

NBER WORKING PAPERS SERIES

TAX POLICY AND THE  
DIVIDEND PUZZLE

B. Douglas Bernheim

Working Paper No. 3434

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
September 1990

I would like to thank Laurie Bagwell, as well as seminar participants at the NBER and Northwestern University for helpful comments. An earlier version of this paper was circulated under the title "Dividends versus Repurchases as Signals of Profitability" (dated July, 1988). Financial support from the National Science Foundation (Grant No. SE5-8821666) is gratefully acknowledged. This paper is part of NBER's research program in Taxation. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

TAX POLICY AND THE DIVIDEND PUZZLE

ABSTRACT

This paper offers a new explanation of the dividend puzzle, based upon a model in which firms attempt to signal profitability by distributing cash to shareholders. I assume that dividends and repurchases are identical, except that dividends are taxed more heavily. Nevertheless, I demonstrate that, under certain plausible conditions, corporations will pay dividends. Indeed, some firms will actually pay dividends, and then retrieve a portion of these payments by issuing new equity (perhaps through a dividend reinvestment plan), despite the fact that this appears to create gratuitous tax liabilities. In addition to providing an explanation for the dividend puzzle, I also derive a number of strong results concerning corporate payout decisions and government tax policy. Some of these results are surprising. For example, the relationship between repurchases and firm quality is hump-shaped. Moreover, despite the fact that a higher dividend tax rate depresses dividend payments, it does not affect either government revenue or welfare.

B. Douglas Bernheim  
Kellogg Graduate School  
of Management  
Northwestern University  
2001 Sheridan Road  
Evanston, IL 60208

## 1. Introduction

Why do companies pay dividends? This question has proven to be one of the most vexing puzzles in economics. Many theories have been proposed, but none has earned general acceptance. While it is not difficult to account for the distribution of some earnings, dividends are treated less favorably than repurchases (even under current law), and therefore appear to be strictly dominated as a mechanism for transferring resources to shareholders.<sup>1</sup> The common practice of paying dividends and issuing new equity simultaneously is especially difficult to understand, since a company could presumably reduce dividends and new equity issues by equal amounts, thereby reducing tax liabilities without altering net distributions.<sup>2</sup>

In this paper, I offer a new explanation of the dividend puzzle, based upon a simple model in which firms attempt to signal profitability by distributing cash to shareholders. Throughout, I assume that dividends and repurchases are identical, except that dividends are taxed more heavily. Nevertheless, I demonstrate that, under certain plausible conditions, corporations will pay dividends, rather than repurchase shares, in order to signal profitability. Indeed, some firms will actually pay dividends, and then retrieve a portion of these payments by issuing new equity (perhaps through a dividend reinvestment plan), despite the fact that this appears to create gratuitous tax liabilities.

In addition to providing an explanation for the dividend puzzle, I also derive a number of strong, and somewhat surprising results concerning corporate payout decisions. Generally, dividend payments rise monotonically with firm quality. In contrast, the relationship between cash distributed through repurchases and firm quality is usually hump-shaped, and the function relating reinvested dividends (new equity) to firm quality may have several

peaks. Under certain plausible conditions, I also obtain pooling at the lower end of the quality spectrum. This implies that there may be a large, heterogeneous population of firms that choose to make no distributions whatsoever. As quality crosses the upper threshold that defines this pool, dividends jump discontinuously to some positive level. Thus, there is a trough in the population distribution of dividends near zero.

The implications for government tax policy are even more surprising. As one might suspect, a higher dividend tax rate depresses dividends, and raises repurchases. I also show that it depresses the level of dividend reinvestment (new equity). Surprisingly, an increase in the tax rate applicable to repurchases has exactly the same effects; that is, it stimulates repurchases, and depresses both the level of dividends and dividend reinvestment. Moreover, in a wide range of circumstances, changes in these tax rates have absolutely no effect on either total (net) distributions to shareholders, shareholder welfare, or government revenue. Although my analysis treats investment as exogenous, it is also clear that changes in these tax rates have no impact on investment incentives. The model therefore isolates a set of plausible conditions under which dividend taxation is economically irrelevant, despite the fact that it affects payout policy.

While the use of a tax-disadvantaged method of distributing cash may at first seem counter-intuitive, there is a very simple explanation. If distributions involve some cost, and if the total cost of distributing a given amount of cash is higher for lower quality firms, then companies can potentially use distributions to signal profitability. If distributions are taxed, then companies will bear a higher cost on each dollar distributed.

However, high quality firms will not need to distribute as much cash to shareholders in order to deter imitation by lower quality firms. Reduced distributions entail both tax and non-tax savings.<sup>3</sup> When signaling costs have a very natural form, the resource savings dominate when tax rates are low, while the tax costs dominate when tax rates are high. Thus, there is an optimal tax rate that allows the firm to signal at minimum total cost. Firms achieve this optimal tax rate by combining repurchases, new equity issues, and dividends appropriately. Many results concerning tax policy follow directly from this observation. In particular, when statutory tax rates change, firms can maintain effective tax rates at optimal levels by adjusting the payout mix.

To obtain the "optimal tax" result, I formulate a standard signaling model similar to that introduced by Bhattacharya [1979]. In particular, paying dividends may expose a firm to various costs if it subsequently runs low on cash. Higher quality firms are assumed to be more liquid, so they are less likely to incur these costs. Thus, distributions discriminate between firms of different quality, and therefore provide a natural basis for signaling. However, in contrast to previous models, I allow firms the option of shutting down. This has profound implications.

If a firm can shut down, then the non-tax costs of signaling must be bounded. Thus, it is impossible for the signaling cost function to be globally convex, as in standard signaling models. Since lower quality firms have less to lose, their cost functions must flatten out more quickly than those of higher quality firms. Thus, though the marginal cost of distributions is initially higher for lower quality firms, limited liability implies that this relationship must eventually reverse. The optimal tax rate

produces a level of distributions that equates the marginal cost of signaling for high and low quality firms.<sup>4</sup>

The paper is organized as follows. Section 2 describes the model. Section 3 investigates the properties of the signaling cost function. Equilibria with two types of firms are derived in section 4. I consider models with arbitrary numbers of firms in section 5. Section 6 clarifies the relationship between this paper and the existing literature. Section 7 concludes with a discussion of implications.

## 2. The Model

In the section, I present a parsimonious model of corporate policy regarding distributions to shareholders. Briefly, this model depicts the following sequence of events. First, the firm undertakes productive activities, and incurs either actual or potential liabilities. Actual liabilities refer to debt. Potential liabilities include damages from lawsuits that might arise in the course of conducting business. Second, managers acquire private information concerning the firm's prospects. Third, they attempt to signal this information by distributing cash to shareholders (implicitly, I assume that operations begin to generate cash flows prior to this stage). Fourth, all uncertainty concerning returns and potential liabilities is resolved. Fifth, the managers decide whether or not to declare bankruptcy. A decision to continue operations with limited liquidity may entail a financial penalty. Finally, operations terminate and the firm is liquidated. All liabilities (actual and potential) come due. If the firm has remained solvent, shareholders receive residual earnings after compensating creditors.

In order to focus exclusively on corporate payout decisions, I do not model the first stage explicitly. Rather, I simply take liabilities and investment as given. Thus, I abstract from the possibility that firms might also signal profitability by manipulating initial capital structure, or by making commitments to new projects.

Let  $\theta$  denote the expected returns from a firm's activities, net of all liabilities (actual and potential), conditional upon the private information received by managers in the second stage. Investors are aware that  $\theta$  belongs to some set  $\Theta$ , and they also know the *ex ante* distribution of types over  $\Theta$ . Throughout much of this paper, I will assume that  $\Theta = (\underline{\theta}, \theta_1, \dots, \theta_N)$ , whereby convention  $\underline{\theta} < \theta_1$  and  $\theta_{n-1} < \theta_n$  for all  $n \in (2, \dots, N)$ . I will also use  $h(\cdot)$  to denote the population density over  $\Theta$ .

In the third stage, managers can choose to distribute cash to shareholders in two different forms (the extension to arbitrary numbers of forms is immediate). Let  $y_j$  denote the amount of cash distributed in the  $j^{\text{th}}$  form. I interpret  $y_1$  as repurchases, and  $y_2$  as dividends. I will use  $y$  to denote the vector  $(y_1, y_2)'$ .  $Y$  will represent total cash distributions, i.e.,  $Y = y_1 + y_2$ . I impose a non-negativity constraint on dividends ( $y_2 \geq 0$ ), and on total distributions ( $Y = y_1 + y_2 \geq 0$ ). I allow for negative repurchases in order to represent new equity issues and dividend reinvestment plans, but rule out the possibility that total distributions are negative. My object in imposing this last constraint is to abstract from decisions about capital structure, in order to focus exclusively on payout decisions.<sup>5</sup>

Distributions of the  $j^{\text{th}}$  type are taxes at the rate  $\tau_j$ . It is important to emphasize that, aside from taxes, all forms of payment are completely equivalent. I will use  $\tau$  to denote the vector  $(\tau_1, \tau_2)$ . For simplicity, I

assume that the tax rate on dividends is positive ( $\tau_2 > 0$ ), while repurchases are untaxed ( $\tau_1 = 0$ ). One might justify the latter assumption by arguing that capital gains taxes are avoidable (see Stiglitz [1983], or Constantinides and Scholes [1980]). I am inclined to discount this argument on the basis of empirical evidence (Poterba [1987]). Alternatively, if repurchases and liquidations are taxed at the same rate, than it is appropriate to set  $\tau_1 = 0$ , and to interpret  $\tau_2$  as the tax rate on dividends measured relative to the tax rate for repurchases. As discussed later, the analysis is not particularly sensitive to the assumption that  $\tau_1$  -- for the most part, one simply needs to assume that dividends are taxed at a higher rate than repurchases.

When uncertainty is resolved in the fourth stage, managers learn the true value of earnings and potential liabilities (although these liabilities do not yet come due). Since  $\theta$  represents expected net returns, actual net returns equal  $\theta + \epsilon$ , where  $\epsilon$  is a random variable with zero mean. Of course, the firm has already distributed cash to shareholders. Its residual return (net of all liabilities) is therefore

$$R = \theta - Y + \epsilon.$$

When  $R$  is sufficiently low, the firm experiences liquidity problems, and must raise capital at very unfavorable terms in order to complete its operations. Specifically (following Battacharya [1979]), I assume that there is some  $\lambda > 0$  such that if  $R < \lambda$ , the firm incurs a cost of  $\beta$  per dollar of shortfall. Thus, the total penalty is  $\beta(\lambda - R)$ .

Management elects to shut the firm down in the fifth stage whenever the

residual value of the firm, net of any costs arising from liquidity problems, is negative. Accordingly, managers declare bankruptcy whenever  $R < \beta(\lambda - R)$ . Alternatively, one can express this condition as

$$R < \beta\lambda(1+\beta)^{-1} = \mu.$$

I assume that, if a firm folds, its shareholders are completely protected from creditors, and indeed creditors are unable to recover any dividends paid prior to its collapse.

When the firm is liquidated in the final stage, payments to shareholders will depend upon expected returns ( $\theta$ ), realizations ( $\epsilon$ ), and previous payments ( $Y$ ). Let  $\rho(Y, \theta, \epsilon)$  denote these terminal returns. On the basis of the preceding discussion, we have

$$\rho(Y, \theta, \epsilon) = \begin{cases} \theta - Y + \epsilon & \text{if } \theta - Y + \epsilon \geq \lambda \\ (\theta - Y + \epsilon)(1 + \beta) - \beta\lambda & \text{if } \lambda > \theta - Y + \epsilon \geq \mu \\ 0 & \text{if } \mu > \theta - Y + \epsilon \end{cases}$$

Assuming for the sake of convenience that the rate of discount is zero, then total payments to shareholders are given by the sum of terminal payments, and after-tax distributions:

$$\rho(Y, \theta, \epsilon) + Y - \tau y.$$

So far, I have discussed the sequence of decisions and events without saying much about the objectives of management. It is natural to assume that management acts to maximize the value of equity. In the fifth stage, I have therefore assumed that managers declare bankruptcy if and only if the firm is insolvent. This is not controversial, since all uncertainty is resolved in

stage four. However, in the third stage, managers have better information than investors. As a result, one could measure the value of equity in two distinct ways: either as market value (which is based on investors' perceptions), or as the managers' assessment of value (which is based on superior information).

I will assume that managers care about both current market value, and their own assessment of value. This can be justified as follows. Managers who act in the interests of current shareholders will certainly wish, ceteris paribus, to maximize the current market value of outstanding shares. Unfortunately, current shareholders are not better informed than other investors -- if management dupes the market (perceived value exceeds real value), it will also dupe the investors it serves. Not knowing that it is in their interests to sell out prior to realization, some fraction of the original shareholders will retain their stock and eventually receive a payment that is below the pre-realization share price. Thus, management should also care about the actual value of the firm. Similar conclusions would follow if one assumed that managers would lose their jobs in the event of bankruptcy, and that this would entail some personal cost (as in Ross [1977]).

Henceforth, I will use  $\hat{V}(y, \tau)$  to denote the cum dividend market valuation of the firm in stage 3. Ultimately, this function will describe the endogenous relationship between payout policy and investors' beliefs generated by a signaling equilibrium. The manager's assessment of value (cum dividend) will be given by a function  $V(y, \theta, \tau)$ . As discussed in the preceding paragraph, I will assume that management cares about both  $\hat{V}(\cdot)$  and

$V(\cdot)$ . In particular, managers will act to maximize the following objective function:

$$(1) \quad \hat{V}(y, \tau) + \alpha V(y, \theta, \tau),$$

where  $\alpha$  is an exogenous parameter.

The function  $V(\cdot)$  will represent an objective estimate of the firm's value under perfect information. That is, it provides an answer to the following hypothetical question: how would the market value this firm if investors were as well informed as managers? To simplify this calculation, I will assume that investors are risk neutral. Thus,

$$(2) \quad \begin{aligned} V(y, \theta, \tau) &= \int_{-\infty}^{\infty} \rho(Y, \theta, \epsilon) f(\epsilon) d\epsilon + Y - r y \\ &= v(Y, \theta) + Y - r y, \end{aligned}$$

where  $f(\epsilon)$  denotes the probability density of  $\epsilon$ . Most of the calculations in subsequent sections will be based on the assumption that

$$(3) \quad \epsilon \sim N(0, \sigma^2).$$

### 3. Properties of the Signaling Cost Function

It is useful to think of the objective function (1) as

$$\hat{V}(y, \tau) - [-\alpha V(y, \theta, \tau)].$$

When management's objectives are written in this way, it is evident that signaling creates a benefit through its effect on market perceptions, and a cost through its impact on actual value. Consequently, we can define the signaling cost function as

$$c(y, \theta, \tau) = -\alpha V(y, \theta, \tau).$$

Note that, as in the standard model, cost depends both upon the value of the signal and the type of firm. The standard assumptions in most signaling models are that  $c_j > 0$ ,  $c_{jj} > 0$ , and  $c_{j\theta} < 0$  (where the subscript  $j$  denotes a partial derivative with respect to  $y_j$ ). These assumptions provide the basis for the "single crossing property" that is typically used to guarantee the existence of signaling equilibria.

For the model considered here, the signaling cost function cannot possibly satisfy these standard assumptions. Since the firm can always elect to shut down, its value is bounded below by 0. Thus,  $c(y, \theta, \tau)$  is bounded above by zero. This rules out the possibility that costs are globally convex. Although the cost function might start out convex at  $y=0$ , it must begin to flatten out before reaching the upper bound.

More importantly, it is extremely unlikely that the marginal cost of signaling would be higher for lower quality firms at all values of  $y$ . Certainly, this condition might hold for small distributions. However, as distributions rise, the total cost for low quality firms will approach its upper bound more quickly. This follows from two separate considerations. First, if marginal cost is initially higher for lower quality firms, total cost will approach the upper bound faster. Second, lower quality firms have less to lose, so the upper bound is lower. Consequently, the cost function for low quality firms must flatten out sooner. This implies that, for large distributions, marginal costs should be higher for high quality firms.

I will now develop these ideas formally. Observe that

$$(4) \quad V_j(y, \theta, \tau) = v_Y(Y, \theta) + (1 - r_j),$$

and

$$(5) \quad v_{jk}(y, \theta, r) = v_{YY}(Y, \theta).$$

Thus, the impact of taxes on marginal cost is independent to type ( $\theta$ ). Moreover, for any given form of payment, marginal cost depends only on total distributions. The second derivative of signaling costs is independent of both taxes and the form of distribution. Like marginal cost, it depends only on total distributions.

In light of these observations, it is possible to deduce the important properties of the signaling cost function by analyzing  $v(\cdot)$ . It is straightforward to verify that

$$(6) \quad v_Y = (1+\beta)F(Y-\theta+\mu) - \beta F(Y-\theta+\lambda) - 1$$

and

$$(7) \quad v_{YY} = (1+\beta)f(Y-\theta+\mu) - \beta f(Y-\theta+\lambda),$$

where  $F(\cdot)$  is the cumulative distribution function associated with  $f(\cdot)$ . The signaling cost function is convex at  $Y$  if and only if

$$(8) \quad \frac{f(Y-\theta+\lambda)}{f(Y-\theta+\mu)} > \frac{1+\beta}{\beta}.$$

(Recall that the cost function is inversely proportional to the value function). Under the assumption that  $\epsilon$  has a normal distribution (expression (3)), it is a simple matter to verify that

$$\frac{f(Y-\theta+\lambda)}{f(Y-\theta+\mu)} = \exp\left\{\frac{1}{2}[2(Y-\theta)(\mu-\lambda) + \mu^2 - \lambda^2]\sigma^{-2}\right\}.$$

Since  $\mu - \lambda < 0$ , this ratio falls monotonically with  $Y$ , and has a limiting value of zero. Thus, even when the cost function starts out convex at  $Y = 0$ , it eventually becomes concave. Moreover, monotonicity of the likelihood ratio implies that there is a single point of inflection,  $\bar{Y}(\theta)$  (note that  $\bar{Y}$  does not depend upon  $\tau$ ).

On the basis of the preceding arguments, I conclude that the function  $V(\cdot)$  has the general shape depicted in figure 1 (here, the subscript "-j" denotes 1 when  $j = 2$ , and 2 when  $j = 1$ ). The slope of this function converges to  $(1 - \tau_j)$  as  $y_j$  get large. The diagram immediately suggests an important problem: if there are no constraints on corporate distributions, the optimal strategy is to pay out as much to shareholders as possible, leaving creditors with worthless claims. In practice, payment of dividends is constrained by the liquid resources of the firm, as well as by debt covenants (not surprisingly, creditors anticipate and attempt to preclude such practices). A contractual limitation on distributions to shareholders would eliminate extreme opportunism, without affecting any of the following analysis (for example, one could choose the level of  $Y$  that minimizes  $V$ ). Moreover, this constraint would not appear to bind in equilibrium, in the sense that the contractual limitation would exceed the amount paid to shareholders under any conceivable realization. The model therefore accounts for Kalay's [1982] puzzling observation that actual dividends are usually strictly less than the maximum amounts specified in debt covenants.

With this specification of signaling costs, what becomes of the single-crossing property? Figure 2 depicts marginal cost ( $-\alpha V_j(\cdot)$ ) as a function of  $y_j$  for two distinct values of  $\theta$ . From equations (4) and (6), we have

$$V_j(y_j, y_{-j}^0, \theta_1, \tau) = V_j(y_j + \theta_2 - \theta_1, y_{-j}^0, \theta_2, \tau).$$

In other words, the marginal cost function is identical for all types of firms, except that higher quality shifts this function to the right. Thus, for each  $\theta_1, \theta_2 \in \Theta$  and  $y_{-j}^0$ , there exists some  $y_j^*$  with the following property: below  $y_j^*$ , the marginal cost of distributing earnings is higher for low quality firms, as in traditional signaling models, but above  $y_j^*$  high quality firms experience higher marginal costs. Because of this reversal, indifference curves cross twice, rather than one. This property is critical in the subsequent analysis.

In deriving many of the results that follow, I exploit certain properties of  $y_j^*$ . In particular, I will focus on pairs of types that include  $\underline{\theta}$ , defined as the lowest value of  $\theta$  in  $\Theta$ . For  $\underline{\theta}$  and any other  $\theta \in \Theta$ ,  $y_j^*$  is determined by the following equation:

$$V_j(y_j^*, y_{-j}^0, \underline{\theta}, \tau) = V_j(y_j^*, y_{-j}^0, \theta, \tau).$$

After substituting (4), one obtains

$$(9) \quad v_Y(y_{-j}^0 + y_j^*, \underline{\theta}) = v_Y(y_{-j}^0 + y_j^*, \theta)$$

Note that only the sum of  $y_{-j}^0$  and  $y_j^*$  appears in equation (9). Thus, to equalize marginal costs between types  $\underline{\theta}$  and  $\theta$ , we need only worry about total distributions -- composition is irrelevant.<sup>6</sup> This greatly simplifies the determination of  $y_j^*$ . Let  $Y^*(\theta)$  denote the value of  $Y$  that satisfies

$$(10) \quad v_Y(Y, \underline{\theta}) = v_Y(Y, \theta).$$

Then  $y_j^* = Y^*(\theta) - y_{-j}^0$ .

Note also that the vector of tax rates,  $\tau$ , does not appear in equation

(9). This makes good sense, since  $r$  makes the same contribution to marginal cost for all types. Thus,  $Y^*(\cdot)$  depends only upon  $\theta$ .

Finally, I will need one additional result concerning the function  $Y^*(\cdot)$ . Implicit differentiation of (10) reveals that<sup>7</sup>

$$Y_{\theta}^*(\theta) = \frac{-v_{YY}(Y^*(\theta), \theta)}{v_{YY}(Y^*(\theta), \underline{\theta}) - v_{YY}(Y^*(\theta), \theta)}$$

Since  $\bar{Y}(\underline{\theta}) < Y^*(\theta) < \bar{Y}(\theta)$ ,  $v_{YY}(Y^*(\theta), \underline{\theta}) > 0$ , and  $v_{YY}(Y^*(\theta), \theta) < 0$  (recall that  $\bar{Y}(\theta)$  is the unique inflection point for  $v(Y, \theta)$ ). It follows immediately that

$$0 < Y_{\theta}^*(\theta) < 1.$$

In words, the level of distributions that equates marginal cost between types  $\underline{\theta}$  and  $\theta$  rises with  $\theta$ , but the increment is less than dollar-for-dollar.

#### 4. Efficient Signaling and the Dividend Puzzle

A signaling equilibrium assigns a market valuation,  $\hat{V}(y, r)$ , to each vector of signals  $y$ , for a particular tax vector  $r$ . Actual value is given by the function  $V(y, \theta, r)$ . In equilibrium, each manager maximizes  $\hat{V}(y, r) + \alpha V(y, \theta, r)$  over  $y$ , subject to  $y_2 \geq 0$  and  $y_1 + y_2 \geq 0$ . The resulting pattern of choices justifies market perceptions, in the sense that  $\hat{V}(y, r)$  is the average value of  $V(\cdot)$  for firms that choose  $y$ .

My object here is to derive the characteristics of the most efficient signaling equilibrium. For the most part, I focus on equilibria with complete separation. Since low quality firms never signal, efficiency is determined exclusively by the payoff to high quality firms. It is therefore

possible to determine the most efficient equilibrium by maximizing the payoff to high quality firms over all possible separating equilibria.<sup>8</sup> Formally, the problem is to solve

$$(11) \quad \hat{\max}_{V(\cdot), y} \hat{V}(y, r) + \alpha V(y, \theta, r)$$

subject to

$$(12) \quad \hat{V}(0, r) + \alpha V(0, \underline{\theta}, r) \geq \hat{V}(y, r) + \alpha V(y, \underline{\theta}, r)$$

$$(13) \quad \hat{V}(y, r) + \alpha V(y, \theta, r) \geq \hat{V}(0, r) + \alpha V(0, \theta, r)$$

$$(14) \quad \hat{V}(y, r) = V(y, \theta, r)$$

$$(15) \quad \hat{V}(0, r) = V(0, \underline{\theta}, r)$$

Equation (12) implies that low quality firms are content to choose 0 rather than  $y$ , while (13) guarantees that high quality firms prefer  $y$  over 0. Equations (14) and (15) guarantee that expectations are confirmed in equilibrium. There are also non-negativity constraints on  $y_1$  and  $y_1 + y_2$ . For expositional clarity, I will begin by ignoring the non-negativity constraints, and then subsequently consider their impact on the solution.

In the appendix (lemma 1), I demonstrate that the solution to the preceding problem has the following properties:

$$(16) \quad y_1 + y_2 = Y^*(\theta),$$

and

$$(17) \quad y_2 = [\tau_2(1+\alpha)]^{-1} ([v^0(Y^*(\theta), \theta, 0) - v^0(0, \underline{\theta}, 0)])$$

$$+ \alpha[V^0(Y^*(\theta), \underline{\theta}, 0) - V^0(0, \underline{\theta}, 0)],$$

where

$$(18) \quad V^0(Y, \theta, q) = v(Y, \theta) + Y - qY.$$

(That is,  $V^0(\cdot)$  describes the perfect information value of the firm under the counterfactual assumption that there is only one form of distribution, which is taxed at the rate  $q$ .)

In deriving the most efficient equilibrium, I have ignored the non-negativity constraints on  $y_2$  and  $y_1 + y_2$ . It is, of course, possible to have  $Y^*(\theta) < 0$ , in which case (16) would imply  $y_1 + y_2 < 0$ . Then, for all  $j$  and any initial  $y_1$ , the marginal cost of increasing  $y_j$  would be greater for high quality firms. Consequently, the total cost associated with any  $y \geq 0$  would also be greater for high quality firms. This implies that it would be impossible to signal value by making payments to shareholders, regardless of whether the firm uses dividends or repurchases.

It is also possible that the right hand side of (17) may be negative. In that case, the managers of high quality firms would signal by distributing resources to shareholders exclusively through repurchases. It is therefore important to isolate conditions under which this expression is positive. I will undertake this task shortly.

#### 4.1 An Interpretation

In this section, I offer an interpretation of the optimal signal derived in the preceding section. Specifically, consider a model in which firms can distribute cash to shareholders in only one form. Under this assumption, there exists an optimal tax rate on distributions, in the sense that this tax

rate gives rise to a separating equilibrium in which the total costs of signaling are smaller than in any other separating equilibrium.<sup>9</sup> To determine the optimal policy with multiple signals, one simply combines dividends, repurchases, and new equity issues in proportions that yield an effective tax rate equal to the optimal rate.

The remainder of this section is devoted to an intuitive development of the optimal tax result. Throughout, I simplify the notation by omitting subscripts on distributions and taxes, treating  $y$  and  $r$  as scalars. Since there is only one form of distribution, this does not introduce any ambiguity.

In a separating equilibrium, high quality firms select a particular level of distributions,  $\hat{y}$ , to signal their type. Together,  $\hat{y}$  and the market perception function  $\hat{V}^o(\cdot)$  must satisfy the following constraints:

$$(19) \quad \hat{V}^o(0, r) + \alpha V^o(0, \underline{\theta}, r) \geq \hat{V}^o(\hat{y}, r) + \alpha V(\hat{y}, \underline{\theta}, r)$$

$$(20) \quad \hat{V}^o(0, r) + \alpha V^o(0, \theta, r) \leq \hat{V}^o(\hat{y}, r) + \alpha V^o(\hat{y}, \theta, r)$$

$$(21) \quad \hat{V}^o(\hat{y}, r) = V^o(\hat{y}, \theta, r)$$

$$(22) \quad \hat{V}^o(0, r) = V^o(0, \underline{\theta}, r)$$

(following the convention introduced earlier, I use the superscript "o" to distinguish  $\hat{V}^o(\cdot)$  and  $V^o(\cdot)$ , the valuation functions with one signal, from  $\hat{V}(\cdot)$  and  $V(\cdot)$ , the valuation functions with two signals). These equations are analogous to (12) through (15). Equations (19) and (20) provide for mutual non-imitation, while (21) and (22) guarantee that expectations are confirmed in equilibrium.

To depict a candidate signaling equilibrium, it is helpful to illustrate the low quality firm's preferences by drawing indifference curves in  $(\hat{V}, y)$  space (see figure 3). The formula for an indifference curve is

$$(23) \quad \hat{V} + \alpha V^0(y, \underline{\theta}, r) = k,$$

where  $k$  is an arbitrary constant. Note that we can rewrite (23) as

$$\hat{V} = k + c(y, \underline{\theta}, r).$$

In other words, to plot an indifference curve, one simply draws the cost function, and adjusts the intercept.

Since my object is to depict the incentive constraint for low quality firms (equation (19)), it is appropriate to set  $k = V^0(0, \underline{\theta}, r)(1+\alpha)$  (here, I have also made use of (22)). For this value of  $k$ , one obtains the indifference curve  $\underline{I}$  depicted in figure 3. Together, equations (22) and (19) imply that  $(\hat{y}, \hat{V}^0(\hat{y}, r))$  must lie to the right of  $\underline{I}$ . Equation (22) places an additional constraint on this point. To represent this constraint, I plot the function  $V^0(y, \theta, r)$  in figure 3.  $(\hat{y}, \hat{V}^0(\hat{y}, r))$  must also lie on this line.

As is clear from the figure, there are many solutions to equations (19), (21) and (22) (anything on  $V^0(y, \theta, r)$  to the right of  $\underline{I}$ ). All of these points are legitimate candidates for separating equilibria. However, they are clearly Pareto ranked. Let  $\hat{y}(\theta, r)$  denote the value of  $\hat{y}$  that solves

$$(24) \quad (1+\alpha)V^0(0, \underline{\theta}, r) = V^0(\hat{y}, \theta, r) + \alpha V^0(\hat{y}, \underline{\theta}, r).$$

In words,  $\hat{y}(\theta, r)$  represents the level of distributions for which the incentive constraint for lower quality firms just binds (see figure 3). Clearly, of all the potential candidates for a signal,  $\hat{y}(\theta, r)$  is most

efficient in the sense that it involves the lowest cost for high quality firms.

So far, I have ignored the incentive constraint for high quality firms (equation (20)). In standard signaling models, the single crossing property renders this constraint redundant -- if low quality agents do not want to imitate high quality agents, then high quality agents will strictly prefer to identify themselves through signaling. Of course, this observation does not apply here, since this model violates the single crossing property. To represent this condition graphically, I draw the indifference curve for the manager of a high quality firm through the point  $(\hat{y}(\theta, \tau), \hat{V}^0(\hat{y}(\theta, \tau), \tau))$ . There are three important cases to consider.

First, the high quality indifference curve may cut the lower quality indifference curve from above at  $(\hat{y}(\theta, \tau), \hat{V}^0(\hat{y}(\theta, \tau), \tau))$  (see the curve labeled  $I_1$  in figure 4). This necessarily implies that it intersects the vertical axis above  $\hat{V}^0(0, \tau)$ . Consequently, the incentive compatibility constraint for high quality workers is satisfied. This is the configuration found in most standard signaling models.

Recall that the indifference curves correspond to cost functions, where the axis has been shifted. The fact that  $I_1$  is flatter than  $I$  therefore implies that, at  $\hat{y}(\theta, \tau)$ , the marginal cost of signaling is lower for the high firm. Thus, this first case arises whenever

$$(25) \quad \hat{y}(\theta, \tau) < Y^*(\theta).$$

Second, the high quality indifference curve may cut the low quality indifference curve from below, but nevertheless intersect the vertical axis above  $\hat{V}^0(0, \tau)$ . This occurs when the high and low quality indifference curves

cross twice, rather than once (see  $I_2$  in figure 4). In this case, the incentive compatibility constraint for high quality firms is still satisfied, even though the configuration is non-standard. In particular, since  $I_2$  is steeper than  $\underline{I}$  at  $\hat{y}(\theta, \tau)$ , the marginal cost of signaling is higher for high quality firms. Thus, this second case arises whenever

$$(26) \quad \hat{y}(\theta, \tau) > Y^*(\theta)$$

and

$$(27) \quad V^0(0, \underline{\theta}, \tau) - V^0(\hat{y}(\theta, \tau), \underline{\theta}, \tau) > V^0(0, \theta, \tau) - \hat{V}^0(\hat{y}(\theta, \tau), \theta, \tau)$$

(the first condition implies that the marginal cost of signaling is higher for higher quality firms, while the second guarantees that the total cost of signaling is higher for lower quality firms).

Finally, the high quality indifference curve may cut the low quality indifference curve from below, and intersect the vertical axis below  $\hat{V}(0, \tau)$  (see  $I_3$  in figure 4). In this case, the incentive compatibility constraint for high quality firms is not satisfied, and there exists no equilibria with complete separation of types. This occurs whenever (25) holds, but (26) does not.

I now turn to the central issue, which concerns the relationship between the equilibria described above and the tax parameter,  $\tau$ . Graphically, as  $\tau$  rises,  $\underline{I}$  rotates up and  $V^0(y, \theta, \tau)$  rotates down, each pivoting on its vertical intercept. Thus, the equilibrium signal ( $\hat{y}$ ) declines. Figure 5 illustrates the determination of  $\hat{y}(\theta, \tau)$  for three tax rates,  $\tau_1, \tau_2$ , and  $\tau^*$ , where  $\tau_1 > \tau^* > \tau_2$ . Both the shape and location of indifference curves depend upon the tax rate. Thus,  $I_k$  and  $I'_k$  represent indifference curves for a high

quality firm when the tax rates is  $\tau_k$ , while  $I_k$  represents an indifference curve of the low quality firm. It is important to bear in mind that  $I_k$ ,  $I'_k$ , and  $\underline{I}_k$  are not valid indifference curves when the tax rate is  $\tau_j$ ,  $j \neq k$ . Clearly, for sufficiently high tax rates,  $\hat{y}(\theta, \tau) < Y^*(\theta)$  (this configuration is generated by  $\tau_1$ ), while for sufficiently low (possibly negative) tax rates,  $\hat{y}(\theta, \tau) > Y^*(\theta)$  (this configuration is generated by  $\tau_2$ ).

How does the tax rate affect management's perception of value? In the appendix (lemma 2) I demonstrate that a higher tax rate benefits high quality firms (and leaves low quality firms unaffected) if and only if  $\hat{y}(\theta, \tau) > Y^*(\theta)$ . This in turn implies that there exists an optimal tax rate that gives rise to the most efficient separating equilibrium. I refer the reader again to figure 5. As one reduces the tax rate from  $\tau_1$ ,  $\hat{y}(\theta, \tau)$  moves to the right. Since  $\hat{y}(\theta, \tau) < Y^*(\theta)$ , this improves the lot of high quality firms. For some particular tax rate between  $\tau_1$  and  $\tau_2$  (call it  $\tau^*$ ), we obtain

$$(28) \quad \hat{y}(\theta, \tau^*) = Y^*(\theta).$$

This implies that the indifference curves of high and low quality firms ( $I^*$  and  $\underline{I}^*$ ) are just tangent at the point  $(\hat{y}(\theta, \tau^*), \hat{V}^0(\hat{y}(\theta, \tau^*), \tau^*))$ . A further reduction of the tax rate would push  $\hat{y}(\theta, \tau)$  above  $Y^*(\theta)$ , and would therefore reduce the value of high quality firms. Hence,  $\tau^*$  is the optimal tax rate.

The intuition for lemma 2 is simple. Taxes do not affect the well-being of low quality firms, and such firms are always indifferent between  $(0, \hat{V}^0(0, \tau))$  and  $(\hat{y}(\theta, \tau), \hat{V}^0(\hat{y}(\theta, \tau), \tau))$ . Suppose then that some low quality firm chooses to imitate the high quality firms. In moving from an equilibrium signal with low taxes to one with high taxes, the resource saving associated with smaller distributions to shareholders would just compensate

these firms for the increase in tax payments. If distributions are relatively less costly on the margin for high quality firms ( $\hat{y}(\theta, \tau) < Y^*(\theta)$ ), then the same tradeoff should leave high quality firms strictly worse off. On the other hand, if distributions are more costly on the margin for high quality firms ( $\hat{y}(\theta, \tau) > Y^*(\theta)$ ), then the same tradeoff leaves high quality firms strictly better off.

Clearly, the optimal tax rate  $\tau^*$  depends upon  $\theta$ . Since lower quality firms must be indifferent between  $(0, \hat{V}^0(0, \tau^*))$  and  $(Y^*(\theta), \hat{V}^0(Y^*(\theta), \tau^*))$ , we have

$$(29) \quad \tau^*(\theta) = [Y^*(\theta)(1+\alpha)^{-1}] \{V^0(Y^*(\theta), \theta, 0) - V^0(0, \underline{\theta}, 0)\} \\ + \alpha [V^0(Y^*(\theta), \underline{\theta}, 0) - V(0, \underline{\theta}, 0)].$$

So far, I have restricted attention to equilibria with complete separation. Whenever condition (26) holds, there are also attractive equilibria with incomplete separation. Figure 6 illustrates this possibility. Since, the high and low quality indifference curves through any point  $(Y^*(\theta), \hat{V})$  are tangent, we can construct an equilibrium as follows.

Let

$$\hat{V}^0(Y^*(\theta), \tau) = V^0(0, \underline{\theta}, \tau)(1+\alpha) - \alpha V^0(Y^*(\theta), \underline{\theta}, \tau)$$

(that is,  $\hat{V}^0(Y^*(\theta), \tau)$  is given by the vertical coordinate of the tangency depicted in figure 6). Also, let  $\eta$  be defined as the solution to

$$(30) \quad \frac{\eta h(\underline{\theta}) V^0(Y^*(\theta), \underline{\theta}, \tau) + h(\theta) V^0(Y^*(\theta), \theta, \tau)}{\eta h(\underline{\theta}) + h(\theta)} = \hat{V}^0(Y^*(\theta), \tau)$$

(recall that  $h(\cdot)$  is defined as the population density over  $\Theta$ ). Suppose for the moment that  $h(\underline{\theta})$  is sufficiently large, so that  $\eta$  lies between 0 and 1. Low quality firms are indifferent between  $(0, \hat{V}^0(0, r))$  and  $(Y^*(\theta), \hat{V}^0(Y^*(\theta), r))$ . I can resolve this indifference by assuming that the fraction  $\eta$  of these firms pay  $Y^*(\theta)$  to shareholders, while  $(1-\eta)$  make no distributions at all. As illustrated in figure 6, all high quality firms strictly prefer  $(Y^*(\theta), \hat{V}^0(Y^*(\theta), r))$  to  $(0, \hat{V}^0(0, r))$ , and therefore choose  $Y^*(\theta)$ . Thus, incentive compatibility holds for both types of firms, and, given (30), beliefs about average quality are self-fulfilling.

When condition (26) holds, imperfectly separating equilibria may coexist with perfectly separating equilibria. We are then confronted with a choice between equilibria. How do we resolve this ambiguity? The partially separating equilibrium is more efficient, in the sense that it weakly Pareto dominates the completely separating equilibrium (high quality firms are better off).<sup>10</sup> On the other hand, standard refinements (such as the Cho-Kreps intuitive criterion) do not rule out complete separation. Moreover, the partially separating equilibrium relies on a very delicate balance of resolving indifference differently for firms of the same type, and may therefore seem fragile in comparison to an equilibrium with complete separation.

Fortunately, consideration of imperfectly separating equilibria does not alter the optimal tax result. In the appendix (lemma 3), I show that the payoffs associated with the equilibrium described above are insensitive to the tax rate. Moreover, when  $r = r^*(\theta)$ , this equilibrium coincides exactly with the optimal separating equilibrium. Consequently, the tax rate  $r^*(\theta)$

gives rise to a separating equilibrium that is as good as the imperfectly separating equilibria that may exist for other tax rates.<sup>11</sup>

Equation (29) (the "optimal tax" formula) suggests a natural interpretation of the efficient equilibrium with multiple signals. In particular, equations (16), (17), and (29) imply that the optimal values of  $y_1$  and  $y_2$  (hereforth written as functions of  $\theta$  and  $\tau_2$ ) are given by the following equations:

$$(31) \quad y_1^*(\theta, \tau_2) = Y^*(\theta) \left(1 - \frac{\tau^*(\theta)}{\tau_2}\right)$$

and

$$(34) \quad y_2^*(\theta, \tau_2) = Y^*(\theta) \tau^*(\theta) / \tau_2$$

In other words, one can simply think of each firm as choosing its tax rate by adopting a particular combination of dividends and repurchases. Naturally, it selects the optimal rate,  $\tau^*(\theta)$ .

#### 4.2 Implications

I now return to the model with multiple signals, and investigate the manner in which the optimal signal varies with firm quality ( $\theta$ ). In doing so, I focus on the "standard" case in which the cost function for the lowest quality firm ( $-\alpha V(y, \underline{\theta}, \tau)$ ) is initially convex (i.e.,  $\ddot{Y}(\underline{\theta}) > 0$ ).

First, consider small values of  $\theta$ . Recall that  $Y^*(\theta) > \ddot{Y}(\underline{\theta})$  for all  $\theta > \underline{\theta}$ . Therefore, as  $\theta$  goes to  $\underline{\theta}$ ,  $Y^*(\theta)$  is bounded away from 0. It is then immediately clear from inspection of equation (29) that

$$\lim_{\theta \rightarrow \underline{\theta}} \tau^*(\theta) < 0.$$

Consequently, managers use repurchases to signal small differences in value.

What happens to  $r^*(\theta)$  as  $\theta$  gets large? First of all, it is clear from equation (2) that for any given  $y$ , one can make  $V(y, \theta, 0)$  arbitrarily large by selecting a large enough value of  $\theta$ . Moreover,  $Y^*(\theta)$  is bounded. In particular, define  $\bar{Y}^*$  as the value of  $Y$  that minimizes  $v(Y, \theta)$ . It is easy to show that  $Y^*(\theta) < \bar{Y}^*$  for all  $\theta > \underline{\theta}$ , and that  $Y^*(\theta)$  has a limiting value of  $\bar{Y}^*$  as  $\theta$  goes to infinity. Consequently, equation (29) implies that  $r^*(\theta)$  rises without bound as  $\theta$  gets large. One can also verify that  $r^*(\cdot)$  is a continuous function. Together, these observations imply that for all  $z > 0$ , there exists  $\theta$  such that  $r^*(\theta) = z$ .

To interpret this result, refer back to equations (31) and (32). Companies pay dividends whenever  $r^*(\theta) > 0$ ; thus, there are values of  $\theta$  for which dividends are part of an optimal signal. Also, repurchases are negative (new equity issues positive) whenever  $r^*(\theta) > r_2$ . From this it follows that there are values of  $\theta$  for which firms pay dividends and issue new equity simultaneously.

Although  $r^*(\theta)$  may not rise monotonically with  $\theta$ , it is negative for small  $\theta$ , and gets very large for big  $\theta$ . Loosely, we would therefore expect companies to use repurchases when signaling small differences in value, a combination of dividends and repurchases when signaling intermediate differences in value, and a combination of dividends and dividend reinvestment options when signaling large differences.

We have therefore reached a rather surprising conclusion. Despite -- or rather, because of -- tax disadvantages, firms should use dividends, and even combinations of dividends and new equity issues to signal sufficiently large differences in value. Although firms would pay lower taxes if they chose to

make the same distributions as repurchases rather than as dividends, this would not suffice to signal profitability. Because of taxes, low quality firms would have an incentive to mimic the repurchase policy, but not the dividend policy. Thus, a switch to repurchases would necessitate making larger payments to shareholders in order to preserve the integrity of the signal. Since larger distributions entail real resource costs, the net result would leave the high quality firms worse off.

It is also important to understand the economic effects of dividend taxes within the context of the current model. From equations (31) and (32), it follows immediately that an increase in the tax rate on dividends reduces both dividends and new equity issues, but raises repurchases. Total distributions always equal  $Y^*(\theta)$ , and are therefore unaffected by the dividend tax. More importantly, a change in the dividend tax rate has absolutely no effect on government revenue or welfare. If the model had included an investment decision, then dividend taxation would not affect the cost of capital. The explanation for this irrelevance result is straightforward. Firms that use dividends also employ either repurchases or new equity issues in order to achieve the optimal effective tax rate,  $\tau^*(\theta)$ . As we change  $\tau_2$ , firms respond by changing the composition of distributions in a way that preserves the same effective tax rate. Changing  $\tau_2$  does not alter the opportunity set of any manager, and therefore cannot effect real outcomes.

Throughout this analysis, I have assumed that  $\tau_1 = 0$ . In practice, it is conceivable that repurchases might be taxed more heavily than liquidations, so that  $\tau_2 > \tau_1 > 0$ . In that case, optimal payout policy would be determined by the following two equations:

$$\tau_1 y_1 + \tau_2 y_2 = \tau^*(\theta) Y^*(\theta)$$

$$y_1 + y_2 = Y^*(\theta)$$

As long as  $\tau^*(\theta) > \tau_1$ , the firm should pay some dividends. When  $\tau^*(\theta) > \tau_2$ , the optimal payout policy includes a dividend reinvestment plan.<sup>12</sup>

The effects of changing  $\tau_1$  are counter-intuitive. Specifically, raising the tax rate for repurchases stimulates repurchases, and depresses both dividends and dividend reinvestments. The explanation for this peculiar result is straightforward. An increase in  $\tau_1$  raises the firm's effective tax rate at its previously optimal payout policy. The firm must then adjust the payout mix to reestablish an effective tax rate of  $\tau^*(\theta)$ . This is accomplished by shifting to a method of distribution that is more lightly taxed.

#### 5. Optimal Signaling with Many Types of Firms

In section 4.2, I demonstrated that firms pay dividends to signal large differences in value. When these differences are sufficiently large, dividends are accompanied by issues of new equity. How do these results generalize to cases in which there are many types of firms? In traditional signaling models, each type of agent is primarily concerned with deterring imitation by the next lowest type. If the number of types is large, then differences between successive types should be small. One might therefore expect all firms to eschew dividends, and to use repurchases in equilibrium. As it turns out, this intuition is based on a false premise. Instead, the results from section 4 generalize in a straightforward way to models with many types of firms.

Recall that, in the optimal separating equilibrium, the indifference curves of the high and low quality firms are tangent. Moreover, the indifference curve for the high quality firm lies above the indifference curve for the low quality firm, except of course at the point of tangency. Thus, if  $\theta_1$ 's efficiently separate themselves from  $\underline{\theta}$ 's, then  $\theta_2$ 's will find it more difficult to deter imitation by the  $\underline{\theta}$ 's than by the  $\theta_1$ 's. Consequently, the binding incentive constraint for the  $\theta_2$ 's will concern the behavior of  $\underline{\theta}$ 's.

To illustrate an equilibrium with three types of firms, I exploit the isomorphism between choosing tax rates, and selecting a particular combination of dividends and repurchases. Let  $\theta = (\underline{\theta}, \theta_1, \theta_2)$ . For  $n=1,2$ , I assign repurchases of  $y_2^*(\theta_n, r_2)$  and dividends of  $y_2^*(\theta_n, r_2)$ , as given by equations (31) and (32) of subsection 4.1. I graph this allocation in figure 7.  $I_n^m$  represents an indifference curve for type  $\theta_n$  firms when the tax rate is  $r^*(\theta_n)$  (likewise,  $I_n^m$  is the corresponding indifference curve for type  $\underline{\theta}$  firms). In the proposed equilibrium, the  $\underline{\theta}$ 's receive point E, the  $\theta_1$ 's receive point A with tax rate  $r^*(\theta_1)$ , and the  $\theta_2$ 's receive point D with tax rate  $r^*(\theta_2)$ . We know by construction that neither the  $\theta_1$ 's or the  $\theta_2$ 's will wish to imitate the  $\underline{\theta}$ 's. Likewise, the  $\underline{\theta}$ 's have no incentive to imitate the  $\theta_1$ 's or  $\theta_2$ 's. The only remaining question is whether the  $\theta_1$ 's would imitate the  $\theta_2$ 's, or vice versa. Note that the  $\theta_2$ 's are completely indifferent between point D with tax rate  $r^*(\theta_2)$  and point C with tax rate  $r^*(\theta_1)$  (recall lemma 3). Because  $I_2^1$  is tangent to  $I_1^1$  at C, the  $\theta_2$ 's strictly prefer C to A when the tax rate is  $r^*(\theta_1)$ . Thus, they prefer their equilibrium allocation to the one assigned to the  $\theta_1$ 's. A completely symmetric argument

also implies that the  $\theta_1$ 's will not imitate the  $\theta_2$ 's. Thus, we have an equilibrium.

Figure 7 suggests that, with arbitrary numbers of types, we can construct equilibria in the following way: for each  $n$ , type  $\theta$  firms distribute the amount  $y_1^*(\theta_n, r_2)$  through repurchases, and  $y_2^*(\theta_n, r_2)$  through dividends. In suggesting this solution, I have implicitly assumed that  $r^*(\theta) > 0$  and  $Y^*(\theta) > 0$  for all  $\theta \in \Theta$  -- this is an important assumption, and I will return to it shortly. However, if this allocation is feasible, then it is an equilibrium. As in the three-type example, mutual non-imitation by each  $(\underline{\theta}, \theta_n)$  pair is assured by construction. In the appendix (lemma 4), I formally prove that this allocation satisfies the mutual non-imitation constraint for all other pairs of types. Thus, it is an equilibrium. Moreover, it is efficient in the sense that each type of firm cannot distinguish itself from inferior types at lower cost in any other signaling equilibrium.

The preceding result has an extremely important and powerful implication. In section 4.2, I discussed the relationships between  $\theta$  and dividends, repurchases, and new equity issues for the case of two types under the assumption that  $Y(\underline{\theta}) > 0$ . As long as we have  $r^*(\theta) > 0$  and  $Y^*(\theta) > 0$  for all  $\theta > \underline{\theta}$ , that discussion also characterizes the relationship between firm type and payout policy in a model with many types of firms. It is worth emphasizing the feature of equilibrium that allows us to carry over the two-type results directly to the multi-type model: each type of firm is primarily concerned with differentiating itself from the lowest type, rather than from the next lowest type. Thus, better quality firms attempt to signal larger differences in value, and therefore rely more heavily on dividends.

The best firms attempt to signal the largest differences in value, and therefore simultaneously pay dividends and use new equity.

What happens when  $r^*(\theta) < 0$  for some  $\theta \in \Theta - \{\underline{\theta}\}$ ? When there are only two types, this presents few problems: high quality firms simply signal with repurchases. However, with more than two types, it greatly complicates the task of finding equilibria. Suppose, for example, that  $r^*(\theta_1) < 0$ . Then the  $\theta_1$ 's will signal by repurchasing shares. The indifference curves of the  $\underline{\theta}$ 's and the  $\theta_1$ 's will then cross at  $\theta_1$ 's equilibrium allocation, and cross again at some higher value of  $Y$ . If  $\theta_2 - \theta_1$  is sufficiently large, the  $\theta_2$ 's will have to concern themselves with imitation by  $\underline{\theta}$ 's, as before. However, for smaller values of  $\theta_2 - \theta_1$ , the  $\theta_2$ 's will have to worry about imitation by the next lowest type, as in the standard model.

This observation raises the following question: is it possible to guarantee  $r^*(\theta) > 0$  for all  $\theta \in \Theta - \{\underline{\theta}\}$ ? I have already argued (section 4.2) that, if  $\bar{Y}(\underline{\theta}) > 0$ , then for small values of  $\theta - \underline{\theta}$ , we have  $r^*(\theta) < 0$ . One would therefore need to assume that  $\theta_1 - \underline{\theta}$  is sufficiently large. With many types of firms, this is a highly objectionable assumption. In particular, if we think of models with large but finite numbers of firms as approximations to the case where  $\Theta$  is a continuum, then  $\theta_1 - \underline{\theta}$  must be very small.

The alternative is to assume that  $\bar{Y}(\underline{\theta}) < 0$ . This implies that the signaling cost function for  $\underline{\theta}$ 's is initially concave. In effect, this simply means that the worst firms are sufficiently bad. In the appendix (lemma 5), I show that, if  $\bar{Y}(\underline{\theta}) < 0$ , then for all  $\theta$  with  $Y^*(\theta) > 0$ ,

$$[V^0(Y^*(\theta), \theta, 0) - V^0(0, \underline{\theta}, 0)] > k$$

for some  $k > 0$ . To understand the importance of this result refer back to the formula for  $r^*(\theta)$  (equation (29)). Recall that  $Y^*(\theta)$  rises monotonically with  $\theta$ , but is bounded above. Thus,  $V^0(Y^*(\theta), \underline{\theta}, 0)$  falls monotonically with  $\theta$ , but is bounded below. Let  $V_{\min}$  be the lower bound. Then it is possible to guarantee  $r^*(\theta) > 0$  for all  $\theta \in \Theta - \{\underline{\theta}\}$  as long as three conditions are satisfied:

$$(i) \quad \bar{Y}(\underline{\theta}) < 0$$

$$(ii) \quad 0 < \alpha < \frac{k}{V(0, \underline{\theta}, 0) - V_{\min}}$$

$$(iii) \quad Y^*(\theta) > 0 \text{ for all } \theta \in \Theta - \{\underline{\theta}\}.$$

Under these conditions, one can construct an equilibrium exactly as suggested at the beginning of this section.

Unfortunately, this does not fully resolve the difficulty. Whenever  $\bar{Y}(\underline{\theta}) < 0$ , then, for  $\theta - \underline{\theta}$  sufficiently small,  $Y^*(\theta) < 0$ . Consequently, conditions (i) and (iii) together imply that  $\theta_1 - \underline{\theta}$  is sufficiently large, which is exactly what we were trying to avoid. However, we can dispense with (iii) without significantly complicating the structure of equilibria.

It is easy to verify that  $Y^*(\cdot)$  is a continuous function. Supposing that  $Y^*(\theta) > 0$  for large  $\theta$ , then under condition (i) there exists  $\hat{\theta}$  such that  $Y^*(\hat{\theta}) = 0$ . Since  $Y^*(\cdot)$  is monotonic,  $Y^*(\theta) < 0$  for all  $\theta < \hat{\theta}$ , and  $Y^*(\theta) > 0$  for all  $\theta > \hat{\theta}$ .

Consider some  $\theta \in (\underline{\theta}, \hat{\theta})$ . Firms of this type have higher marginal costs of signaling than do type  $\underline{\theta}$ 's for all  $y \geq 0$ . Consequently, these firms cannot separate themselves from the  $\underline{\theta}$ 's. In equilibrium, we will have

pooling at the lower end of the quality spectrum. That is, no firm with  $\theta < \hat{\theta}$  will make any distributions. Let

$$\hat{\theta} = \{\theta \in \Theta \mid \theta \leq \hat{\theta}\}.$$

The market will correctly perceive that the average value of firms in the low quality pool is

$$\underline{V} = \left[ \sum_{\theta \in \hat{\theta}} h(\theta) \right]^{-1} \sum_{\theta \in \hat{\theta}} V^0(0, \theta, 0) h(\theta)$$

When  $\underline{\theta}$ 's choose  $y=0$ , their perceived value is  $\underline{V}$ , rather than  $V(0, \underline{\theta}, 0)$  as before.<sup>13</sup>

Higher quality firms must differentiate themselves from all types that end up in the low quality pool. Since the  $\underline{\theta}$ 's have lower marginal costs of signaling than any other type in  $\hat{\theta}$ , non-imitation by the  $\underline{\theta}$ 's implies non-imitation by all  $\theta \in \hat{\theta}$ . Thus, the analysis goes through exactly as before, except that one replaces  $V(0, \underline{\theta}, \tau)$  with  $\underline{V}$  in equation (15).

As in section 4, one can show that the optimal signal for all  $\theta > \hat{\theta}$  involves

$$(33) \quad y_1 + y_2 = Y^*(\theta).$$

However, because I have changed a constant in the incentive constraint for the  $\underline{\theta}$ 's, I must replace equation (32) with

$$(34) \quad y_2 - [\tau_2(1+\alpha)]^{-1} \{V^0(y_1 + y_2, \theta, 0) - \underline{V}\} \\ + \alpha \{V^0(y_1 + y_2, \underline{\theta}, 0) - V^0(0, \underline{\theta}, 0)\}$$

The solution to (33) and (34) is given by

$$(35) \quad y_1^{**}(\theta, \tau_2) - Y^*(\theta) \left(1 - \frac{\tau^{**}(\theta)}{\tau_2}\right)$$

and

$$(36) \quad y_2^{**}(\theta, \tau_2) - Y^*(\theta) \tau^{**}(\theta) / \tau$$

where

$$(37) \quad \tau^{**}(\theta) = [Y^*(\theta)(1+\alpha)]^{-1} \{V^0(Y^*(\theta), \theta, 0) - \underline{V}\} \\ + \alpha[V^*(\theta, \underline{\theta}, 0) - V(0, \underline{\theta}, 0)]$$

Now suppose that type  $\theta$  firms distribute  $y_1^{**}(\theta, \tau)$  through repurchases and  $y_2^{**}(\theta, \tau_2)$  through dividends for each  $\theta \in \Theta - \hat{\Theta}$  (recall that if  $\theta \in \hat{\Theta}$ , the firm does not signal). To establish that this is an equilibrium, I must first verify that these choices are feasible. When  $\theta \in \Theta - \hat{\Theta}$ ,  $Y^*(\theta) > 0$ . Analogously to lemma 5, one can show that  $V(Y^*(\theta), \theta, 0) - \underline{V} > k^*$  for all  $\theta > \hat{\theta}$ , where  $k^* > 0$  (in this case,  $k^* = V(0, \hat{\theta}, 0) - \underline{V}$ ). It follows from equation (37) that as long as  $\alpha$  is sufficiently small,  $\tau^{**}(\theta) > 0$  for all  $\theta \in \Theta - \hat{\Theta}$ . In particular, I replace condition (ii) with

$$(ii)' \quad 0 < \alpha < \frac{k^*}{V(0, \underline{\theta}, 0) - \underline{V}_{\min}}$$

Together,  $Y^*(\theta) > 0$  and  $\tau^{**}(\theta) > 0$  imply that  $y_2^{**}(\theta, \tau_2) > 0$  and  $y_1^*(\theta, \tau_2) + y_2^*(\theta, \tau_2) > 0$ , so the prescribed choices are feasible.

By construction, the mutual non-imitation constraints for each pair  $(\theta, \theta')$ ,  $\theta \in \hat{\Theta}$ ,  $\theta' \in \Theta - \hat{\Theta}$ , are satisfied. Through an argument entirely analogous to that used in the proof of lemma 4, one can also show that these

constraints are satisfied when  $\theta, \theta' \in \theta - \hat{\theta}$ . Since the equilibrium prescribes pooling for members of  $\hat{\theta}$ , there is no need to consider the case of  $\theta, \theta' \in \hat{\theta}$ . Consequently, I have established that the prescribed actions do in fact constitute an equilibrium. Moreover, in this equilibrium, all high quality firms (those with  $\theta \in \theta - \hat{\theta}$ ) separate themselves from all inferior types as efficiently as possible.

### 5.1 Implications

I will now consider the relationship between firm quality and payout policy, under assumptions (i) and (ii)'. Since  $r^{**}(\theta) > 0$  for all  $\theta > \hat{\theta}$ , all firms with  $\theta \in \theta - \hat{\theta}$  will pay some dividends. Combining (36) and (37), I obtain

$$y^{**}(\theta, r_2) = [\tau_2(1+\alpha)]^{-1} \{ [V^0(Y^*(\theta), \theta, 0) - \underline{v}] + \alpha [V^0(Y^*(\theta), \underline{\theta}, 0) - V^0(0, \underline{\theta}, 0)] \}.$$

Two important conclusions follow from inspection of this formula. First, dividends tend to rise with quality. As I pointed out in section 4.2,  $Y^*(\theta)$  is bounded, but  $V^0(\cdot)$  is not. Consequently, dividends must increase with  $\theta$  when  $\theta$  is sufficiently large. Moreover, when  $\alpha$  is small, dividends rise monotonically with  $\theta$ , even when  $\theta$  is small. Second, dividends change discontinuously as  $\theta$  moves past  $\hat{\theta}$ . For  $\theta \leq \hat{\theta}$ , firms do not pay dividends. However,

$$\lim_{\theta \downarrow \hat{\theta}} y^{**}(\theta, r_2) = [\tau_2(1+\alpha)]^{-1} [V^0(0, \hat{\theta}, 0) - \underline{v}] > 0.$$

Thus, we should not observe any firms that pay trivial levels of dividends.

Next, consider repurchases and new equity issues. In section 4.2, I argued that, with two types of firms and  $\bar{Y}(\theta) > 0$ , lower quality firms would repurchase shares, while higher quality firms would issue new equity. With  $\bar{Y}(\theta) < 0$ , this conclusion must be modified. Note in particular that

$$\lim_{\theta \downarrow \hat{\theta}} r^{**}(\theta) = \infty.$$

Thus, low quality firms pay dividends and issue new equity. Indeed, as  $\theta$  approaches  $\hat{\theta}$ , net distributions ( $Y^*(\theta)$ ) go to 0 -- dividends and new equity issues become offsetting.

This observation has another interesting implication: repurchases and new equity issues are both non-monotonic in quality. In particular,  $r^{**}(\theta)$  goes to infinity both as  $\theta$  falls to  $\hat{\theta}$ , and as  $\theta$  becomes arbitrarily large. Loosely,  $r^{**}(\theta)$  must be U-shaped. Thus, we may well have equilibria where firms issue new equity at both ends of the quality spectrum in  $\theta - \hat{\theta}$ , where only intermediate quality firms repurchase shares, and where the relationship between quality and repurchases is hump-shaped.

One final property of the relationship between firm quality and distributions deserves mention. Specifically, net distributions are bounded. This conclusion follows from two facts: first, net distributions are equal to  $Y^*(\theta)$  for  $\theta \in \theta - \hat{\theta}$ , and second,  $Y^*(\theta)$  is bounded. Thus, the mapping from quality to payout policy compresses the quality distribution, in the sense that net distributions never rise above a fixed level, regardless of how large  $\theta_N$  might be.

In section 4.2, I argued that an increase in either the dividend tax rate or the tax rate for repurchases would reduce dividends, reduce new

equity issues, raise repurchases, and have absolutely no effect on net distributions, government revenue, or welfare (nor would it affect investment if investment was modeled explicitly). All of these results carry over directly to models with many types of firms. The arguments are exactly as before.

Ordinarily, it is difficult to solve for signaling equilibria in models with continua of types. Here, this presents few problems. In particular, since the relevant incentive constraint always concerns  $\theta$ , rather than the next lowest  $\theta$ , one does not need to solve a differential equation in order to characterize the equilibrium. Indeed, practically nothing in the preceding discussion depended on the assumption that  $\theta$  is finite. With a continuum of agents, I can simply let

$$\underline{v} = \left[ \int_{\underline{\theta}} h(\theta) d\theta \right]^{-1} \int_{\underline{\theta}} v^0(0, \theta, 0) h(\theta) d\theta,$$

and proceed exactly as before.

## 5.2 A Numerical Example

In this section, I exhibit signaling equilibria for particular values of the underlying parameters. Since many of the relevant variables (e.g.,  $Y^*(\theta)$ ) are determined implicitly, I was unable to derive a closed-form, analytic solution. Instead, I have solved for equilibria numerically. In doing so, I have used a discrete approximation to the normal distribution, allowing  $\epsilon$  to take on any of 1,000 distinct values between  $-5\sigma_{\epsilon}$  and  $5\sigma_{\epsilon}$ .

I consider three different sets of parameter values. For the first or "base case" set, I take  $\underline{\theta} = 100$ ,  $\theta_N = 300$ , and  $N = 100$ . The lower bound on  $\theta$  is effectively a normalization. Experimentation revealed that an upper limit

of 300 allowed me to exhibit the most interesting properties of equilibrium, given the values of the other parameters. I used a large value of  $N$  in order to approximate the case of  $\theta = [100, 300]$ . The base case also assumes that  $\sigma_\epsilon = 35$ ,  $\lambda = 150$ ,  $r = 0.3$  (there is a 30% tax rate imposed on dividends),  $\beta = 0.25$  (companies pay a 25% premium when raising cash in order to weather a liquidity crisis), and  $\alpha = 9$  (managers attach 9 times as much weight to the actual value of the firm as they do to the firm's current market value). I also consider a "high tax case," with  $r = 0.5$  (the dividend tax rate is 50%), and a "low penalty case" with  $\beta = 0.1$  (companies pay a 10% premium when raising cash in order to weather a liquidity crisis).

One aspect of these calculations deserves emphasis. Earlier in this section, I pointed out that one must use sufficiently small values of  $\alpha$  in order to assure  $r^{**}(\theta) > 0$  for all  $\theta \in \theta - \hat{\theta}$ . This raises the possibility that the equilibria discussed in that section exist only when managers are unrealistically short-sighted. In practice, I ran into problems only for extremely large values of  $\alpha$ . The base case uses  $\alpha = 9$ , so that managers attach much less importance to current market values than to actual values. Even so, the non-negativity constraints on  $r^{**}(\theta)$  are satisfied.

Figure 8 depicts dividends as a function of quality. Note first that I obtain a heterogeneous pool of non-signalers at the lower end of the quality spectrum. At that point, dividends jump upward discontinuously, so that one does not observe any firms paying negligible dividends. For larger values of  $\theta$ , dividends rise monotonically. Raising the tax rate depresses dividends, but does not change the identity of firms that pay dividends. Lower liquidity penalties result in higher levels of dividends (higher quality

firms must substitute tax costs for non-tax costs in order to discourage imitation by lower quality firms).

Figure 9 depicts repurchases as a function of quality. Only intermediate quality firms repurchase shares. Moreover, the relationship between repurchases and quality is hump-shaped, as predicted. A higher dividend tax rate stimulates repurchases, while a lower liquidity penalty causes managers to reduce repurchases.

As shown in figure 10, the relationship between new equity issues and quality has an extremely unusual shape (as predicted in section 5.1). For the lowest values of  $\theta$ , firms issue no new equity. New issue jump up discontinuously around  $\theta = 115$  (as with dividends), but then decline monotonically when quality gets sufficiently high. A higher dividend tax rate depresses new equity issues, while a lower liquidity penalty stimulates them.

Figure 11 illustrates the relationship between net distributions to shareholders and firm quality. There is once again a heterogeneous pool of low quality firms that make no net distributions. However, as  $\theta$  rises, net distributions become positive, rise continuously, and approach an asymptote (as noted earlier, they are bounded). The dividend tax rate has no impact on net distributions.

Figure 12 displays market valuation as a function of firm quality. Note that market valuation is identical for all members of the low quality pool. It jumps discontinuously when  $\theta$  leaves this pool, and rises monotonically with  $\theta$  thereafter. The dividend tax rate has no effect on market value.

## 6. Relationship to Previous Literature

There is a close correspondence between lemma 1 and a result previously obtained by Milgrom and Roberts [1986]. These authors considered a model in which agents could either signal conventionally, or throw resources away observably ("burn money"). They showed that the conventional signal will only be used up to the point where it differentiates between types. If high quality agents must incur additional costs in order to deter imitation, then they will throw away money to make up the difference.

It is possible to cast lemma 1 in these terms. In particular, one can think of total distributions as the conventional signal, and total tax liabilities as "burned money." Firms increase the amount of money distributed to shareholders up to the point where this no longer discriminates between high and low quality firms; past that point, they throw money away by paying taxes on these distributions. They control the amount thrown away by combining dividends and repurchases appropriately.

Despite this observation, one should not be tempted to think of dividends merely as a method of burning money (analogous to advertising in the Milgrom-Roberts model). Rather, it is a linear combination of dividends and new equity issues that is equivalent to burning money. Both forms of distribution discriminate between high and low quality firms, whereas burning money does not.

Although applications of signaling theory in the area of corporate payout policy have become increasingly common (see Bhattacharya [1979, 1980], Hakansson [1982], Miller and Rock [1984], Kumar [1988], Kumar and Spatt [1987], and John and Nachman [1987]), few authors have ventured explanations

for the practice of signaling with dividends, rather than repurchases. There are, however, some notable exceptions.

Ofer and Thakor [1987] formulate a model in which there are explicit costs associated with repurchases. Unfortunately, it is not clear that these costs are large enough in practice to overcome the tax disadvantages of dividends. Moreover, the model does not account for the practice of paying dividends and issuing new equity simultaneously.

John and Williams [1985], Ambarish, John, and Williams [1987], and Williams [1988] analyze versions of a model which accounts for simultaneous use of dividends and new equity issues. In this model, firms plan to raise capital through equity markets, and therefore wish to minimize dilution by maximizing share price. Since dilution is more damaging to firms with favorable private information, it is possible to sustain a signaling equilibrium in which investors interpret any costly activity as a signal of profitability. There is, however, no particular reason to incur the required cost by distributing cash to shareholders, rather than by undertaking some other purely dissipative activity (this contrasts with the current model, where dividends do distinguish between firms of different quality). More importantly, this model can only explain the use of dividends in cases where new equity issues exceed distributions to shareholders. It cannot, for example, account for a dividend reinvestment plan in which new equity merely reduces the net amount distributed. Generally speaking, it is not at all clear that the equity market is an important source of net capital for corporations.<sup>14</sup> It is therefore very unlikely that companies pay taxable dividends in order to reduce the costs of raising new capital through equity

markets. Rather -- and this is the heart of the puzzle -- equity transactions are apparently used to reduce net distributions to shareholders.

Other non-signaling explanations for the dividend puzzle rely on assumptions that endow dividends with intrinsic advantages over other forms of distributions (see e.g., Brennan and Thakor [1989], or Bagwell and Judd [1988]). The paper by Bagwell and Judd is somewhat unique, in that it isolates some special conditions under which firms would engage in the apparently pointless practice of exposing corporate resources to taxation by paying dividends and issuing new equity simultaneously.

#### 7. Implications and Conclusions

The analysis of sections 4 and 5 is consistent with the fact that firms pay dividends, despite obvious tax disadvantages. In addition, it accounts for the practice of paying dividends and issuing new equity simultaneously (in the U.S., this is usually accomplished through dividend reinvestment plans). In contrast to other work (for example, Ofer and Thakor [1987] or Brennan and Thakor [1989]), these results do not depend on some assumed disadvantage of repurchases. Here, there are no differences between dividends and repurchases other than tax treatment. Apparently inefficient signals turn out to be efficient because of a "double crossing" property, which is in turn generated by limited liability.

In addition, the model generates a number of ancillary predictions that are consistent with casual observation, as well as with formal empirical evidence. First, it predicts that firms should "fine tune" their use of dividends, repurchases, and new equity issues in order to achieve an optimal tax rate. In particular, one would expect companies to manipulate the terms (discounts and limits) of dividend reinvestment plans in order to achieve the

desired level of reinvestment. In fact, the terms of these plans do vary widely (see Scholes and Wolfson [1989]). Second, the model predicts that there will be a potentially large pool of firms that fail to signal, despite the fact that the managers of some of these firms have better private information than others. In practice, a non-trivial fraction of firms pay no dividends and make no repurchases.<sup>15</sup> It is difficult to believe that the managers of so many firms have equivalent private information. Third, the model predicts that dividends should jump discontinuously from zero to some positive number as quality moves continuously across some threshold. In other words, even when there are many different types of firms, we should observe a trough in the distribution of dividends just above zero. Casual empiricism suggests that the distribution of dividends (as a fraction of earnings or value) is in fact bimodal, with peaks at 0 and some positive number, and with few companies making very small payments. Fourth, the model predicts that share price should rise in response to the announcement of a dividend increase or of plans to repurchase shares. This is consistent with existing evidence (see e.g., Ofer and Siegel [1986] or Dann [1981]).<sup>16</sup> Fifth, the model predicts that higher dividend taxes should depress dividends. Once again, there is considerable support for this prediction (see e.g., Poterba and Summers [1985]).

The model also generates a set of predictions that could be tested empirically, but concerning which there is (to my knowledge) no existing evidence. First, higher dividend taxes should stimulate repurchases, and depress the use of new equity issues (particularly dividend reinvestment plans) as a means of reducing net distributions.<sup>17</sup> Second, dividend taxes should have no effect on net distributions to shareholders, or on total

government revenues. Third, higher tax rates for repurchases (measured relative to effective tax rates for retained earnings) should stimulate repurchases, depress dividends and dividend reinvestments, and leave net distributions and government revenue unaffected. Fourth, the marginal effect on share price of increasing an announced increment to dividends should be positive. Fifth, the marginal effect on share price of increasing an announced repurchase may be positive or negative. It should be positive for lower quality firms, and negative for higher quality firms.<sup>18</sup> A similar observation applies for the use of dividend reinvestment plans.

The model has at least one implication that runs counter to some existing evidence. In particular, it predicts that dividend taxes are economically irrelevant (in a more fully elaborated model, they would not affect investment). Poterba and Summers [1983] have found, to the contrary, that dividend taxes do appear to affect investment.

## FOOTNOTES

1. Some analysts have pointed out that the IRS might well treat any regular distributions as taxable dividends. This argument fails to explain the widespread and consistent reluctance of firms to experiment with repurchases, and certainly cannot account for the robustness of dividend policy despite the failure of the IRS to tax repurchases as dividends in recent years.
2. In the U.S., companies often pay dividends and issue new equity simultaneously through the use of dividend reinvestment plans.
3. The observation that dividend taxation produces these opposing effects originally appeared in Gordon and Malkiel [1981]. They also conjectured that the resource savings might dominate the unit tax costs, but did not investigate this possibility formally.
4. The argument here is similar to the analysis of Milgrom and Roberts [1986].
5. Technically, the value of low quality firms might rise monotonically as one reduced  $y_1$  below 0. This suggests that low quality firms could benefit from injections of new equity. In practice, market imperfections place limits on firms' abilities to raise additional capital. In the interests of clarity and tractability, I have simply taken this limit as exogenous.
6. Composition affects total costs because tax rates on different forms of distribution may differ. But on the margin, only the  $j_{\text{FH}}$  tax rate matters, and this is common for all types.
7. In deriving this equation, I have used the fact that  $v_{Y\theta}(Y, \theta) = -v_{\theta Y}(Y, \theta)$  -- see equation (6).
8. This is also the equilibrium that would be isolated by the intuitive criterion of Cho and Kreps [1987].
9. The optimal tax result has been derived simultaneously and independently by Rotemberg [1988].
10. The careful reader may have notice that there is actually a continuum of imperfectly separating equilibria, corresponding to pairs  $(\hat{y}, \hat{V}(\hat{y}, \tau))$  lying on  $I_1$ , with  $\eta$  set appropriately. The tangency depicted in figure 6 actually defines the most efficient of these equilibria.
11. In principle, a simple pooling equilibrium could Pareto dominate the optimal separating equilibrium. The ranking of these two options depends critically on the population distribution. It is worth mentioning, however, that the pooling equilibrium does not satisfy the

Cho-Kreps intuitive criterion, whereas the imperfectly separating equilibrium and the optimal separating equilibrium do.

12. Of course, the tax system might no longer be symmetric. In particular, we might have  $r_1 > 0$  when  $y_1 > 0$ , and  $r_1 = 0$  when  $y_1 < 0$ . This alters the analysis in a perfectly straightforward way.
13. There is another class of possible outcomes that I do not consider here. Suppose that  $\hat{\theta} = (\underline{\theta}, \theta_1, \theta_2)$ . Then, in principle, the  $\theta_2$ 's might be willing and able to separate themselves from the  $\theta_1$ 's, even though they would be imitated by the  $\theta$ 's. In other words, there may be an equilibrium in which the  $\theta_1$ 's don't signal, and where the  $\underline{\theta}$ 's and  $\theta_2$ 's from a pool, and signal by making some positive distribution. While this would change the structure of equilibria for  $\theta \in \hat{\theta}$ , it would not alter any qualitative results pertaining to the behavior of higher quality firms (those with  $\theta \in \theta - \hat{\theta}$ ). In particular, those firms would still have to deter imitation by the  $\underline{\theta}$ 's, and would do so as efficiently as possible.
14. In 1989, new equity issues (other than initial public offerings) amounted to only \$22.9 billion, in comparison to roughly \$122 billion worth of dividends. To put this in perspective, note that the tax payments on dividends may have exceeded new equity issues. Moreover, stock repurchases and cash layouts have exceeded new equity issues in every year since 1984. The statistics are taken from Winkler [1990], Council of Economic Advisers [1990], and Bagwell and Shoven [1989].
15. Between 10% and 20% of the firms on the New York Stock Exchange neither pay dividends nor repurchase shares in any given year. See Barclay and Smith [1988].
16. The model also appears to predict that share price should rise when new equity issues are announced. This is contrary to the evidence (see e.g., Asquith and Mullins [1986]). However, one must distinguish new issues that raise capital from new issues that reduce net distributions (such as dividend reinvestment plans). This model only concerns the latter.
17. One must, however, be careful in interpreting the relevant data, since the explosion of repurchases during the 80's may reflect non-tax factors (e.g., it may represent attempts to deflect hostile takeovers).
18. Quality is not observable. Fortunately, the model also predicts that dividends rise monotonically with quality, so dividends could in principle be used as a proxy for quality.

## APPENDIX

Lemma 1: Consider the problem of maximizing (11) subject to (12), (13), (14), and (15). The solution satisfies equations (16) and (17).

Proof: I will simplify this problem by substituting for  $\hat{V}(\cdot)$  in the objective function, (12) and (13) using (14) and (15). Some additional manipulations (using equation (2)) yield the following equivalent problem:

$$\max_{y_1, y_2} (1+\alpha)v(y_1 + y_2, \theta) + (y_1 + y_2) - \tau_2 y_2$$

subject to

$$(A.1) \quad (1+\alpha)v(0, \underline{\theta}) \geq v(y_1 + y_2, \theta) + \alpha v(y_1 + y_2, \underline{\theta}) \\ + (1+\alpha)(y_1 + y_2 - \tau_2 y_2)$$

$$(A.2) \quad (1+\alpha)v(Y, \theta) + (y_1 + y_2) - \tau_2 y_2 \geq v(0, \underline{\theta}) + \alpha v(0, \theta)$$

I now argue that, in any solution, (A.1) must bind. Suppose not. If I raise  $y_1$  by  $\epsilon$  and lower  $y_2$  by  $\epsilon$ , then  $y_1 + y_2$  remains constant, but  $(y_1 + y_2) - \tau_2 y_2$  rises. This raises the value of the objective function, and relaxes the incentive constraint for high quality firms (equation (A.2)). For small  $\epsilon$ , (A.1) is still satisfied.

In solving this problem, I can therefore use the fact that (A.1) binds. I will proceed on the assumption that (A.2) does not bind. It is easy to verify that my solution is consistent with this assumption.

Differentiation of the Lagrangian for the preceding optimization problem yields the following first-order conditions:

$$(1+\alpha)[v_Y(y_1 + y_2, \theta) + (1-\tau_j)] = \lambda[v_Y(y_1 + y_2, \theta) + \alpha v_Y(y_1 + y_2, \underline{\theta}) \\ + (1+\alpha)(1-\tau_j)]$$

for  $j=1,2$ , where  $\lambda$  is the multiplier associated with (A.1). Some tedious manipulation of these two conditions reveals that

$$v_Y(y_1 + y_2, \theta) = v_Y(y_1 + y_2, \underline{\theta}).$$

But by the definition of  $Y^*(\theta)$ , this immediately implies (16).

Now I make use of equation (A.1). Since this constraint necessarily binds, I have

$$(A.3) \quad y_2 = [\tau_2(1+\alpha)]^{-1} ([v^0(y_1 + y_2, \theta, 0) - v^0(0, \underline{\theta}, 0)] \\ + \alpha[v^0(y_1 + y_2, \underline{\theta}, 0) - v^0(0, \underline{\theta}, 0)])$$

Substituting (16) into (A.3), I obtain (17).

Q.E.D.

Lemma 2: Suppose as in section 4.2 that there is only one method of making distributions to shareholders. Suppose also that high quality firms signal by distributing  $\hat{y}(\theta, \tau)$  (see equation (24)). Then an increase in the tax rate benefits high quality firms if and only if  $\hat{y}(\theta, \tau) > Y^*(\theta)$ .

Proof: I derive an expression for the derivative of  $\hat{y}$  with respect to the tax rate by differentiating equation (24) to obtain:

$$(A.4) \quad \hat{y}_\tau(\theta, \tau) = \frac{(1+\alpha)\hat{y}(\theta, \tau)}{v_y^0(\hat{y}(\theta, \tau), \theta, \tau) + \alpha v_y^0(\hat{y}(\theta, \tau), \underline{\theta}, \tau)}$$

Since dividends involve a cost within the range that is relevant to this analysis,  $v_y^0(\cdot) < 0$ , so  $\hat{y}_\tau(\theta, \tau)$  is negative. The equilibrium level of utility for the managers of high quality firms is given by

$$(A.5) \quad U(\theta, \tau) = (1+\alpha)v^0(\hat{y}(\theta, \tau), \theta, \tau)$$

Differentiation of (A.5) reveals that

$$(A.6) \quad U_r(\theta, \tau) = \alpha(1+\alpha)\hat{y}(\theta, \tau) \left[ \frac{V_y^0(\hat{y}(\theta, \tau), \theta, \tau) - V_y^0(\hat{y}(\theta, \tau), \underline{\theta}, \tau)}{V_y^0(\hat{y}(\theta, \tau), \theta, \tau) + \alpha V_y^0(\hat{y}(\theta, \tau), \underline{\theta}, \tau)} \right]$$

Thus,  $U_r(\theta, \tau)$  has the same sign as  $\alpha[V_y^0(\hat{y}(\theta, \tau), \theta, \tau) - V_y^0(\hat{y}(\theta, \tau), \theta, \tau)]$ , which equals the difference between the marginal cost of signaling for high and low quality firms. When  $\hat{y}(\theta, \tau) < Y^*(\theta)$ , marginal costs are higher for low quality firms, so higher taxes make these firms worse off. Conversely, when  $\hat{y}(\theta, \tau) > Y^*(\theta)$ , marginal costs are higher for high quality firms, so higher taxes make these firms better off.

Q.E.D.

Lemma 3: Suppose as in section 4.2 that there is only one method of making distributions to shareholders. Suppose that high quality firms signal by distributing  $Y^*(\theta)$ , and consider the imperfectly separating equilibrium where the fraction  $\eta$  of low quality firms also signal (see equation (30)). Then payoffs are insensitive to the tax rate.

Proof: Clearly, managers of low quality firms always receive the payoff  $(1+\alpha)V^0(0, \underline{\theta}, \tau)$ , regardless of whether or not they signal. The managers of high quality firms receive a level of utility equal to

$$(A.7) \quad U^*(\theta, \tau) = \hat{V}^0(Y^*(\theta), \tau) + \alpha V^0(Y^*(\theta), \theta, \tau)$$

I obtain an expression for  $\hat{V}^0(Y^*(\theta), \tau)$  from the indifference condition of low quality firms:

$$(A.8) \quad \hat{V}^0(Y^*(\theta), \tau) = (1+\alpha)V^0(0, \underline{\theta}, \tau) - V^0(Y^*(\theta), \underline{\theta}, \tau)$$

Substitution of (A.8) into (A.7) yields

$$U^*(\theta, \tau) = (1+\alpha)v(0, \underline{\theta}) + \alpha[v(Y^*(\theta), \theta) - v(Y^*(\theta), \underline{\theta})].$$

Note that this expression does not depend upon  $r$ .

Q.E.D.

For the next result, I need some definitions. Let

$$(A.9) \quad y^*(\theta, r) = (y_1^*(\theta, r_2), y_2^*(\theta, r_2))$$

(recall that  $r = (r_1, r_2)$ , and that I have assumed  $r_1 = 0$ ). Also let

$$(A.10) \quad \hat{W}(\theta, r) = V(y^*(\theta, r), \theta, r)$$

Finally, for  $y, t \in \mathbb{R}^2$  and  $w, h \in \mathbb{R}$ , let

$$(A.11) \quad U(y, w, h, t) = w + \alpha V(y, h, t).$$

Lemma 4: For all  $\theta, \theta' \in \Theta - \{\theta\}$  with  $\theta \neq \theta'$ ,

$$U(y^*(\theta', r), \hat{W}(\theta', r), \theta', r) > U(y^*(\theta, r), \hat{W}(\theta, r), \theta', r)$$

Remark: This lemma implies that type  $\theta'$  does not have an incentive to imitate type  $\theta$  for any  $\theta \in \Theta - \{\theta\}$ .

Proof: Define

$$\Gamma(\theta, \theta') = U(y^*(\theta, r), \hat{W}(\theta, r), \theta', r) - U(y^*(\theta', r), \hat{W}(\theta', r), \theta', r).$$

This represents the gain to type  $\theta'$  from imitating type  $\theta$ . From equations

(A.9), (A.10), (A.11), (2), (31), and (32), we have

$$(A.12) \quad \begin{aligned} \Gamma(\theta, \theta') &= \alpha [v(Y^*(\theta), \theta') - v(Y^*(\theta'), \theta')] \\ &\quad + [v(Y^*(\theta), \theta) - v(Y^*(\theta'), \theta')] \\ &\quad + (1-\alpha) [Y^*(\theta)(1-r^*(\theta)) - Y^*(\theta')(1-r^*(\theta'))]. \end{aligned}$$

By construction,

$$U(Y^*(0, \tau), \hat{W}(\theta, \tau), \underline{\theta}, \tau) = U(Y^*(\theta', \tau), \hat{W}(\theta', \tau), \underline{\theta}, \tau)$$

which implies (using (A.9), (A.10), (A.11), (2) (31) and (32)),

$$(A.13) \quad [v(Y^*(\theta), \theta) - v(Y^*(\theta'), \theta')] + (1+\alpha)[Y^*(\theta)(1-\tau^*(\theta)) - Y^*(\theta')(1-\tau^*(\theta'))] \\ - \alpha[v(Y^*(\theta'), \underline{\theta}) - v(Y^*(\theta), \underline{\theta})].$$

Substitution of (A.13) into (A.12) yields

$$(A.14) \quad \alpha^{-1}\Gamma(\theta, \theta') = [v(Y^*(\theta), \theta') - v(Y^*(\theta'), \theta')] \\ - [v(Y^*(\theta), \underline{\theta}) - v(Y^*(\theta'), \underline{\theta})].$$

Suppose  $\theta > \theta'$ . Then  $Y^*(\theta) > Y^*(\theta')$ . Note that, by (A.14),

$$(A.15) \quad \alpha^{-1}\Gamma(\theta, \theta') = \int_{Y^*(\theta')}^{Y^*(\theta)} [v_Y(Y, \theta') - v_Y(Y, \underline{\theta})] dY$$

Recall that for all  $Y > Y^*(\theta')$ ,  $v_Y(Y, \underline{\theta}) > v_Y(Y, \theta')$ . Combining this with

(A.15), we have

$$\alpha^{-1}\Gamma(\theta, \theta') < 0.$$

A completely symmetric argument holds for  $\theta < \theta'$ .

Q.E.D.

Lemma 5: Suppose  $\tilde{Y}(\underline{\theta}) < 0$ . Then for all  $\theta$  with  $Y^*(\theta) > 0$ ,

$$V^0(Y^*(\theta), \theta, 0) - V^0(0, \underline{\theta}, 0) > k$$

for some  $k > 0$ .

Proof: If  $\tilde{Y}(\underline{\theta}) < 0$ , then for  $\theta$  close to  $\underline{\theta}$ ,  $Y^*(\theta) < 0$ . Since  $Y^*(\theta)$  is continuous (this is easy to check), if there exists  $\theta$  with  $Y^*(\theta) > 0$ , then there also exists  $\hat{\theta}$  with  $Y^*(\hat{\theta}) = 0$ . Note that

$$(A.16) \quad V^0(Y^*(\hat{\theta}), \hat{\theta}, 0) - V^0(0, \underline{\theta}, 0) = k > 0$$

Since  $Y_\theta^*(\cdot) > 0$  (see section 3),  $Y^*(\theta) < 0$  for all  $\theta < \hat{\theta}$ . Therefore, we need only consider  $\theta > \hat{\theta}$ .

Trivially,

$$v_\theta(Y, \theta) = -v_Y(Y, \theta).$$

Thus,

$$(A.17) \quad V_\theta^0(Y, \theta, 0) = 1 - V_Y^0(Y, \theta, 0)$$

From (A.17), it follows that

$$\begin{aligned} \frac{d}{d\theta} V^0(Y^*(\theta), \theta, 0) &= V_Y^0(Y^*(\theta), \theta, 0) Y_\theta^*(\theta) + V_\theta^0(Y^*(\theta), \theta, 0) \\ &= -1 + V_Y^0(Y^*(\theta), \theta, 0) [Y_\theta^*(\theta) - 1]. \end{aligned}$$

But  $V_Y^0(Y^*(\theta), \theta, 0) < 1$  and  $0 < Y_\theta^*(\theta) < 1$  (see section 3), from which it follows that

$$(A.18) \quad \frac{d}{d\theta} V^0(Y^*(\theta), \theta, 0) > 0.$$

Together, (A.16) and (A.18) imply that for all  $\theta > \hat{\theta}$ ,

$$V(Y^*(\theta), \theta, 0) - V(0, \underline{\theta}, 0) > k$$

as desired.

Q.E.D.

## REFERENCES

- Ambarish, Ramasastry, Kose John, and Joseph Williams [1987], "Efficient Signalling with Dividends and Investments," Journal of Finance 42, June, 321-343.
- Asquith, Paul and David W. Mullins [1986], "Equity Issues and Offering Dilution," Journal of Financial Economics 15, 61-89.
- Bagwell, Laurie Simon and Kenneth L. Judd [1988], "Transaction Costs and Corporate Control," mimeo, Northwestern University.
- Bagwell, Laurie Simon and John B. Shoven [1989], "Cash Distributions to Shareholders," Journal of Economic Perspectives 3, Summer, 129-40.
- Barclay, M. and C. W. Smith [1988], "Corporate Payout Policy: Cash Dividends versus Open Market Share Repurchases," Journal of Financial Economics 22, 61-82.
- Bhattacharya, Sudipto [1979], "Imperfect Information, Dividend Policy, and the 'Bird in the Hand' Fallacy," Bell Journal of Economics 10 (1).
- \_\_\_\_\_ [1980], "Nondissipative Signalling: Structure and Dividend Policy," Quarterly Journal of Economics 95 (1), 1-24.
- Brennan, Michael J. and Anjan V. Thakor [1989], "Shareholder Preferences and Dividend Policy," mimeo, UCLA.
- Cho, In-Koo and David M. Kreps [1987], "Signaling Games and Stable Equilibria," Quarterly Journal of Economics 102, May, 179-221.
- Constantinides, George and Myron Scholes [1980], "Optimal Liquidation of Assets in the Presence of Personal Tax: Implications for Asset Pricing," Journal of Finance 35, 1980, 439-452.
- Council of Economic Advisors [1990], Economic Report of the President, U.S. Government Printing Office: Washington D.C.
- Dann, L. [1981], "Agency Stock Repurchases: An Analysis of Returns to Bondholders and Stockholders," Journal of Financial Economics 9, June, 113-38.
- Hakansson, Nils H. [1982], "To Pay or Not to Pay Dividend," Journal of Finance 37, 415-428.
- John, Kose and David C. Nachman [1987], "Multiperiod Financing Strategies and Dividend Management," CRSP Working Paper No. 210.
- John, Kose and Joseph Williams [1985], "Dividends, Dilution, and Taxes: A Signalling Equilibrium," Journal of Finance 40, 1053-1070.
- Kalay, Avner [1982], "Stockholder-Bondholder Conflict and Dividend Constraints," Journal of Financial Economics 10, 211-23

- Kumar, Praveen [1986], "Shareholder-Manager Conflict, Costly Monitoring and the Information Content of Dividends," mimeo, Carnegie Mellon University.
- Kumar, Praveen and Chester Spatt [1987], "Asymmetric Information and the Smoothing of Dividends," mimeo, Carnegie Mellon University.
- Milgrom, Paul and John Roberts [1986], "Price and Advertising Signals of Product Quality," Journal of Political Economy 94, 796-821.
- Miller, Merton H. and Kevin Rock [1985], "Dividend Policy under Asymmetric Information," Journal of Finance 40, 1031-1051.
- Ofer, Aharon R. and Daniel R. Siegel [1986], "Testing for the Information Content of Corporate Financial Policy Using Expectations Data: An Empirical Investigation of Special Dividends," BRC Working Paper #144, Northwestern University.
- Ofer, Aharon R. and Anjan V. Thakor [1989], "A Theory of Stock Price Responses to Alternative Corporate Cash Disbursement Methods: Stock Repurchases and Dividends," Journal of Finance 42, June, 365-394.
- Poterba, James [1987], "How Burdensome are Capital Gains Taxes?" Journal of Public Economics 33, 157-172.
- Poterba, James M. and Lawrence H. Summers [1983], "Dividend Taxes, Corporate Investment, and 'Q'," Journal of Public Economics 22, 135-167.
- Poterba, James M. and Lawrence H. Summers [1985], "The Economic Effects of Dividend Taxation," in Recent Advances in Corporate Finance, Edward I. Altman and Mark G. Subrahmanyam (eds.), Homewood, IL: Richard D. Irwin, 227-284.
- Rotemberg, Julio J. [1988], "Pareto Improving Distortionary Taxes in the Presence of Signaling," Alfred P. Sloan School of Management Working Paper No. 2039-88.
- Scholes, Myron S. and Mark A. Wolfson [1989], "Decentralized Investment Banking: The Case of Discount Dividend-Reinvestment and Stock-Purchase Plans," Journal of Financial Economics 23, 7-35.
- Stiglitz, Joseph [1983], "Some Aspects of Capital Gains Taxation," Journal of Public Economics 21, 257-94.
- Williams, Joseph [1988], "Efficient Signalling with Dividends, Investment, and Stock Repurchases," Journal of Finance 43, July, 737-747.
- Winkler, Matthew [1990], "Merrill Tops Underwriters of Stock and Bond Issues," Wall Street Journal, January 2, p. R4.

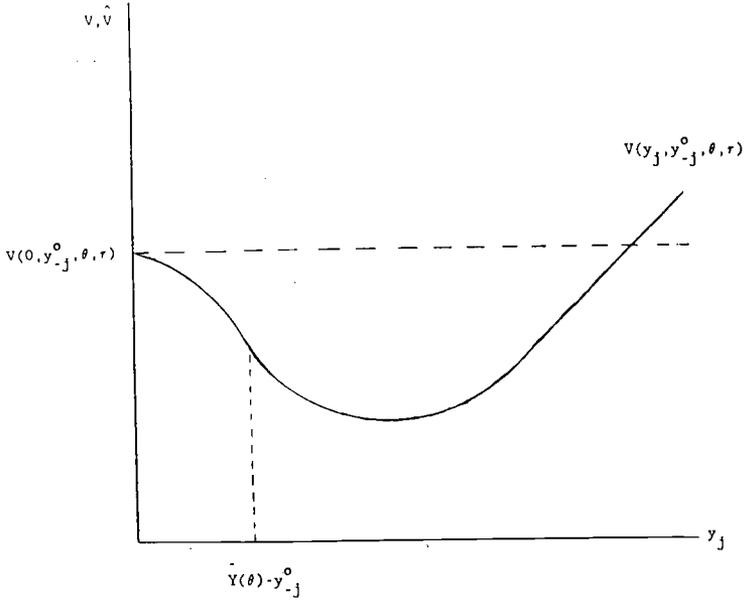


Figure 1

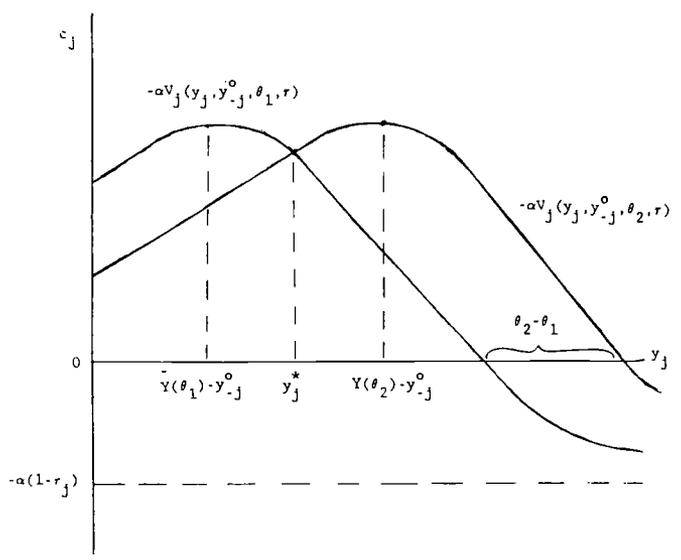


Figure 2

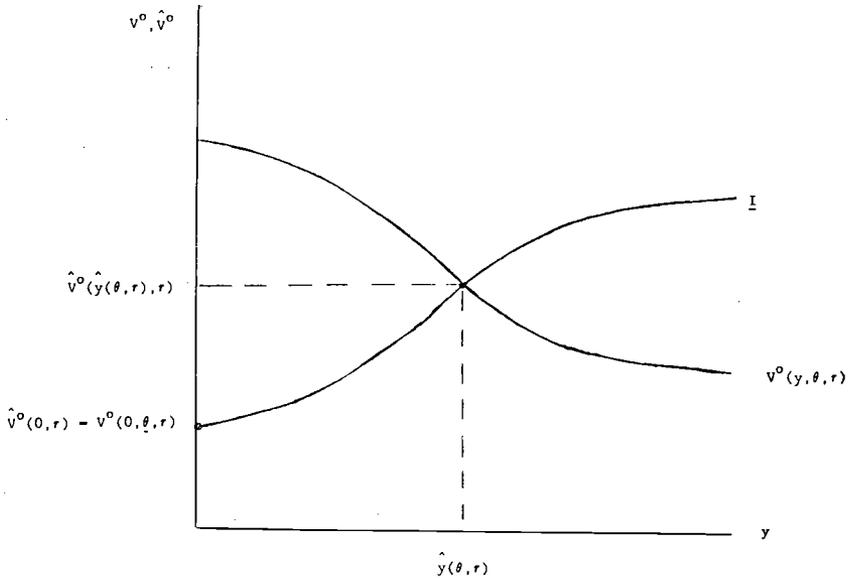


Figure 3

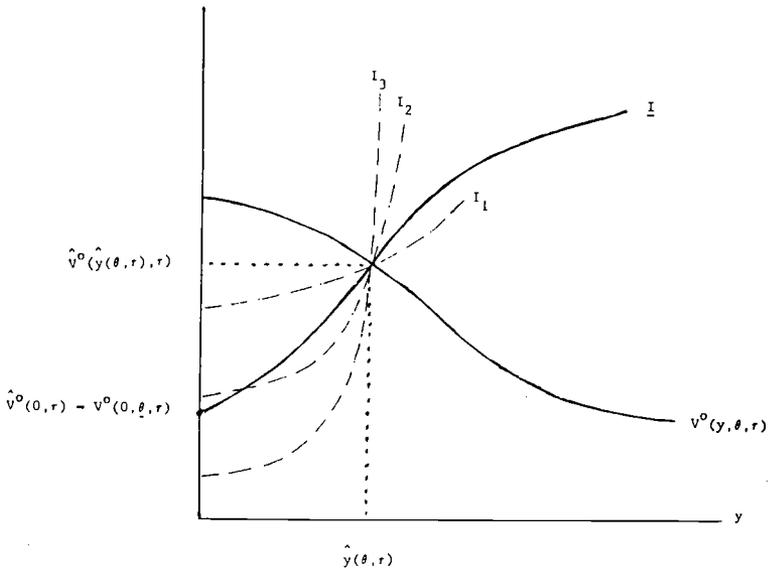


Figure 4

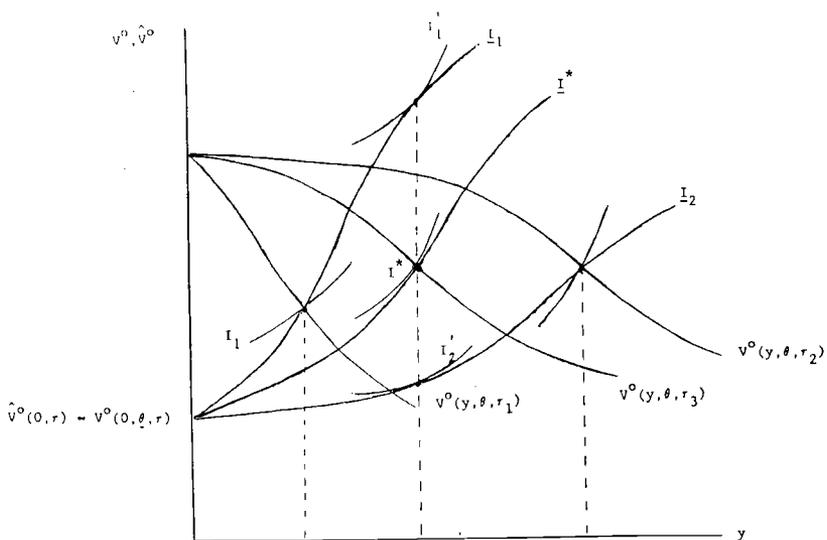


Figure 5

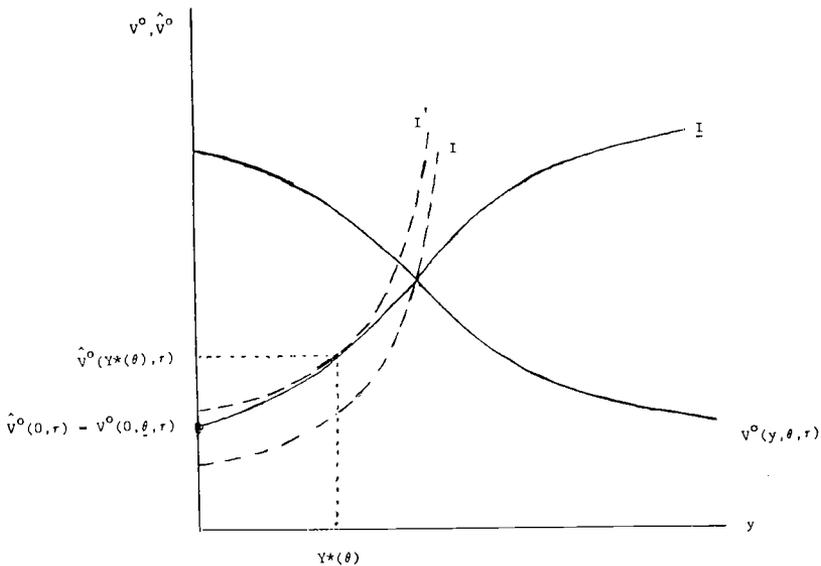


Figure 6

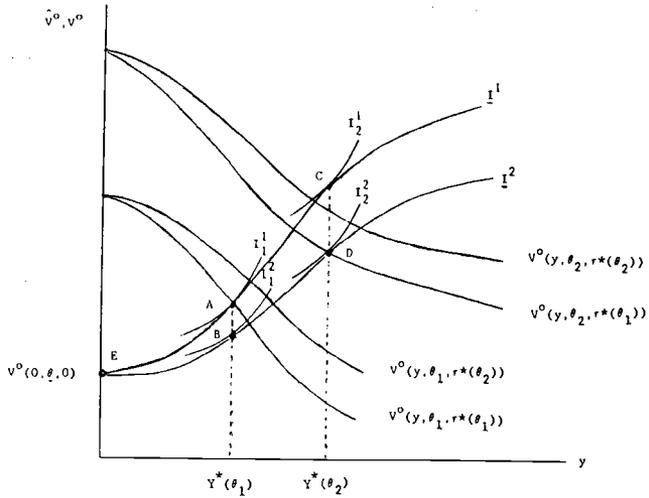


Figure 7

Figure 8: Dividends  
as a function of quality

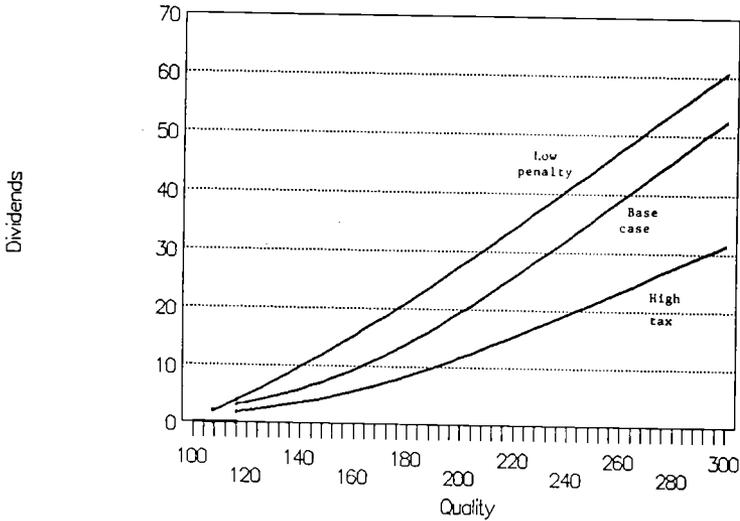


Figure 9: Repurchases  
as a function of quality

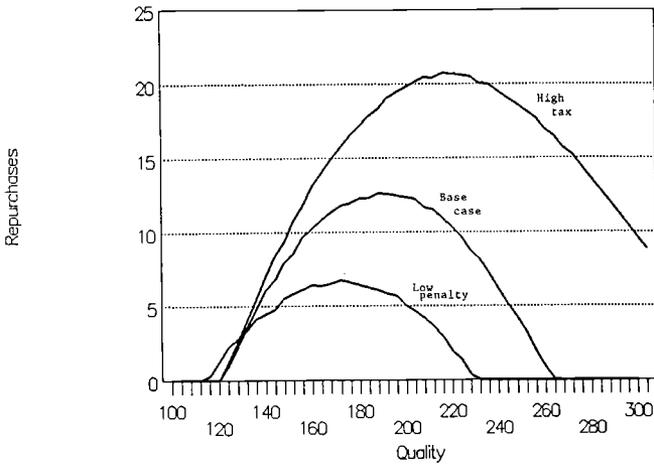


Figure 10: New Equity  
as a function of quality

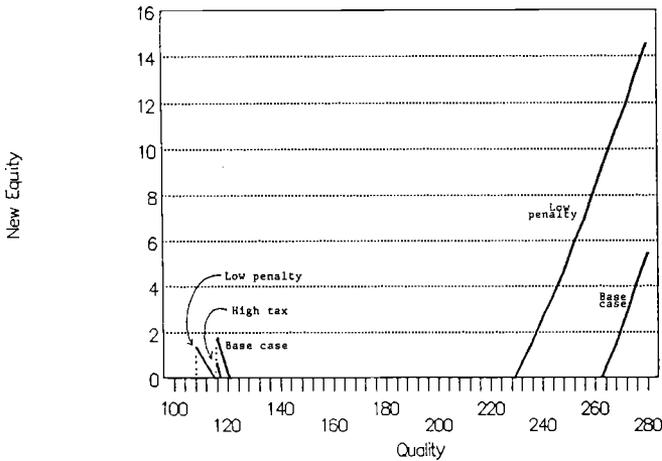


Figure 11: Total Distributions  
as a function of quality

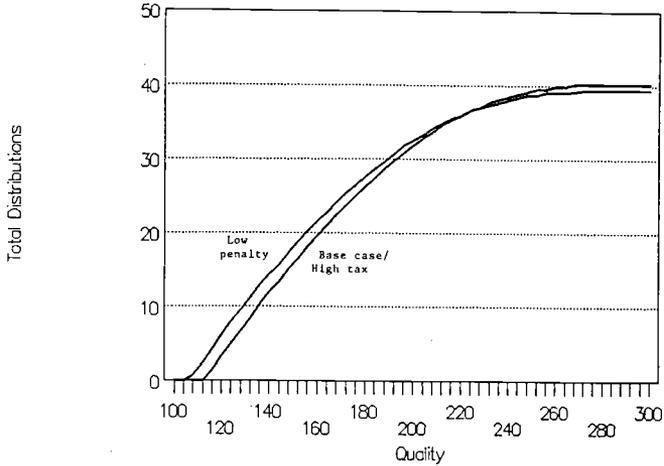


Figure 12: Market Value  
as a function of quality

