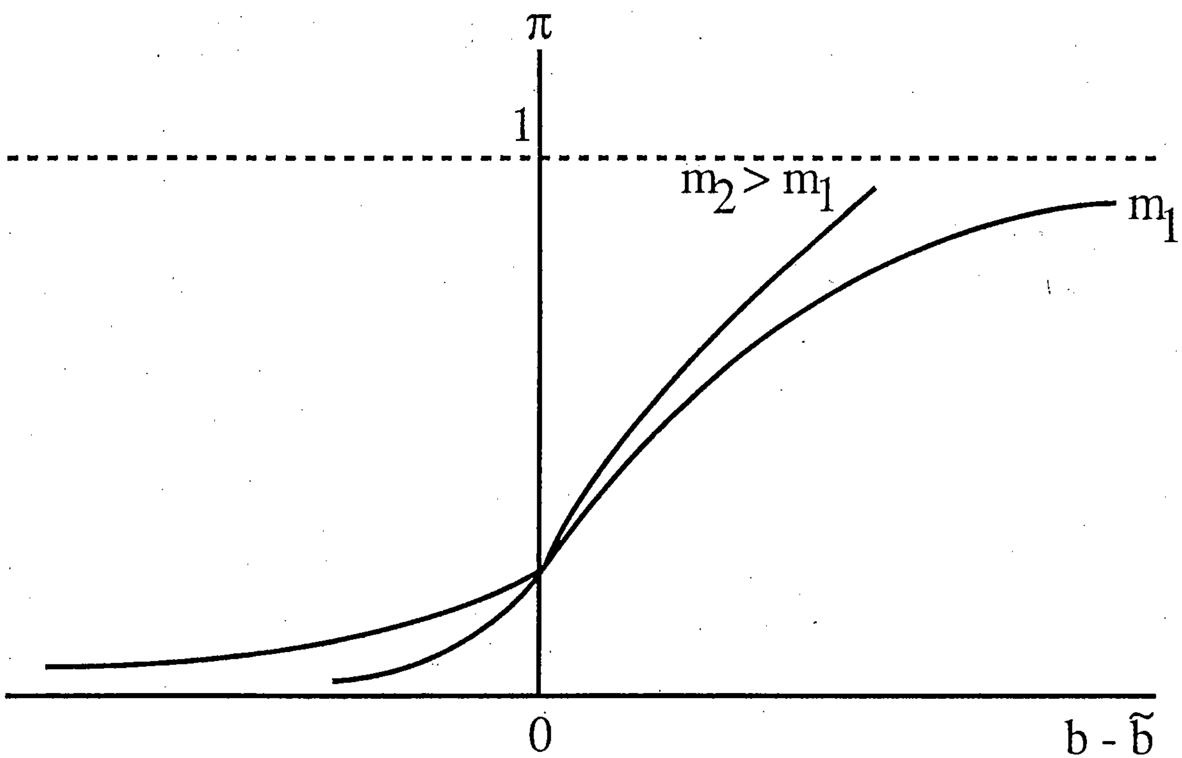

FIGURE 1

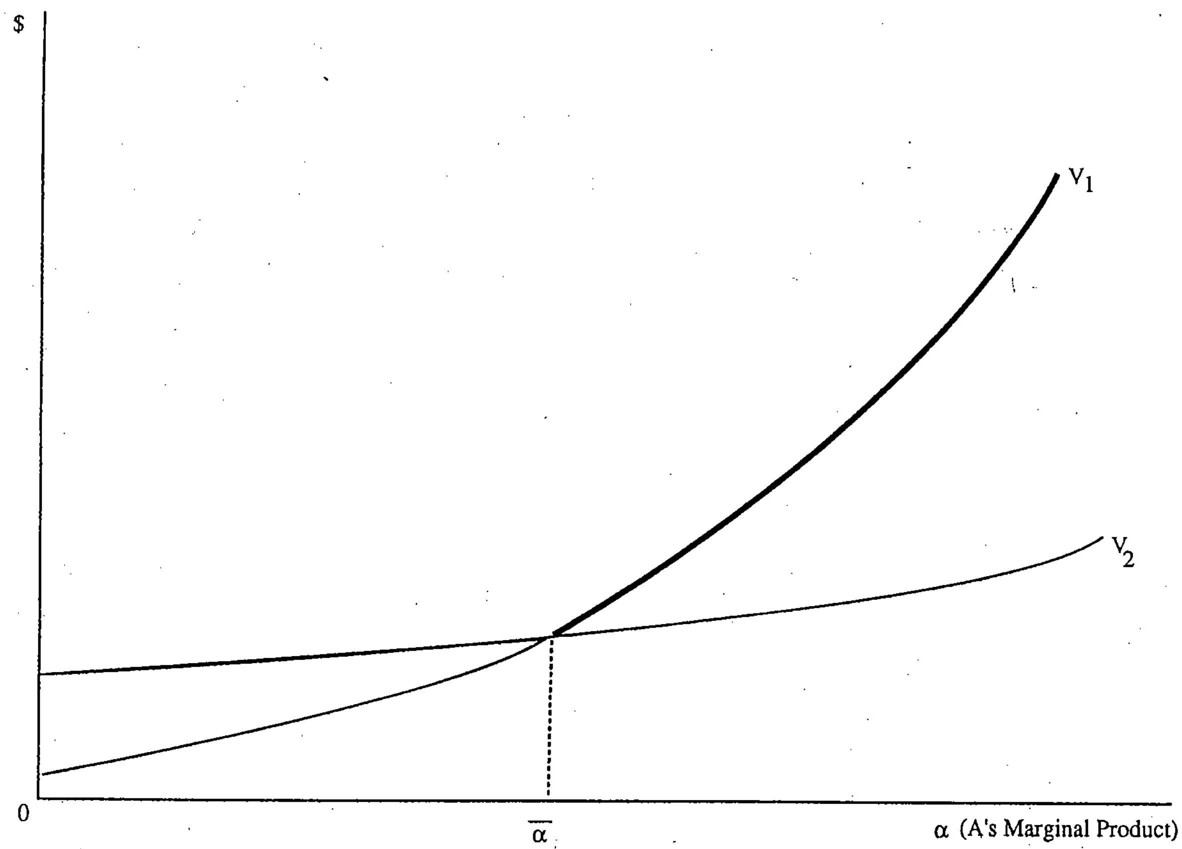

FIGURE 2

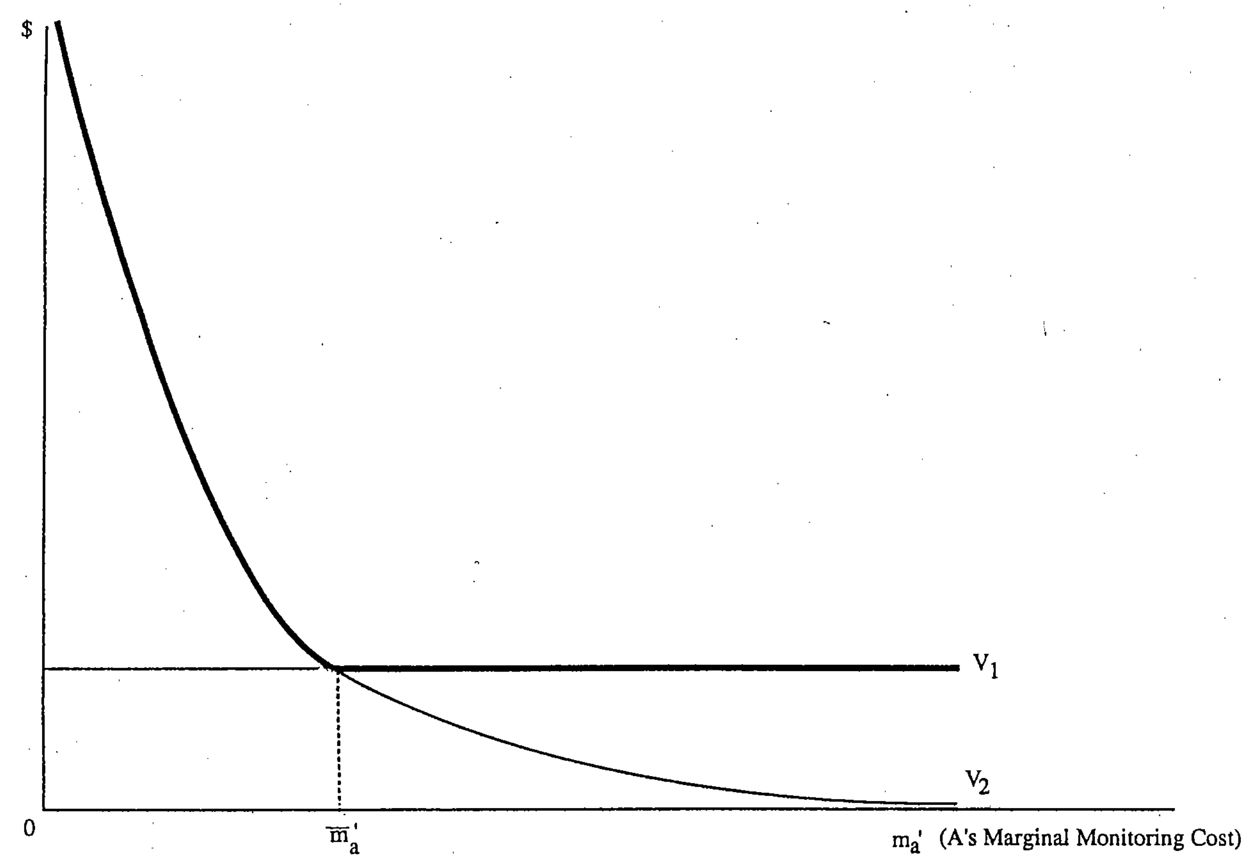

FIGURE 3

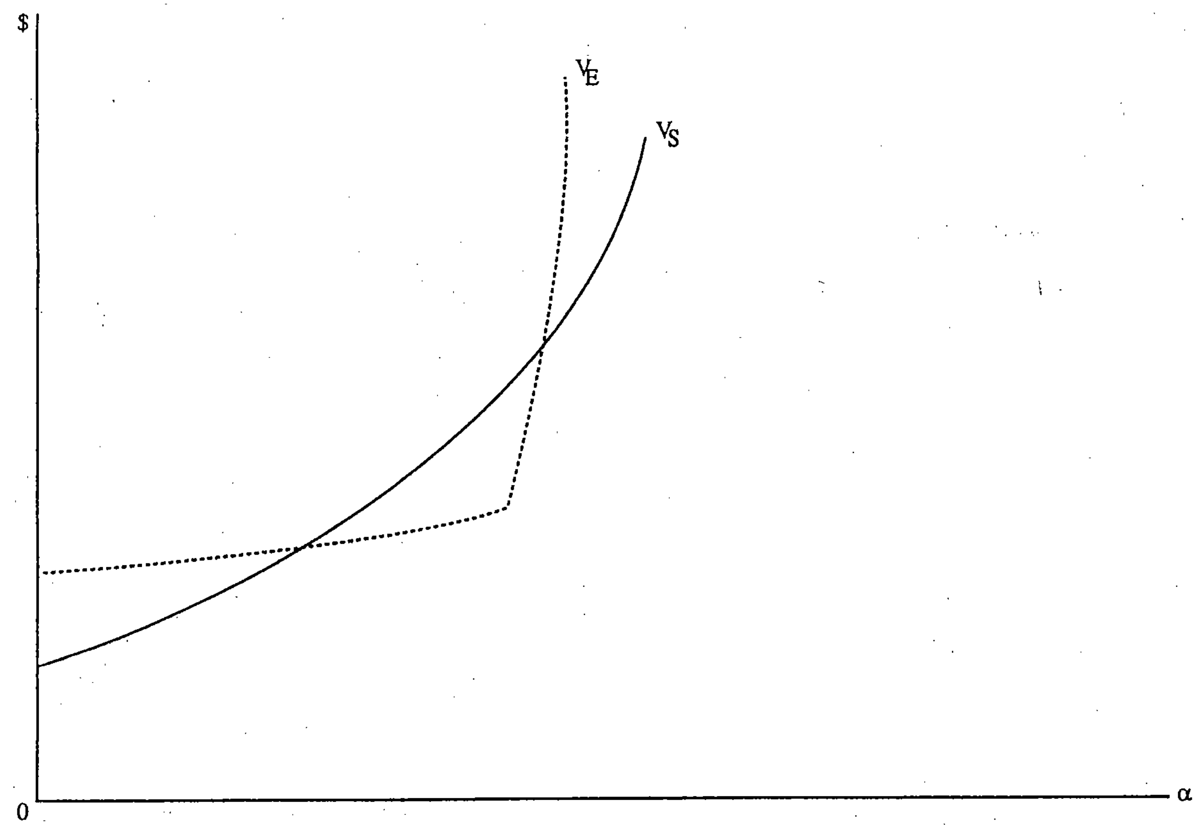

FIGURE 4