Tax Policy and the Dividend Puzzle

B. Douglas Bernheim


Stable URL:
http://links.jstor.org/sici?sici=0741-6261%28199124%2922%3A4%3C455%3ATPATDP%3E2.0.CO%3B2-R

*The RAND Journal of Economics* is currently published by The RAND Corporation.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/rand.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
Tax policy and the dividend puzzle

B. Douglas Bernheim*

This article offers a new explanation of the dividend puzzle, based upon a model in which firms 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 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.

1. Introduction

Why do companies pay dividends? This question has proved to be one of the most vexing puzzles in economics. 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.¹ The common practice of paying dividends and issuing new equity simultaneously is especially difficult to understand, since a company could reduce dividends and new equity issues by equal amounts, thereby reducing tax liabilities without altering net distributions.²

In this article I offer an explanation of the dividend puzzle, based upon a model in which firms signal profitability by distributing cash to shareholders. Throughout, I assume that dividends and repurchases are identical, except that dividends are taxed more heavily.

---

* Princeton University and National Bureau of Economic Research.

I would like to thank Laurie Bagwell, as well as seminar participants at the NBER, Northwestern University, Texas A&M University, the University of Illinois at Urbana-Champaign, and Princeton University for helpful comments. An earlier version of this article was circulated under the title “Dividends versus Repurchases as Signals of Profitability” (dated July, 1988). Financial support from the National Science Foundation (grant SES-8821666) is gratefully acknowledged.

¹ Some analysts have argued that firms fear the Internal Revenue Service would treat any regular distributions as taxable dividends. This fails to explain the widespread reluctance of firms to experiment with repurchases, as well as the robustness of dividend policy even after the IRS exhibited a reluctance to tax repurchases as dividends.

² In the United States, companies often pay dividends and issue new equity simultaneously through the use of dividend reinvestment plans.
Nevertheless, I demonstrate that under certain plausible conditions, corporations will pay dividends, rather than repurchase shares. 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.

The model has striking implications 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.

I also examine implications for government tax policy. A higher dividend tax rate depresses dividends, raises repurchases, and depresses the level of dividend reinvestment (new equity). Surprisingly, an increase in the tax rate applicable to repurchases has exactly the same effects. Moreover, in a wide range of circumstances, changes in tax rates have no effect on either total (net) distributions to shareholders, shareholder welfare, or government revenue. 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 seem counterintuitive, there is a simple explanation. If distributions involve some cost, and if this cost 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 nontax savings. When signalling 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.

The article is organized as follows. Section 2 describes the model. Section 3 investigates the properties of the signalling 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 describes some related work, and Section 7 concludes.

2. The model

The model depicts a stylized sequence of events. First, the firm undertakes productive activities and incurs either actual or potential liabilities (including debt, liability exposures, etc.). Second, managers acquire private information concerning the firm's prospects. Third, they attempt to signal this information by distributing cash to shareholders. Fourth, all uncertainty concerning net returns is resolved. Fifth, the managers decide whether or not to declare bankruptcy. Continuing operations with limited liquidity may entail a financial penalty. Finally, operations terminate, all liabilities come due, and the firm is liquidated.

In order to focus exclusively on payout decisions, I do not model the first stage explicitly.

\[\text{Gordon and Malkiel (1981) noted that dividend taxation produces these opposing effects. They conjectured that the resource savings might dominate, but did not investigate this possibility formally.}\]
Rather, I simply take capital structure and investments as given, thereby abstracting from the possibility that firms might also signal profitability by manipulating these variables.

Let \( \theta \) denote the net expected returns from a firm's activities, conditional upon private information received in the second stage. Investors know the ex ante distribution of types, \( h(\theta) \), over the type-set \( \Theta \). Throughout much of this article, I will assume that \( \Theta = \{ \theta_0, \theta_1, \ldots, \theta_N \} \), where, by convention, \( \theta_0 < \theta_1 \) and \( \theta_{n-1} < \theta_n \) for all \( n \in \{2, \ldots, N\} \).

In the third stage, managers can distribute cash to shareholders in two different forms (the extension to arbitrary numbers of forms is immediate). Let \( y_1 \) and \( y_2 \) denote, respectively, the levels of repurchases and dividends. Define \( y = (y_1, y_2)^T \) and \( Y = y_1 + y_2 \). I impose a nonnegativity constraint on dividends \( (y_2 \geq 0) \), and on total distributions \( (Y \geq 0) \). 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.\(^4\)

Distributions of the \( j \)th type are taxed at the rate \( \tau_j \). Aside from taxes, all forms of payment are completely equivalent. Let \( \tau = (\tau_1, \tau_2) \). For simplicity, I assume that the tax rate on dividends is positive \( (\tau_2 > 0) \) and that 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, then 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. The analysis is qualitatively unchanged with \( \tau_1 > 0 \) as long as \( \tau_1 < \tau_2 \).

In the fourth stage, the firm realizes net returns of \( \theta + \epsilon \), where \( \epsilon \) is a random variable with zero mean. Since it has already distributed \( Y \), its residual resources are \( R = \theta - Y + \epsilon \). When \( R \) is sufficiently low, the firm experiences liquidity problems and must raise capital at unfavorable terms in order to continue operations. Following Bhattacharya (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. Its 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 \( R < \beta(\lambda - R) \), or \( R < \beta(\lambda(1 + \beta)^{-1}) = \mu \). I assume that if a firm folds, its shareholders are completely protected from creditors.

When the firm is liquidated in the final stage, shareholders receive a payment of \( \rho(Y, \theta, \epsilon) \), defined as follows:

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

Assuming for simplicity that the discount rate is zero, the total payoff to shareholders equals the sum of terminal payments and after-tax distributions, \( \rho(Y, \theta, \epsilon) + Y - \tau Y \).

Next I describe managerial objectives. I assume that management acts to maximize the value of equity. In the fifth stage, they 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

\(^4\) Technically, the value of low-quality firms might rise monotonically as one reduced \( y_1 \) below zero. 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.
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, 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. Thus, management should also care about the actual value of the firm. Similar conclusions would follow if one assumes that managers lose their jobs in the event of bankruptcy (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. The manager's assessment of value (cum dividend) will be given by a function \( V(y, \theta, \tau) \). Managers act to maximize

\[
\hat{V}(y, \tau) + \alpha V(y, \theta, \tau),
\]

where \( \alpha \) is an exogenous parameter.

The function \( V(\cdot) \) represents an objective estimate of the firm's value under perfect information. For simplicity, I will assume that investors are risk neutral. Thus,

\[
V(y, \theta, \tau) = \int_{-\infty}^{\infty} \rho(Y, \theta, \epsilon)f(\epsilon)\,d\epsilon + Y - \tau y
= v(Y, \theta) + Y - \tau y,
\]

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

\[
\epsilon \sim N(0, \sigma^2).
\]

3. Properties of the signalling cost function

It is useful to think of the objective function (1) as \( \hat{V}(y, \tau) - c(y, \theta, \tau) \), where \( c(y, \theta, \tau) = -\alpha V(y, \theta, \tau) \) embodies the cost of signalling. In standard signalling models, one assumes that \( c_\tau > 0, c_{\theta} > 0, \) and \( c_{\theta \theta} < 0 \) (where the subscript \( j \) denotes a partial derivative with respect to \( y_j \)). These assumptions provide the basis for the “single crossing property.”

For the model considered here, the signalling cost function cannot possibly satisfy these standard assumptions. Since the firm can always elect to shut down, its value is bounded below by zero. This rules out the possibility that costs are globally convex. It is also unlikely that the marginal cost of signalling would be higher for lower-quality firms at all values of \( y \). Suppose this condition holds for small distributions. As distributions rise, the total cost for low-quality firms will approach its upper bound more quickly, for two reasons. First, since marginal cost is initially higher for lower-quality firms, total cost will approach any fixed level 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. Differentiation of \( V \) yields

\[
V_j(y, \theta, \tau) = v_\theta(Y, \theta) + (1 - \tau_j),
\]

\[
V_{jk}(y, \theta, \tau) = v_{\theta\theta}(Y, \theta).
\]

Thus, the impact of taxes on marginal cost is independent of type \( (\theta) \). Moreover, for any given form of payment, marginal cost depends only on total distributions and type; likewise for the second derivative of cost.
Given equations (4) and (5), one can deduce the key properties of the cost function by analyzing the first and second derivatives of $v(\cdot)$:

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

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

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

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

Assuming that $\epsilon$ has a normal distribution (expression (3)), we have

$$\frac{f(Y - \theta + \lambda)}{f(Y - \theta + \mu)} = \exp\left\{ \frac{1}{2}(2(Y - \theta)(\mu - \lambda) + \mu^2 - \lambda^2\sigma^2) \right\}. \quad (9)$$

Since $\mu - \lambda < 0$, this ratio falls monotonically with $Y$, and it has a limiting value of zero. Thus, the cost function must eventually become 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 $\tau$).

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. If there are no constraints on corporate distributions, the optimal strategy is to pay out an infinite amount of cash. In practice, payments are constrained by the liquid resources of the firm, as well as by contractual limitation on distributions (debt covenants). For example, covenants that prohibit managers from choosing a value of $Y$ in excess of that which minimizes $V$ eliminate extreme opportunism. Moreover, in equilibrium this 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.

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

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

FIGURE 1
THE VALUE FUNCTION
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 \( y_j^c \), there exists some \( y_j^* \) such that below \( y_j^* \) the marginal cost of distributing earnings is higher for low-quality firms, but above \( y_j^* \) high-quality firms have higher marginal costs. Because of this reversal, indifference curves cross twice, rather than once.

In deriving many of the results that follow, I exploit certain properties of \( y_j^* \). Note that for \( \theta \) and any other \( \theta \in \Theta \), \( y_j^* \) satisfies

\[
V_j(y_j^*, y_j^c, \theta, \tau) = V_j(y_j^*, y_j^c, \theta, \tau).
\]

After substituting (4), one obtains

\[
v_Y(y_j^c + y_j^*, \theta) = v_Y(y_j^c + y_j^*, \theta).
\]

(9)

Only the sum of \( y_j^c \) and \( y_j^* \) appears in equation (9). Thus, to equalize marginal costs between types \( \theta \) and \( \theta \), we need only worry about total distributions—composition is irrelevant.\(^5\) This greatly simplifies the determination of \( y_j^* \). Let \( Y^*(\theta) \) denote the value of \( Y \) that satisfies

\[
v_Y(Y, \theta) = v_Y(Y, \theta).
\]

(10)

Then \( y_j^* = Y^*(\theta) - y_j^c \). Note that the vector of tax rates, \( \tau \), does not appear in equation (10). This makes sense, since \( \tau \) has the same impact on marginal cost for all types. Thus, \( Y^*(\cdot) \) depends only upon \( \theta \).

I will need one additional result concerning \( Y^*(\cdot) \). Through implicit differentiation, it is easily verified that \( 0 < Y^*_\theta(\theta) < 1 \). In words, the level of distributions that equates marginal cost between types \( \theta \) and \( \theta \) rises with \( \theta \), but the increment is less than dollar-for-dollar.

**4. Efficient signalling and the dividend puzzle**

- A signalling equilibrium assigns a market valuation, \( \hat{V}(y, \tau) \), to each vector of signals \( y \), for a particular tax vector \( \tau \). Actual value is given by the function \( V(y, \theta, \tau) \). In equilibrium, each manager maximizes \( \hat{V}(y, \tau) + \alpha V(y, \theta, \tau) \) over \( y \), subject to \( y_2 \geq 0 \) and

\(^5\) Composition affects total costs because tax rates on different forms of distribution may differ. But on the margin, only the \( j \)th tax rate matters, and this is common for all types.
\( y_1 + y_2 \geq 0 \). The resulting pattern of choices justifies market perceptions, in the sense that \( \hat{V}(y, \tau) \) is the average value of \( V(\cdot) \) for firms that choose \( y \).

In this section, I confine attention to models with two types of firms \((\Theta = \{\theta, \bar{\theta}\})\) and analyze efficient separating equilibria. Since low-quality firms never signal, efficiency is determined exclusively by the payoff to high-quality firms: the most efficient separating equilibrium maximizes the payoff to high-quality firms over the set of all possible separating equilibria.\(^6\) Formally, the problem is to solve

\[
\max_{\hat{V}(\cdot), \tau} \hat{V}(y, \tau) + \alpha V(y, \theta, \tau)
\]

subject to

\[
\hat{V}(0, \tau) + \alpha V(0, \theta, \tau) \geq \hat{V}(y, \tau) + \alpha V(y, \theta, \tau) \quad (12)
\]

\[
\hat{V}(y, \tau) + \alpha V(y, \theta, \tau) \geq \hat{V}(0, \tau) + \alpha V(0, \theta, \tau) \quad (13)
\]

\[
\hat{V}(y, \tau) = V(y, \theta, \tau)
\]

\[
\hat{V}(0, \tau) = V(0, \theta, \tau).
\]

Equations (12) and (13) are incentive compatibility conditions. Equations (14) and (15) guarantee that expectations are confirmed in equilibrium. There are also nonnegativity constraints on \( y_1 \) and \( y_1 + y_2 \). For expositional clarity, I shall begin by ignoring the nonnegativity 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:

\[
y_1 + y_2 = Y^*(\theta)
\]

and

\[
y_2 = [\tau_2(1 + \alpha)]^{-1}\{[V^\circ(Y^*(\theta), \theta, 0) - V^\circ(0, \theta, 0)]
\]

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

where

\[
V^\circ(Y, \theta, q) = v(Y, \theta) + Y - qY.
\]

\((V^\circ(\cdot))\) describes the perfect information value of the firm under the counterfactual assumption that all distributions are taxed at the rate \( q \).

How is this solution affected by the nonnegativity constraints? The constraint \( y_1 + y_2 \geq 0 \) binds only if \( Y^*(\theta) < 0 \). In that case, for all \( j \) and any initial \( y_{-j} \), the marginal cost of increasing \( y_j \) would be greater for high-quality firms. Consequently, the total cost associated with any \( y > 0 \) 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.

The constraint \( y_2 \geq 0 \) binds whenever the right-hand side of (17) is 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 shall undertake this task shortly.

\[\Box\text{ An interpretation.}\] In this section, I offer the following interpretation of the optimal signal derived above. Suppose that firms can distribute cash to shareholders in only one

\(^6\) This equilibrium survives various refinements, such as the intuitive criterion and the D1 criterion (see Cho and Kreps (1987)). Some equilibria with incomplete separation may also survive these criteria. However, consideration of such equilibria does not alter my conclusions. See Bernheim (1990) for a discussion.
form. Then 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 signalling are smaller than in any other separating equilibrium, for any other tax rate.\(^7\) When the firm can distribute cash in several different ways, it simply combines these 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 \(\tau\) as scalars. Since there is only one form of distribution, this does not introduce 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:

\[
\hat{V}^o(0, \tau) + \alpha V^o(0, \theta, \tau) \geq \hat{V}^o(\hat{y}, \tau) + \alpha V(\hat{y}, \theta, \tau) \tag{19}
\]

\[
\hat{V}^o(0, \tau) + \alpha V^o(0, \theta, \tau) \leq \hat{V}^o(\hat{y}, \tau) + \alpha V^o(\hat{y}, \theta, \tau) \tag{20}
\]

\[
\hat{V}^o(\hat{y}, \tau) = V^o(\hat{y}, \theta, \tau) \tag{21}
\]

\[
\hat{V}^o(0, \tau) = V^o(0, \theta, \tau). \tag{22}
\]

(Following the convention introduced earlier, I use the superscript \(\text{"o"}\) 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.) Equations (19) and (20) provide for mutual nonimitation, while (21) and (22) guarantee that expectations are confirmed in equilibrium.

To exhibit a candidate 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

\[
\hat{V} + \alpha V^o(y, \theta, \tau) = k, \tag{23}
\]

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

\[
\hat{V} = k + c(y, \theta, \tau).
\]

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

**FIGURE 3**

A CANDIDATE EQUILIBRIUM

\(^7\) The optimal tax result has been derived simultaneously and independently by Rotemberg (1988).
To depict the incentive constraint for low-quality firms (equation (19)), one sets
\[ k = V^\alpha(\hat{y}, \theta, \tau)(1 + \alpha) \] (here, I have also made use of (22)). For this value of \( k \), one obtains the indifference curve \( I \) depicted in Figure 3. Together, equations (22) and (19) imply that \((\hat{y}, \hat{V}^\alpha(\hat{y}, \theta, \tau))\) must lie to the right of \( I \). Equation (22) places an additional constraint on this point. To represent this constraint, I plot the function \( V^\alpha(y, \theta, \tau) \) in Figure 3. \((\hat{y}, \hat{V}^\alpha(\hat{y}, \theta, \tau))\) must also lie on this line.

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

\[ (1 + \alpha)V^\alpha(0, \theta, \tau) = V^\alpha(\hat{y}, \theta, \tau) + \alpha V^\alpha(\hat{y}, \theta, \tau). \]

(24)

In words, \( \hat{y}(\theta, \tau) \) represents the level of distributions for which the incentive constraint for lower-quality firms just binds (see Figure 3). Of all the potential candidates for a signal, \( \hat{y}(\theta, \tau) \) 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 signalling models, the single crossing property renders this constraint redundant. Recall, however, that this model violates single crossing. To represent (20) graphically, I draw the indifference curve for the manager of a high-quality firm through the point \((\hat{y}(\theta, \tau), \hat{V}^\alpha(\hat{y}(\theta, \tau), \theta, \tau))\). There are three important cases to consider.

First, the high-quality indifference curve may cut the lower-quality indifference curve from above at \((\hat{y}(\theta, \tau), \hat{V}^\alpha(\hat{y}(\theta, \tau), \tau))\) (see the curve labeled \( I_1 \) in Figure 4). This necessarily implies that it intersects the vertical axis above \( \hat{V}^\alpha(0, \tau) \). Consequently, the incentive compatibility constraint for high-quality firms is satisfied. This is the configuration found in most standard signalling 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 signalling is lower for the high-quality firm. Thus, this first case arises whenever

\[ \hat{y}(\theta, \tau) < Y^*(\theta). \]

(25)

Second, the high-quality indifference curve may cut the low-quality indifference curve from below and nevertheless intersect the vertical axis above \( \hat{V}^\alpha(0, \tau) \). 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 nonstandard. In particular, since \( I_j \) is steeper than \( I \) at \( \hat{y}(\theta, \tau) \), the marginal cost of signaling is higher for high-quality firms. Thus, this second case arises whenever

\[
\hat{y}(\theta, \tau) > Y^*(\theta)
\]

and

\[
V^o(0, \theta, \tau) - V^o(\hat{y}(\theta, \tau), \theta, \tau) > V^o(0, \theta, \tau) - V^o(\hat{y}(\theta, \tau), \theta, \tau).
\]

Finally, the high-quality indifference curve may cut the low-quality indifference curve from below and intersect the vertical axis below \( \bar{V}(0, \tau) \) (see \( I_3 \) in Figure 4). In this case, the incentive constraint for high-quality firms is not satisfied, and there exists no equilibrium with complete separation.\(^8\) This occurs whenever (26) holds but (27) does not.

Now consider the effects of changing \( \tau \). Graphically, as \( \tau \) rises, \( I \) rotates up and \( V^o(y, \theta, \tau) \) 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 \( \tau_1 > \tau^* > \tau_2 \). \( I_k \) and \( I_k' \) represent indifference curves for a high-quality firm when the tax rate is \( \tau_k \), while \( I_k'' \) represents an indifference curve of the low-quality firm. (Recall that \( I_k, I_k', \) and \( I_k'' \) are not valid indifference curves when the tax rate is \( \tau_j, j \neq k \).)

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? A higher tax rate benefits high-quality firms (and leaves low-quality firms unaffected) if and only if \( \hat{y}(\theta, \tau) > Y^*(\theta) \). The intuition for this result is straightforward. Taxes do not affect the well-being of low-quality firms, and such firms are always indifferent between \( (0, \hat{V}^o(0, \tau)) \) and \( (\hat{y}(\theta, \tau), \hat{V}^o(\hat{y}(\theta, \tau), \tau)) \). Consider a low-quality firm that imitates 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 this imitator for the increase in tax payments. If distributions are relatively less costly on the

---

\(^8\) There may, however, exist equilibria with incomplete separation. As noted in footnote 6, consideration of such equilibria does not alter my conclusions.
margin for high-quality firms \((\hat{j}(\theta, \tau) < Y^*(\theta))\), 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 \((\hat{j}(\theta, \tau) > Y^*(\theta))\), then the same tradeoff leaves high-quality firms strictly better off.

The existence of an optimal tax rate follows immediately. I refer the reader again to Figure 5. As one reduces the tax rate from \(\tau_1\), \(\hat{j}(\theta, \tau)\) moves to the right. Since \(\hat{j}(\theta, \tau) < Y^*(\theta)\), this improves the lot of high-quality firms. For some tax rate between \(\tau_1\) and \(\tau_2\) (call it \(\tau^*\)), we obtain

\[ \hat{j}(\theta, \tau^*) = Y^*(\theta). \]  

(28)

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

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

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

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

(29)

This 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 \(y_2\) (henceforth written as functions of \(\theta\) and \(\tau_2\)) are given by the following equations:

\[ y_1^*(\theta, \tau_2) = Y^*(\theta)\left(1 - \frac{\tau^*(\theta)}{\tau_2}\right) \]  

(30)

and

\[ y_2^*(\theta, \tau_2) = Y^*(\theta)\frac{\tau^*(\theta)}{\tau_2}. \]  

(31)

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, \(\tau^*(\theta)\).

Why dividends? Lemma 1 is closely related to a result of Milgrom and Roberts (1986). Their model of pricing and advertising allows firms to either signal conventionally or throw resources away observably. The conventional signal is used only up to the point where it ceases to be less costly for high-quality firms. Beyond that, firms "burn money" to deter imitation.

It is possible to cast Lemma 1 in these terms. 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 is equally costly for high- and low-quality firms; past that point, they throw money away by exposing distributions to higher taxation. One should not, however, equate dividends with money burning. Rather, firms burn money through a particular linear combination of dividends and repurchases that has the effect of increasing tax payments while holding total distributions constant. Both dividends and repurchases discriminate between high- and low-quality firms, whereas burning money does not.

Even so, dividends are used in this model only because firms want to burn money. This is troublesome, since there are many other ways to dissipate resources, such as advertising, charitable contributions, or even demolition of valuable structures. Why should firms pay dividends rather than engage in some other dissipative activity?
In this section, I propose an answer to this question, predicated on the supposition that many dissipative activities provide managers with utility or disutility. Some managers may feel strongly about certain social causes or desire reputations as philanthropists and humanitarians. Others may crave public attention. These kinds of considerations alter the effective costs of activities like charity, advertising, and demolition.

It is important to emphasize that the mere introduction of managerial preferences over dissipative activities does not resolve the indeterminacy noted earlier. Management does not simply pick its favorite activity. To illustrate, suppose that there are $J$ different methods of dissipation. Let $y_j$ denote resources devoted to the $j$th dissipative activity. Let $1 - \tau_j$ represent the fraction of $y_j$ that is actually pocketed by shareholders. For some activities, $\tau_j$ is a tax rate, but more generally it is simply a rate of dissipation. If, for example, shareholders do not care directly about charitable contributions, then the associated $\tau_j$ is unity. Continue to interpret $y_j$ as repurchases and to assume that $\tau_1 = 0$. Suppose as before that repurchases can be negative (new equity) but that managers must respect $y_j \geq 0$ for $j \geq 2$. To keep things simple, suppose that managerial preferences are linear in $y$, so that the $j$th activity confers marginal utility of $(1 + \alpha)\gamma_j$ (the multiplicative term $1 + \alpha$ normalizes $\gamma_j$ in a convenient way). Management's objective function is then to maximize

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

where $y = (y_1, \ldots, y_J)$, $\gamma = (\gamma_1, \ldots, \gamma_J)$, and where the function $V$ has been modified in the obvious way. For simplicity, I will also assume that $\tau_j - \gamma_j > 0$ for all $j$, so that management would not undertake any dissipative activity in the absence of signalling.

With this modification, one can analyze the signalling problem exactly as before. As long as $\tau^*(\theta) > 0$, optimal signals are characterized by two conditions:

$$\sum_{j=1}^{J} y_j = Y^*(\theta) \quad \text{and} \quad (\tau - \gamma)y = \tau^*(\theta)Y^*(\theta).$$

When $J > 2$, these equations have an infinite number of solutions. This simply reflects the indeterminacy described at the outset of this section: there are many ways to throw away a given amount of money.

It is, however, rather unrealistic to assume that all managers care equally about all activities. Moreover, managers' preferences are not directly observable by stockholders. When managers have private information about the extent to which they like some activity, it becomes more difficult to signal through that activity.

I demonstrate this point as follows. Assume that, for each firm (manager), $\gamma_j$ belongs to some set $\Gamma_j (\Gamma = \lambda_j^{\gamma_j} \Gamma_j)$. Let $\gamma_j^v$ be the largest value of $\gamma_j$ in $\Gamma_j$, and let $\gamma^u = (\gamma_1^u, \ldots, \gamma_J^u)$. Investors do not observe $\gamma$ for any firm, but they know the joint distribution $\gamma$ and $\theta$. For simplicity, I assume that this distribution places positive probability on $(\theta, \gamma^u)$.

I shall continue to study equilibria with complete separation of types. In the current context, type refers to the pair $(\theta, \gamma)$. Shareholders care only about $\theta$ and $y$; $\gamma$ does not matter to them directly. Consequently, the manager of a high-quality firm need worry only about deterring imitation by all varieties of low-quality firms, $(\theta, \gamma)$ for $\gamma \in \Gamma$. Given the structure of managerial preferences, the optimal signal solves the following problem:

$$\max_{\hat{V}(\cdot, \cdot), y} \hat{V}(y, \tau) + \alpha V(y, \theta, \tau) + (1 + \alpha)\gamma y$$

subject to the incentive compatibility conditions,

$$\hat{V}(0, \tau) + \alpha V(0, \theta, \tau) \geq \hat{V}(y, \tau) + \alpha V(y, \theta, \tau) + (1 + \alpha)\gamma y \geq \hat{V}(0, \tau) + \alpha V(0, \theta, \tau),$$

as well as equations (14), (15), and nonnegativity constraints on $y_j$. 
Let $\tilde{y}$ denote the solution to this problem. It is easy to show that $\tilde{y}$ must also solve $\min_y (\tau - \gamma) y$, subject to $ey = e\tilde{y}$ (where $e$ is a vector of ones), $(\tau - \gamma^u) y = (\tau - \gamma^u) \tilde{y}$, and $y_j \geq 0$ for $j > 1$. If this were not the case, then one could rearrange the dissipative activities in a way that held $Y$ and $(\tau - \gamma^u) y$ constant while reducing $(\tau - \gamma) y$. Clearly, this would raise the value of the objective function in the previous problem, without violating either incentive constraint.

For simplicity, I shall suppose not only that $\tau_1 = 0$, but also that $\Gamma_1 = \{0\}$ (so that there is no private information about the utility received from repurchases). Using the result of the preceding paragraph, one can then easily show that, for $j > 1$, $y_j > 0$ only if $j$ solves

$$\min_j \frac{\tau_j - \gamma_j}{\tau_j - \gamma^u_j}.$$

Several important implications follow from this result. First, different firms may use different activities to signal profitability. For distinct $\gamma \in \Gamma$, different values of $j$ may minimize $(\tau_j - \gamma_j)/(\tau_j - \gamma^u_j)$. Thus, some firms may pay dividends while others advertise or make charitable contributions. Second, a signal is more likely to be used by a larger fraction of firms when shareholders have better information about its impact on managerial utility. If managers have a great deal of private information, then $\gamma^u_j - \gamma_j$ will be large for many firms, whereas if managers have little private information, $\gamma^u_j - \gamma_j$ will tend to be small for most firms. Finally, all else equal, the $j$th activity is less likely to be used as a signal when $\tau_j - \gamma^u_j$ is small.

The second and third implications mentioned above both suggest that dividends are a more likely candidate for signalling than alternatives like charitable contributions, advertising, and so forth. As mentioned earlier, these other activities have significant collateral effects for management. Some managers may well care a great deal about social causes, public visibility, or reputation for philanthropy, which suggests that $\tau_j - \gamma^u_j$ may be very low for these activities. Moreover, the degree to which managers care about such things probably varies widely. In contrast, dividends have few collateral effects about which management might care. There is little to counterbalance the tax disadvantages of dividends or to introduce uncertainty about management’s preferences.\footnote{9}

I have described conditions under which dividends would emerge as the most common signal of profitability. Except in extreme cases, some firms will use other methods of signalling even when these conditions are satisfied. For example, the managers who receive the most utility from public recognition will still find it desirable to signal profitability through highly visible dissipative activities. Thus, the behavior of flamboyant, publicity-seeking executives is consistent with the model and does not undermine the analysis as a theory of dividends.

\section*{Implications}

Different values of $\theta$ correspond to distinct payout patterns. To illustrate, fix $\theta$, and suppose that the cost function for the lowest-quality firm is initially convex ($\bar{Y}(\theta) > 0$). First consider small values of $\theta$. Recall that $Y^*(\theta) > \bar{Y}(\theta)$ for all $\theta > \theta$. Therefore, as $\theta$ goes to $\theta$, $Y^*(\theta)$ is bounded away from $0$. It follows from equation (29) that, for sufficiently small values of $\theta - \theta$, $\tau^*(\theta)$ is negative. Consequently, managers use repurchases to signal small differences in value. Next consider large values of $\theta$. As $\theta$ goes to infinity, $V(y, \theta, 0)$ rises without bound (see equation (2)). Define $Y^*$ as the value of $Y$ that minimizes $V(Y, \theta)$. It is easy to show that $Y^*(\theta) < \bar{Y}^*$ for all $\theta > \theta$ and that $Y^*(\theta)$ has a limiting value of $\bar{Y}^*$. Consequently, equation (29) implies that $\tau^*(\theta)$ rises without
bound as $\theta$ gets large. Now refer to equations (30) and (31). Repurchases are negative (new equity issues positive) whenever $\tau^*(\theta) > \tau_2$. It follows that for sufficiently large $\theta$, firms pay dividends and issue new equity simultaneously. For intermediate values of $\theta$, $0 < \tau^*(\theta) < \tau_2$, so firms pay dividends and repurchase shares.

Firms could, of course, reduce investors’ tax liabilities by making the same net distribution exclusively through a share repurchase. But 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.

I turn next to the effects of dividend taxation. From equations (30) and (31), an increase in the tax rate on dividends reduces both dividends and new equity issues, but it raises repurchases. Total distributions always equal $Y^*(\theta)$ and are therefore unaffected by the dividend tax. More important, a change in the dividend tax rate has no effect on government revenue or welfare. In a model with an investment decision, 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 the manager, and therefore it cannot affect real outcomes.

So far, I have assumed that $\tau_1 = 0$. In practice, 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:

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

and

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

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.10

The effects of changing $\tau_1$ are counterintuitive. Assuming that dividends are positive, raising the tax rate for repurchases stimulates repurchases and depresses both dividends and dividend reinvestment. There is a simple explanation for this peculiar result. Since $\tau_1 < \tau_2$, an increase in $\tau_1$ raises the firm’s effective tax rate at its previous 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 signalling with many types of firms

Derivation of equilibria. In standard signalling models, the behavior of each type is determined by the incentive constraint of the next lowest type. If we take $|\theta|$ to be large, then it is natural to assume that differences between successive types ($\theta_n - \theta_{n-1}$) are small. Given the results of Section 4, one might then expect all firms to distribute cash exclusively through repurchases. As it turns out, this intuition is based on a false premise.

Recall that, in the optimal separating equilibrium, the indifference curves of the high- and low-quality firms are tangent—the indifference curve for the high-quality firm never crosses below the indifference curve for the low-quality firm. Thus, if $\theta_1$’s efficiently separate themselves from $\theta_2$’s, then $\theta_2$’s will find it more difficult to deter imitation by the $\theta_1$’s than

---

10 Of course, the tax system might no longer be symmetric. In particular, we might have $\tau_1 > 0$ when $y_1 > 0$, and $\tau_1 = 0$ when $y_1 < 0$. This alters the analysis in a perfectly straightforward way.
by the $\theta_1$'s. Consequently, the binding incentive constraint for the $\theta_2$'s will concern the behavior of $\theta$'s—not the $\theta_1$'s.

Figure 6 illustrates an equilibrium with three types of firms ($\Theta = \{\theta, \theta_1, \theta_2\}$). For $n = 1, 2$, I assign repurchases of $y^*_1(\theta_2, \tau_2)$ and dividends of $y^*_2(\theta_n, \tau_2)$, as given by equations (30) and (31). $I^*_n$ represents an indifference curve for type $\theta_n$ firms when the effective tax rate is $\tau^*(\theta_n)$ (likewise, $I^*_m$ is the corresponding indifference curve for type $\theta$ firms). In the proposed equilibrium, the $\theta$'s receive point $E$, the $\theta_1$'s receive point $A$ with tax rate $\tau^*(\theta_1)$, and the $\theta_2$'s receive point $D$ with tax rate $\tau^*(\theta_2)$. By construction, neither the $\theta_1$'s or the $\theta_2$'s wish to imitate the $\theta$'s, and the $\theta$'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. It is easy to check that the $\theta_2$'s are indifferent between point $D$ with tax rate $\tau^*(\theta_2)$ and point $C$ with tax rate $\tau^*(\theta_1)$. Because $I_2^1$ is tangent to $I_1^1$ at $C$, the $\theta_2$'s strictly prefer $C$ with $\tau^*(\theta_1)$ to $A$ with $\tau^*(\theta_1)$. A completely symmetric argument also implies that the $\theta_1$'s will not imitate the $\theta_2$'s. Thus, we have an equilibrium.

With arbitrary numbers of types, one can try to construct equilibria in an analogous way: for each $n$, type $\theta_n$ firms distribute $y^*_1(\theta_n, \tau_2)$ through repurchases and $y^*_2(\theta_n, \tau_2)$ through dividends. As long as $\tau^*(\theta) > 0$ and $Y^*(\theta) > 0$ for all $\theta \in \Theta$, this allocation is in fact an equilibrium. As in the three-type example, mutual nonimitation by each $(\theta, \theta_n)$ pair is assured by construction. In the Appendix (Lemma 2), I prove that the mutual nonimitation constraints are also satisfied for all other pairs of types. This equilibrium is also efficient in the sense that each type of firm distinguishes itself from inferior types at the lowest possible cost.

Unfortunately, the assumption that $\tau^*(\theta) > 0$ and $Y^*(\theta) > 0$ for all $\theta$ is problematic. Suppose first that $\tilde{Y}(\theta) > 0$. With many types, it is natural to assume that $\theta_1 - \theta$ is small. But then, as argued in Section 4, we must have $\tau^*(\theta_1) < 0$.

What happens when $\tau^*(\theta) < 0$ for some $\theta \in \Theta - \{\theta\}$? When there are only two types, high-quality firms simply signal with repurchases. However, with more than two types, the analysis becomes much more complex. If $\tau^*(\theta_1) < 0$, the $\theta_1$'s signal by repurchasing shares. The indifference curves of the $\theta$'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 $\theta$'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 standard models.

FIGURE 6
EQUILIBRIA WITH THREE TYPES
The alternative is to assume that

\[(i) \quad \tilde{Y}(\theta) < 0.\]

In that case, for \(\theta - \tilde{\theta}\) sufficiently small, \(Y^*(\theta) < 0\), which again contradicts our premise. Fortunately, this does not significantly complicate the analysis.

It is easy to show that \(Y^*(\cdot)\) is continuous. If \(Y^*(\theta) > 0\) for large \(\theta\), then under condition (i) there exists \(\tilde{\theta}\) such that \(Y^*(\tilde{\theta}) = 0\). Since \(Y^*(\cdot)\) is monotonic, \(Y^*(\theta) < 0\) for all \(\theta < \tilde{\theta}\), and \(Y^*(\theta) > 0\) for all \(\theta > \tilde{\theta}\).

Consider some \(\theta \in (\tilde{\theta}, \hat{\theta})\). Firms of this type have higher marginal costs of signalling than do type \(\theta'\)'s for all \(y \geq 0\). Consequently, these firms cannot separate themselves from the \(\theta'\)'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 \(\hat{\Theta} = \{\theta \in \Theta \mid \theta \leq \tilde{\theta}\}\). The market will correctly perceive that the average value of firms in the low-quality pool is

\[V = \left[ \sum_{\theta \in \hat{\Theta}} h(\theta) \right]^{-1} \sum_{\theta \in \hat{\Theta}} V^c(0, \theta, 0) h(\theta). \quad (32)\]

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

As in Section 4, one can show that the optimal signal that differentiates each \(\theta > \hat{\theta}\) from \(\hat{\theta}\) involves

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

However, because I have changed a constant in the incentive constraint for the \(\theta'\)'s, I must replace equation (17) with

\[y_2 = [\tau_2 (1 + \alpha)]^{-1} \{[V^c(y_1 + y_2, \theta, 0) - V]\]

\[+ \alpha[V^c(y_1 + y_2, \hat{\theta}, 0) - V^c(0, \hat{\theta}, 0)]\}. \quad (34)\]

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

\[y_1^{**}(\tau_2) = Y^*(\theta) \left(1 - \frac{\tau^*(\theta)}{\tau_2}\right) \quad (35)\]

and

\[y_2^{**}(\tau_2) = Y^*(\theta) \tau^*(\theta)/\tau_2, \quad (36)\]

where

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

\[+ \alpha[V^c(Y^*(\theta), \hat{\theta}, 0) - V^c(0, \hat{\theta}, 0)]\}. \quad (37)\]

Suppose that type \(\theta \in \Theta - \hat{\Theta}\) firms distribute \(y_1^{**}(\tau, \tau)\) through repurchases and \(y_2^{**}(\tau, \tau_2)\) through dividends (recall that if \(\theta \in \hat{\Theta}\), the firm does not signal). To establish that this is an equilibrium, I first verify feasibility. When \(\theta \in \Theta - \hat{\Theta}\), \(Y^*(\theta) > 0\). In the Appendix (Lemma 3), I show that \(V^c(Y^*(\theta), \theta, 0) - V > V(0, \hat{\theta}, 0) - V > 0\) for \(\theta > \hat{\theta}\). It follows from equation (37) that as long as \(\alpha\) is sufficiently small, \(\tau^*(\theta) > 0\) for all \(\theta \in \Theta - \hat{\Theta}\). The following is a sufficient condition:

\[(ii) \quad 0 < \alpha < \frac{V(0, \hat{\theta}, 0) - V}{V(0, \hat{\theta}, 0) - V_{\min}},\]

where \(V_{\min} = \min_{y > 0} V^c(Y^*(\theta, 0), 0)\). \(^{11}\) Together, \(Y^*(\theta) > 0\) and \(\tau^*(\theta) > 0\) imply that \(y_2^{**}(\theta, \tau_2) > 0\) and \(y_1^{**}(\theta, \tau_2) + y_2^{**}(\theta, \tau_2) > 0\), as required.

\(^{11}\) When \(V_{\min} = V(0, \hat{\theta}, 0), \tau^*(\theta) > 0\) for all \(\theta > \hat{\theta}\) follows for all \(\alpha > 0\).
By construction, the mutual nonimitation constraints for each pair \((\theta, \theta')\), \(\theta \in \hat{\Theta}, \theta' \in \Theta - \hat{\Theta}\), are satisfied. Analogously to Lemma 2, one can show that these constraints are satisfied for \(\theta, \theta' \in \Theta - \hat{\Theta}\). Since pooling is prescribed for members of \(\hat{\Theta}\), there is no need to consider the case of \(\theta, \theta' \in \hat{\Theta}\). Consequently, I have established that, under conditions (i) and (ii), the proposed actions do in fact constitute an equilibrium. Moreover, in equilibrium, all high-quality firms (those with \(\theta \in \Theta - \hat{\Theta}\)) separate themselves from all inferior types as efficiently as possible. Neither of the two key conditions (i or ii) is particularly demanding—in words, they simply require that managers place sufficient weight on market perceptions and that the lowest-quality firm be sufficiently bad.

This analysis extends easily to cases in which \(\Theta\) is a continuum. Since the relevant incentive constraint always concerns \(\hat{\theta}\), rather than the next lowest \(\theta\), one does not need to solve a differential equation, as in more standard models. In effect, one merely replaces the summations in (32) with integrals and proceeds as before.

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

\[
y^*_2(\theta, \tau_2) = [\tau_2(1 + \alpha)]^{-1} \{V^\circ(Y^*(\theta), \theta, 0) - V^\circ\} + \alpha[V^\circ(Y^*(\theta), \theta, 0) - V^\circ(0, \theta, 0)].
\]

Two conclusions follow. First, dividends tend to rise with quality. Since \(Y^*(\theta)\) is bounded while \(V^\circ(\cdot)\) is not, dividends increase with \(\theta\) when \(\theta\) is sufficiently large. For small \(\alpha\), dividends also rise monotonically with \(\theta\) even when \(\theta\) is small. Second, dividends change discontinuously as \(\theta\) moves past \(\hat{\theta}\). For \(\theta \approx \hat{\theta}\), firms do not pay dividends. However,

\[
\lim_{\theta \to \hat{\theta}} y^*_2(\theta, \tau_2) = [\tau_2(1 + \alpha)]^{-1}[V^\circ(0, \hat{\theta}, 0) - V^\circ] > 0.
\]

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

Next consider repurchases and new issues. In Section 4 I argued that with \(N = 1\) and \(\tilde{Y}(\theta) > 0\), lower-quality firms would repurchase shares, while higher-quality firms would issue new equity. With \(\tilde{Y}(\theta) < 0\), this conclusion must be modified. Note that \(\lim_{\theta \to \hat{\theta}} r^*(\theta) = \lim_{\theta \to -\infty} r^*(\theta) = \infty\). Thus, firms pay dividends and issue new equity at both ends of the quality spectrum in \(\Theta - \hat{\Theta}\). As \(\theta\) approaches \(\hat{\theta}\), net distributions (\(Y^*(\theta)\)) go to zero—dividends and new equity issues become offsetting. Only firms of intermediate quality repurchase shares. These observations suggest that the relationship between quality and new equity issues is bimodal, while the relationship between quality and repurchases is hump-shaped. Numerical computations confirm this expectation (see Bernheim (1990)).

Finally, note that net distributions are bounded. This conclusion follows from two facts: net distributions are equal to \(Y^*(\theta)\) for \(\theta \in \Theta - \hat{\Theta}\), and \(Y^*(\theta)\) 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 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 no effect on net distributions, government revenue, or welfare (nor would it affect investment if investment were modelled explicitly). All of these results carry over directly to models with many types of firms.

## 6. Relationship to Previous Literature

- Although applications of signalling theory in the area of corporate payout policy have become increasingly common (see Bhattacharya (1979, 1980), Hakansson (1982), Miller and Rock (1985), Kumar (1988), Kumar and Spatt (1987), and John and Nachman
few authors have ventured explanations for the practice of signalling with dividends rather than repurchases. There are, however, some notable exceptions.

John and Williams (1985) analyze a model that accounts for the simultaneous use of dividends and new equity issues. In their model, firms plan to raise equity capital, and therefore they wish to minimize dilution by maximizing share price. Since dilution is more damaging to firms with favorable private information, it is possible to sustain equilibria in which investors interpret costly dividends as signals of profitability. Ambarish, John, and Williams (1987) endogenize investment within this model and study signalling through the optimal dividend-investment mix. Under somewhat restrictive assumptions, Williams (1988) extends the analysis to cases where the set of possible quality types is a continuum.

There are several important differences between the current article and those by Ambarish, John, and/or Williams (henceforth AJW). First, the AJW model places several implicit restrictions on transactions with incumbent shareholders. One consequence is that dividends and repurchases are not equivalent (apart from taxes). Firms cannot arrange a payout policy so as to produce any given effective tax rate on distributions without also changing other real variables. As a result, managers' opportunity sets are not invariant with respect to the dividend tax rate. Changes in this tax rate therefore affect the real characteristics of equilibrium. A second related point concerns dividend reinvestment plans. Since dividend reinvestment raises capital without introducing new shareholders, it does not create dilution. Consequently, the AJW model cannot explain the practice of simultaneously paying dividends and issuing new equity to incumbent shareholders through dividend reinvestment plans. More generally, the importance of dilution is questionable. Since equity markets do not provide corporations with a significant net flow of resources, it seems unlikely that companies pay taxable dividends in order to reduce the costs of raising equity capital. Finally, AJW do not consider the possibility (discussed in Section 4) that firms can deplete resources through a variety of activities, and that any such activity might substitute for, or even improve upon, taxable dividends.

Other explanations for the dividend puzzle rely on assumptions that endow dividends with intrinsic advantages over other forms of distributions (see, e.g., Ofer and Thakor (1987), Brennan and Thakor (1989), or Bagwell and Judd (1988)). It is, however, difficult to identify factors that are important enough in practice to overcome the tax disadvantages of dividends. Moreover, most of these articles do not attempt to explain the practice of paying dividends and issuing new equity simultaneously (the article by Bagwell and Judd is somewhat unique in this respect).

---

12 Total distributions to incumbent shareholders are limited by their liquidity demand. Firms are not allowed to increase repurchases by offering a price in excess of the current share price. Also, firms cannot vary both dividends to and repurchases from incumbents without changing investment. In fact, large repurchases necessitate low investment, whereas large dividends do not. This is restrictive, since in principle shares could be repurchased from incumbents while new equity is issued to new investors. Finally, firms cannot freely vary new capital raised from incumbents (e.g., by adjusting the terms on dividend reinvestment plans). Indeed, firms can only raise capital by selling stock to new investors.

13 From equation (15) in Ambarish, John, and Williams (1987), it is clear that $t$, the tax rate on dividends, affects equilibrium investment. Likewise, equation (12) of Williams (1988) implies that the tax rate on dividends, $\gamma$, affects total revenue, $\gamma D^*(x)$.

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 probably 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 One might object that the current model is similarly predicated on the assumption that the nontax costs of financial distress are large enough to counterbalance the tax disadvantages of dividends. This objection is misguided. In the current model, when the costs of illiquidity are low, high-quality firms must incur substantial tax liabilities in order to deter imitation. Thus, dividends tend to be higher when $\beta$ is lower.
Another related strand of literature considers the possible irrelevance of dividend taxation. Under Auerbach's (1979) "tax capitalization" hypothesis, dividend taxation does not affect investment; however, in contrast to the current model, dividend policy is not sensitive to the dividend tax rate, and increases in dividend taxation augment government revenues. Other authors (e.g., Miller and Scholes (1978) and Black and Scholes (1974)) have argued that dividend taxes are avoidable, and therefore completely irrelevant. In contrast to the current article, their analysis implies that dividend taxation should not affect payout policy and that it should raise no revenue.

7. Implications and conclusions

- My analysis is consistent with the fact that firms pay dividends despite tax disadvantages. It also explains the practice of paying dividends and issuing new equity simultaneously (e.g., through dividend reinvestment plans). In addition, it generates a number of ancillary predictions that are consistent with both casual observation and 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. 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 practice, the terms of these plans vary widely (Scholes and Wolfson, 1989). Second, the model predicts the existence of a heterogeneous, low-quality pool of firms that distribute nothing. This prediction is consistent with empirical patterns.\textsuperscript{16} Third, the model suggests that dividends should jump discontinuously from zero to some positive number as quality moves across some threshold. In practice, there appears to be a trough in the distribution of the dividend-to-price ratio near zero. Fourth, the model predicts that share price should rise in response to the announcement of a dividend increase or plans to repurchase shares. This is consistent with existing evidence (see, e.g., Ofer and Siegel (1986) or Dann (1981)).\textsuperscript{17} Fifth, the model predicts that higher dividend taxes should depress dividends. There is considerable support for this prediction (see Poterba and Summers (1985)).

The model also generates a set of testable predictions 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).\textsuperscript{18} 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 dividends should always be positive, but the marginal effect of increasing repurchases may be positive or negative, depending on firm quality (it should be positive for lower-quality firms and negative for higher-quality firms).\textsuperscript{19} A similar observation applies for the use of dividend reinvestment plans.

The analysis has at least one implication that runs counter to some existing evidence. In particular, in a more fully elaborated model, it would imply that dividend taxes should

\textsuperscript{16} 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).

\textsuperscript{17} 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.

\textsuperscript{18} One must, however, be careful in interpreting the relevant data, since the explosion of repurchases during the 1980s may reflect nontax factors (e.g., it may represent attempts to deflect hostile takeovers).

\textsuperscript{19} 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.
not affect investment. This prediction is inconsistent with the findings of Poterba and Summers (1983).

Appendix

Proofs of Lemmas 1, 2, and 3 follow.

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 \( \hat{\varphi}(\cdot) \) 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

\[
(1 + \alpha)[v(y_1 + y_2, \theta) + \alpha v(y_1 + y_2, \theta) + (1 + \alpha)(y_1 + y_2 - \tau_2 y_2)] \geq v(0, \theta) + \alpha v(0, \theta).
\]

(1 + \alpha)[v(y_1 + y_2, \theta) + (y_1 + y_2) - \tau_2 y_2] \geq v(0, \theta) + \alpha v(0, \theta).

I now argue that, in any solution, (1A) 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 constraint (A2). For small \( \epsilon \) (A1) is still satisfied.

I will proceed on the assumption that (A2) does not bind. It is easy to verify that my solution is consistent with this assumption. Differentiation of the Lagrangian yields the following first-order conditions:

\[
(1 + \alpha)[v(y_1 + y_2, \theta) + (1 - \tau_2)] = \lambda[v(y_1 + y_2, \theta) + \alpha v(y_1 + y_2, \theta) + (1 + \alpha)(1 - \tau_2)]
\]

for \( j = 1, 2 \), where \( \lambda \) is the multiplier associated with (A1). Some tedious manipulation of these two conditions reveals that \( \nu_j(y_1 + y_2, \theta) = v(y_1 + y_2, \theta) \). But by the definition of \( Y^\star(\theta) \), this immediately implies (16).

Now I make use of equation (A1). Since this constraint necessarily binds, I have

\[
y_2 = \{\tau_2(1 + \alpha)^{-1} \{[V^\alpha(y_1 + y_2, \theta, 0) - V^\alpha(0, \theta, 0)] + \alpha[V^\alpha(y_1 + y_2, \theta, 0) - V^\alpha(0, \theta, 0)]\} \}
\]

(A3)

Substituting (16) into (A3), I obtain (17). Q.E.D.

For the next result, I need some definitions. Let

\[
y^\star(\theta, \tau) = (y^\star_1(\theta, \tau), y^\star_2(\theta, \tau)).
\]

(A4)

(Recall that \( \tau = (\tau_1, \tau_2) \), and that I have assumed \( \tau_1 = 0 \)) Also let

\[
\tilde{W}(\theta, \tau) = V(y^\star(\theta, \tau), \theta, \tau).\]

(A5)

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

\[
U(y, w, h, t) = w + \alpha V(y, h, t).
\]

(A6)

Lemma 2. For all \( \theta, \theta' \in \Theta \cup \{\emptyset\} \) with \( \theta \neq \theta' \),

\[
U(y^\star(\theta', \tau), \tilde{W}(\theta', \tau), \theta', \tau) > U(y^\star(\theta, \tau), \tilde{W}(\theta, \tau), \theta, \tau).
\]

Proof: Define

\[
\Gamma(\theta, \theta') = U(y^\star(\theta, \tau), \tilde{W}(\theta, \tau), \theta', \tau) - U(y^\star(\theta', \tau), \tilde{W}(\theta', \tau), \theta', \tau).
\]

This represents the gain to type \( \theta' \) from imitating type \( \theta \). From equations (A4), (A5), (A6), (2), (30), and (31), we have

\[
\Gamma(\theta, \theta') = \alpha[v(Y^\star(\theta, \theta') - v(Y^\star(\theta', \theta'))] + \alpha[Y^\star(\theta)(1 - \tau^\star(\theta)) - Y^\star(\theta')(1 - \tau^\star(\theta'))]
\]

\[
+ (1 + \alpha)[Y^\star(\theta)(1 - \tau^\star(\theta)) - Y^\star(\theta')(1 - \tau^\star(\theta'))].
\]

(A7)

By construction,

\[
U(y^\star(\theta, \tau), \tilde{W}(\theta, \tau), \theta, \tau) = U(Y^\star(\theta', \tau), \tilde{W}(\theta', \tau), \theta', \tau),
\]

which implies (using (A4), (A5), (A6), (2), (30) and (31))

\[
[v(Y^\star(\theta, \theta') - v(Y^\star(\theta', \theta'))] + (1 + \alpha)[Y^\star(\theta)(1 - \tau^\star(\theta)) - Y^\star(\theta')(1 - \tau^\star(\theta'))]
\]

\[
= \alpha[v(Y^\star(\theta', \theta') - v(Y^\star(\theta', \theta'))].
\]

(A8)
Substitution of (A8) into (A7) yields

\[ \alpha^{-1} Y(\theta, \theta') = [v(Y(\theta, \theta')) - v(Y(\theta', \theta'))] - [v(Y(\theta, \theta)) - v(Y(\theta', \theta))]. \]  

(A9)

Suppose \( \theta > \theta' \). Then \( Y(\theta) > Y(\theta') \). Note that, by (A9),

\[ \alpha^{-1} Y(\theta, \theta') = \int_{Y(\theta)}^{Y(\theta')} [v_Y(Y, \theta') - v_Y(Y, \theta)] dY. \]  

(A10)

Recall that for all \( Y > Y(\theta') \), \( v_Y(Y, \theta') > v_Y(Y, \theta) \). Combining this with (A10), we have \( \alpha^{-1} Y(\theta, \theta') < 0 \). A symmetric argument holds for \( \theta < \theta' \). \textit{Q.E.D.}

Lemma 3. Suppose \( \bar{Y}(\theta) < 0 \). Then for all \( \theta \) with \( Y(\theta) > 0 \),

\[ V^*(Y(\theta), \theta, 0) - V = V(0, \hat{\theta}, 0) - V > 0. \]

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

\[ V^*(Y(\tilde{\theta}), \tilde{\theta}, 0) - V = V(0, \hat{\theta}, 0) - V > 0. \]  

(A11)

Since \( Y(\cdot) > 0 \) (see Section 3), \( \bar{Y}(\theta) < 0 \) for all \( \theta < \tilde{\theta} \). Therefore, we need only consider \( \theta > \tilde{\theta} \). Trivially, \( v_Y(Y, \theta) = -v_Y(Y, \theta) \). Thus,

\[ V^*(Y, \theta, 0) = 1 - V^*(Y, \theta, 0). \]  

(A12)

From (A12), it follows that

\[ \frac{d}{d\theta} V^*(Y(\theta), \theta, 0) = V^*(Y(\theta), \theta, 0) Y_p(Y(\theta) + V^*(Y(\theta), \theta, 0) = 1 + V^*(Y(\theta), \theta, 0) [Y_p(\theta) - 1]. \]

But \( V^*(Y(\theta), \theta, 0) < 1 \) and \( 0 < Y_p(\theta) < 1 \) (see Section 3), which implies

\[ \frac{d}{d\theta} Y_p(Y(\theta), \theta, 0) > 0. \]  

(A13)

Together, (A11) and (A13) imply that for all \( \tilde{\theta} > 0 \),

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

as desired. \textit{Q.E.D.}

References


